The Fundamental Program of the Calculus

February 5, 2007. BETA Version 0.1

Alan Smithee

DIRECT CORRESPONDENCE TO ALAN SMITHEE
E-mail address: calculusprogram@gmail.com

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

Key words and phrases. derivative, Dini derivatives, fundamental theorem of the calculus, Lebesgue integral, Henstock-Kurzweil integral, Stieltjes integral, variation of a function, Lebesgue measure, variational measures, Vitali covering theorem.

Abstract. Submitted, for your consideration, a set of calculus notes from some unlikely future, covering the essential elements of the theory of the derivative, integral, and measure on the real line.

Contents

Preface

On a holiday (to Vienna and the Czech Republic) I encountered a time-travelling student from the 22nd century who was taking his spring break in our era before returning to resume his calculus studies. This fact might not have arisen were it not for a situation, embarrassing to me, in which I offered to help him with his calculus homework and found that I was entirely unfamiliar with the material, this in spite of many years of teaching the subject.

He was equally embarrassed since it was considered, naturally, an ethical breach to be travelling with copies of time-sensitive material. Nonetheless, he graciously gave me a copy of his classroom notes and it is these notes that I present here as being of some possible interest to the mathematical community in general and the designers of calculus curricula in particular.

Not fully confident in my ability to assess the notes correctly, I approached several of my colleagues. The first, a young woman in whom I had the greatest confidence, took them from me and returned them later in the afternoon. "Curious," she said. I could elicit no further opinions from her and she wandered off deep in thought, about something else no doubt. I could not help but agree with this one word commentary and my confidence in the notes was somewhat shaken.

I considered then that I needed to find a more outspoken, but equally distinguished colleague. I approached one, timidly with the notes in hand, and asked for his assessment. He glanced impatiently at his watch but agreed to take a look at them.

As he read some passages from the notes, he began sputtering and grew incandescent with rage. "What . . . what? Riemann's unfortunate integral! The Lebesgue integral? Not the correct integral? Correct integral indeed!"

He then threw the notes at me and yelled "The writer is an ignoramus. The Lebesgue integral *is* the correct version of the integral, as anyone who knows anything at all about mathematics would surely realize. Does he know nothing of modern probability theory, functional analysis, harmonic analysis, ..." He stormed off continuing to list applications of the Lebesgue integral as he disappeared from view. I could not help but agree with him and my confidence in the notes was quite shaken.

I decided then that, since these were in fact calculus notes, I should approach my good friend who was recognized as the best teacher in the department and who had spoken at numerous conferences on mathematical education and calculus reform. He was pleased to be consulted and agreed to read the notes overnight. The next morning he returned them to me with an indulgent smile.

"Rather interesting, but seriously flawed. This is no way to teach the calculus! The most important idea to get across is the mean-value theorem of course. Look here ... no fanfare, and the theorem is never used. Never used! Imagine. Granted we never give a full proof, but to treat it this way ... well!

"Then these covering relations. I confess to being absolutely baffled by these things, full covers, fine covers and so on. But really ... partitions are covering relations? These covering relation things really obscure the true nature of a partition. You see really what the partition is, is the set of points where the subdivisions are made. In studies students have been videotaped in small group settings and the consensus has emerged that partitions are clearly understood if presented in this way, especially with appropriate teaching aids."

He then rushed to a blackboard where he described to me with various pictures how partitions were sets of points, and then there were associated points too, and the fineness of the partition was given by certain distances, not at all the strange notion of "fineness" in the notes, and so on. I could not follow the argument and got lost in the maze of pictures, but I could not help but agree that his was by far the better way to teach the subject. He ambled off continuing to list further pedagogical defects. By now my confidence in the notes was entirely shattered.

Not knowing what to do next I considered destroying all of the material. But I reasoned that, just as historians are always searching for documents from the past to analyse, so too there might be a dual version of history in which documents from the future play a role. Thus I offer the notes here, with only some minor editing to make the language sound rather more contemporary.

I thought for a time that I could perform an historically useful function and track the source of the ideas and refer to the authors of these ideas. I have been assured by an expert that, while the ideas are well-known, the authorship is murky and many are certain to take offense. Most likely, he said, all of the integration and variational ideas can be found at least in the papers (published and unpublished) of Ralph Henstock from the 1950s and 1960s. The task, however, of reading the many papers and finding which ideas are not so attributable would take more dedication to the project than I can imagine applying.

Alan Smithee

Direct correspondence to Alan Smithee at calculusprogram@gmail.com.

Comments of the Reviewer

This manuscript should not be accepted. Indeed it should be destroyed!

At a first glance the submission appears to be a text in elementary real analysis intended for undergraduate mathematics majors who have just completed their calculus courses. It covers much of the elements for the theory of differentiation and integration in dimension one. All the basics are there: sequences, series, open and closed sets, derivatives, upper and lower derivatives, Dini derivatives, Lebesgue measure and integral, Lebesgue differentiation theorem, Baire category theorem, Vitali covering theorem, the fundamental theorem of the calculus.

But a closer look reveals unusual and disturbing features. The History section is not at all history: it seems to be some kind of polemic. The Riemann integral makes no appearance, nor do improper integrals. Definitions are by no means the conventional, currently acceptable ones. Standard compactness arguments have been replaced by appeals to the little known Cousin lemma. The Lebesgue integral is defined, not by the usual route, but directly from some kind of coverings. The Vitali covering theorem is barely recognizable. The integral itself is again defined by these coverings and is neither the Riemann integral (as one would hope in an undergraduate class), nor the Lebesgue integral (as one would demand in a graduate class), but the obscure and arcane Henstock-Kurzweil integral.

To tell the truth, these covering relations monstrosities that dominate the discussion give me a headache. Intervals and points, and ordered pairs of them, and collections of those, and filters of those in turn do not clarify anything and just leave the reader spinning. Am I really expected to consider the filter of Cousin covers as a differentiation basis that is the central concern of elementary analysis?

I think we do not have to take seriously the introduction where the author (Alan Smithee indeed?) describes the notes as prepared by a student travelling from some distant future. Clearly our unfortunate era of unending imperial wars, suicide bombers, and economic woes could not (will not?) offer any interest to tourists from the 22nd century. Nor do we have to take seriously that this offering is in any sense a real textbook. So what is it?

There is a familiar conjurer's trick here, often played before, whereby an order of concepts appears in which some later theorem becomes an early definition and some old definition reemerges later as a leisurely theorem. The magician prepares his apparatus backstage and then, to the amazement of the audience, can pull rabbits out of hats. Here there are plenty of rabbits.

But do we want the traditional order and presentation of our subject matter to be treated in this cavalier fashion? Take the measure theory for example. I remember a happy year learning this from Halmos's text and would never impose any other presentation on my students. I remain as addicted to what my professor called "Halmos's greatest contribution to mathematics" (black box) as I am to that way of developing the Lebesgue integral. Should I entertain for a moment a strange development based on covering relations?

I expect this author is greatly amused by all of this in a solipsistic sense. But the rest of us must live in the real world where we are obliged to use standard terminology, the usual arguments, and the popular textbooks. Treat it as a curiosity and go back to your usual calculus books and the usual analysis texts (baby Rudin?) for any serious teaching of integration theory.

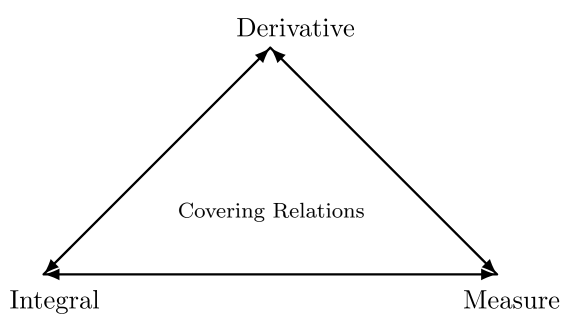
The Reviewer

History

This seems not to be a genuine history in any sense that I would understand. Perhaps it is just some sketchy notes taken by an undergraduate student during a calculus lecture and presents the odd perspective of that teacher on his subject. ...A. S.

The Fundamental Program of the Calculus. The connections among the derivative/integral/measure concepts on the real line are traditionally called "the fundamental program of the calculus." The fundamental program of the calculus, as originally outlined by Newton, is to determine the correct relation between the concepts of integral and derivative which Newton saw as mutually inverse. A third aspect of the program was revealed by Lebesgue in the early 20th century who developed the appropriate measure theory for the program and established the first connections between the measure and the derivative. The correct version of the integral appeared only in the second half of the 20th century and by the end of that century the fundamental program of the calculus was essentially complete, three hundred years after Newton's first insights.

We now recognize these three fundamental concepts (derivative, integral, measure) as three different aspects of a single structure—the notion of a covering relation. The fact that the three concepts are intimately related is then transparent. This triangle should help us visualize the structure of the theory and keep us on track in the search to complete the program:



Newton. Newton, by the end of the 17th century, had described the derivative and integral in clear, but not formally rigorous terms. He conceived of the relation between them as mutually inverse. In our language, an integral

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

for Newton inherits its meaning from the relation $F'(x) = f(x)$ and is assigned the value

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

His intention is better revealed by taking the following as Newton's definition, a definition he would doubtless approve:

> *The relation*
>
> $$\int_a^b f(x)\,dx = F(b) - F(a)$$
>
> *signifies that $F'(x) = f(x)$ at all points of the interval $[a, b]$ excepting some set of negligible points and that the function F does not grow on that negligible set of points.*

For the 18th century one could have used "negligible set" to indicate a finite set. Then continuous functions are the ones that do not grow on finite sets and the definition reads: $F'(x) = f(x)$ at all points of the interval $[a, b]$ excepting a finite set of points and the function F is continuous on $[a, b]$. The mean-value theorem of the calculus justifies the definition.

By the late 19th century a larger class of negligible sets (the sets of zero content) were being used and the class of Lipschitz functions employed for indefinite integrals. Much confusion remained and many problems led to technical difficulties that could be resolved only with later developments.

By the early 20th century the necessary apparatus was in place for a full realization of Newton's definition. The negligible sets are taken as sets of measure zero for Lebesgue's new theory of measure and the class of functions that do not grow on such sets was described by Vitali's notion of absolute continuity. Clearly much analysis of sets, measures, and functions is needed before any real progress with Newton's original conception can be made.

Berkeley. The rest of the 18th century sees a development of the computational and application aspects of Newton's ideas, but little formal development. There are significant criticisms, most notably from the philosopher Berkeley, about the rigorous level of the subject. Following an era that had rediscovered the ancient Greek mathematician Euclid and his contemporaries, the calculus seemed to be lacking in precisely that high level of reliability and authority that ancient geometry afforded.

Berkeley's criticism is worth recounting since it tells much about the thinking at the time as well as including a moral lesson for philosophers thinking to comment in areas in which they have little actual knowledge.

> "Now the other Method of obtaining a Rule to find the Fluxion of any Power is as follows. Let the Quantity x flow uniformly, and be it proposed to find the Fluxion of x^n. In the same time that x by flowing becomes $x + o$, the Power x^n becomes $\overline{x+o}|^n$, *i.e.* by the Method of infinite Series
>
> $$x^n + nox^{n-1} + \frac{nn-n}{2}\,oox^{n-2} + \&c.$$
>
> and the Increments
>
> $$o \text{ and } nox^{n-1} + \frac{nn-n}{2}\,oox^{n-2} + \&c.$$

> are to one another as
>
> $$1 \text{ to } nx^{n-1} + \frac{nn-n}{2}\, ox^{n-2} + \&c.$$
>
> the Increments vanish, and their last Proportion will be 1 to nx^{n-1}. But it should seem that this reasoning is not fair or conclusive. For when it is said, let the Increments vanish, *i.e.* let the Increments be nothing, or let there be no Increments, the former Supposition that the Increments were something, or that there were Increments, is destroyed, and yet a Consequence of that Supposition, *i.e.* an Expression got by virtue thereof, is retained. Which, by the foregoing Lemma, is a false way of reasoning. Certainly when we suppose the Increments to vanish, we must suppose their Proportions, their Expressions, and every thing else derived from the Supposition of their Existence to vanish with them." ... [From *The Analyst*," by George Berkeley.]

This offers us an interesting glimpse into the past. There are, indeed, serious flaws in the reasoning, but not the ones pointed out by Berkeley. First, not having a proper definition for the limit process, we see that Berkeley has been led astray by interpreting the process of "letting the increments vanish" as meaning "let the increments be set equal to zero." The error is Berkeley's not Newton's; a calculus student making the same error would be considered naive. But a serious error does enter here. There are two limiting processes, one involves letting the increments vanish and the other is the infinite sum (itself definable only as a limiting process). Changing the order of these two processes (as occurs here) is not justified and, in general, can lead to error. Thus the formula is very much in doubt, but not for the reason that Berkeley has given.

The moral, if we can draw one, is that a lack of rigorous definitions and adequate theory leads to serious errors and a limitation on the methods. Perhaps most 18th century mathematicians were unimpressed with criticisms of the foundations of the calculus. After all the methods were undeniably successful. What they may have failed to appreciate, however, was how this lack of proper foundations was a true impediment to future progress. Indeed, there was a general *fin de siecle* attitude among scholars that the calculus was a completed subject by 1800, an attitude that was utterly wrong.

Cauchy. A clarification of both the limit process underlying the derivative and the integral itself is given in the early 19th century by Cauchy. Cauchy offered little in the way of encouragement or pleasant attitudes to his contemporaries, but left a legacy of good mathematical works in compensation.

An integral $\int_a^b f(x)\,dx$ can be computed by finding the function F that has f as its derivative. The value is $F(b) - F(a)$. The mean value theorem helps a little, since then $F(b) - F(a) = f(\xi)(b-a)$ for at least one value ξ in $[a, b]$. Similarly, subdividing the interval into smaller pieces by

$$a = a_0 < a_1 < a_2 < \dots a_{n-1} < a_n = b$$

one sees that

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

for certain choices of ξ_i in the interval $[a_{i-1}, a_i]$. If we make arbitrary choices of ξ_i' in the interval $[a_{i-1}, a_i]$, then the altered sum approximates (perhaps) the integral:

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

with an error that cannot exceed

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

[Here, as usual, $\omega f([c,d])$ denotes the oscillation of f on an interval $[c,d]$, the amount by which f might vary in the interval.]

Cauchy restricted attention to continuous functions and was able to realize the integral for all such functions as a limiting process involving "Cauchy sums," i.e., sums of the form

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

taken over partitions

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

of an interval $[a,b]$. This constructs the integral in a sequence of operations. Thus an integral, while originally an operation inverse to differentiation, now has a distinct definition and a constructive procedure attached to its computation.

Cauchy's procedure here is restricted to continuous functions. He considers, too, functions with a finite number of discontinuities, and offers other procedures for these as well.

Riemann's Unfortunate Integral. Riemann, in mid 19th century, gave a series of lectures wherein he took Cauchy's integration ideas and formalized them. The integral was defined, exactly following Cauchy, as a limit process applied to Cauchy sums over a partition. But, whereas Cauchy had intended the procedure to apply to continuous functions, Riemann examined the full class of functions for which the process would work.

Now it is a legitimate academic exercise to take an earlier technique and ask just how far it may be applied. But for an empty exercise, as this turned out to be, to then become a method is most unfortunate. For it to be widely adopted as the correct integration theory is beyond unfortunate.

Riemann's considerable prestige, no doubt, added to the influence that this integral had on the calculus. But the integral, in retrospect, can be seen as a mistake. Riemann had not spotted the simple and correct definition used here in our notes, but selected a narrower interpretation closer to Cauchy's original method. This method works for continuous functions and for bounded functions having negligibly many discontinuities, but does not cover the full scope of Newton's program. Riemann's definition, followed unimaginatively by so many others of the time, thus doomed the calculus program to a period of stagnation and confusion.

The flaws in this theory appeared soon enough. It was recognized, by Cauchy in fact, that an integral should have the extension property of Theorem 14.5. During that century this is incorrectly assumed to be a procedure that must be applied to an integral in order to *extend* it to handle unbounded functions, rather than an essential feature of the integral itself. The Riemann integral does not have this extension property.

Another development had a serious impact on a number of mathematicians. That was the construction of a closed subset P of $[0,1]$ that has measure arbitrarily close to 1 and yet P contains no intervals. From that one can easily construct a bounded derivative f discontinuous at each point of P. Such a function cannot be Riemann integrable. This indicates clearly that the Riemann integral fails to promote the fundamental program of Newton, who would have had no hesitation in integrating such a function.

Lebesgue's "Integral." Lebesgue offered in 1901, as motivation for working on his integration techniques, precisely this example of a bounded derivative too discontinuous to be Riemann integrable. Newton's calculus program alone would demand that such a function should be treatable by any respectable integration method.

By 1901 there was a clearer understanding that a proper development of the calculus required a measure theory, a way of assigning a generalized length to sets of real numbers. The measure theory that seemed appropriate to the study of Riemann's integral was developed in the last part of the 19th century by Peano, Jordan, and Cantor. This measure, which we might denote as $\mathcal{P}J$, is related to ours by the formula

$$\mathcal{P}J(E) = \mathcal{L}(\overline{E}).$$

In particular this measure is unable to distinguish between a closed set and a dense subset.

Lebesgue focussed on the Peano-Jordan-Cantor measure theory and correctly saw that the appropriate measure theory for the full program of the calculus must have certain essential properties. His development of the measure theory leads to the study of absolutely integrable functions (soon after called Lebesgue-integrable) but only in their measure aspects. Thus his "integral" is not an integral as such, but more correctly viewed as the right-hand side of the expression in Theorem 8.9, namely

$$\int_a^b f(x)\,dx = \int_{[a,b]} [f(x)]^+\,dx - \int_{[a,b]} [f(x)]^-\,dx \tag{0.1}$$

where the integrals on the right side of the identity are given by elaborate measure-theoretic constructions, no longer by limits of Cauchy sums.

This approach forces the theory of the integral to a study of functions for which at least one (but usually both) of the terms in the difference in identity (0.1) is finite. Such an integral, while powerful enough for most applications, does not complete the calculus program for there are derivatives $F' = f$ with both of these terms infinite.

Curiously Lebesgue did feel obliged to show that his integral could be obtained as a limit of Cauchy sums, just as tradition seemed to demand. But the limit process was not transparent enough to serve as a definition. No mathematician at the time even considered that a simple and natural definition of an integral from Cauchy sums was available that would completely answer the needs of the calculus program as well as include the integral described by Lebesgue's constructions.

Denjoy. In the first half of the 20th century Lebesgue's measure-theoretic viewpoint dominated analysis and the connections between the derivative and his integral were established directly with considerable skill and insight, not at all by means of the more elementary methods of this course.

Although an interpretation of the integral by an identity such as (0.1) cannot handle all derivatives, it was widely considered in those years that the program was now complete anyway.

Several authors, however, continued the program. Denjoy in 1916 realized that, to integrate all derivatives, an integral must possess both of the extension properties of Theorems 14.5 and 14.6. A natural method to describe such an integral is to continue to extend Lebesgue's integral in a transfinite sequence of such extensions until the process stabilizes. The resulting integral is known as Denjoy's integral, or more commonly in that period as the Denjoy-Perron integral because of a description of the same integral by Perron of an entirely different nature.

Denjoy claimed for his integral a constructibility and severely lampooned Perron's integral as a meaningless artifact. We would have to agree with Denjoy in that he does accurately set out the constructive nature of the integration process. But his methods proved too arcane to be considered a theory of integration. By 1916 only the more ambitious mathematicians were willing to embrace even Lebesgue's methods, let alone those of Denjoy.

The Integral. Throughout the 20th century Lebesgue's integral and Lebesgue's measure-theoretic methods, applied to higher dimensions and various abstract settings, took a prominent place among the main tools of analysis. The Denjoy-Perron integral was largely viewed as an historical curiosity, too baroque for most tastes and not of much use. It described exactly the integral which the Newton program had demanded, but, by doing so in a cumbersome and technically inferior manner, doomed it seemingly to the footnotes of history.

The correct integral, in its simplicity, was discovered independently by Henstock and Kurzweil in the mid 1950s. Henstock was investigating an integration theory developed by Ward, while Kurzweil was working on a problem in differential equations. This integral was quickly identified as equivalent to the Denjoy-Perron integral. Its transparent nature immediately recommends itself as the proper formal way to study the program. The variational techniques that we now use were developed by Henstock and, by the end of the century, it was recognized that these techniques allowed the fundamental program to be completed in the natural and simple way we now present in calculus courses.

The Fundamental Covering Lemma of the Calculus The integral and much of the elementary calculus development depends on a simple covering argument, now known as the *fundamental covering lemma of the calculus*, or sometimes as the Cousin covering lemma.

The historical name originates from a paper in the late 19th century by the Belgian mathematician, Pierre Cousin. It seems entirely unreasonable to attribute the lemma to him, but so far no earlier volunteer has stepped forward. For a while it looked like his more famous French contemporary, Edouard Goursat, could claim the honors for a paper published in 1900 but Cousin's paper predates his by a little bit. The French and the Italians have long hoped to claim the honors for this little lemma, but so far the Belgians retain the prize.

The lemma is both elementary and fundamental to the whole calculus program. Hardly any calculus presentations of the 20th century give it such a prominent role; few even mention it. Curiously, though, the lemma has a peculiar and persistent habit of rediscovery. An avid reader of late 20th century copies of *American Mathematical Monthly* will find that it reappears with nearly the regularity and twice the frequency of the cicadas. Such papers emerge as authors discover how most of the elementary calculus compactness proofs can be easily framed as applications of the lemma. None seem to be aware of the unoriginality of their contribution. One assumes that a certain class of mathematician of the time was fond of publishing in the Monthly but not much interested in reading any but the most recent issues.

Failure to use the lemma obscures the calculus program and leads to the standard oversight of the integration theory—using the incorrect integral. For more advanced subjects the compactness argument of choice is patterned after the Bolzano-Weierstrass property of sequences or the Heine-Borel property of open coverings. No analogue of Cousin coverings can survive or prove as useful except on the real line. This tends to reduce use of the lemma since mathematicians would prefer to prepare their students for advanced ideas rather than seek for the clearest and simplest presentation of the material. Some mathematicians have claimed that the lemma is perhaps *too* successful in early courses: the students become so enamored in using the machinery of the lemma that they avoid learning any other techniques.

Calculus Curricula. For most of the 21st century, the standard calculus course remained in a frozen and fossilized state, ignoring not only these developments but virtually all of the mathematics of the 20th century. Calculus students were taught only the integration theories of Cauchy and Riemann. The measure theory naturally associated with the study (that theory developed by Peano, Jordan and Cantor) was carefully avoided. Lebesgue's measure theory and associated "integral" were available only in advanced courses, presented at a significant level of abstraction and with only slight attention to the calculus implications.

Behaving improperly. Here is a bizarre ritual from calculus courses of the last century, that never fails to amuse our students. A student of that era might be required to determine an integral such as

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

Our students would instantly recognize that the function

$$F(x) = 2\sqrt{x}$$

has $F'(x)$ equal to the integrand at every point with the sole exception $x = 0$, which immediately shows the integral exists with value $F(1) - F(0) = 2$. But in the past the calculus student was expected, first, to work on an interval $[t, 1]$ with $t > 0$, show that the integrand $f(x) = 1/\sqrt{x}$ is continuous there. That justified an application of Riemann's integral. Then he should notice that f is unbounded on $(0, 1]$, and would require "an improper integral"' so that the integral assumes the value

$$\lim_{t\searrow 0}[2 - 2\sqrt{t}] = 2.$$

No other method or fewer steps would have been considered acceptable.

Fear and confusion in graduate school. To compound the confusion, the student would later go on to study Lebesgue's integral. The Lebesgue integral does handle unbounded functions, but these manipulations, that were required to be performed as a freshman student, are no longer legitimate. Often unknown to the student was that some unbounded functions that could be handled by elementary calculus methods, could not be handled by Lebesgue's method.

Typically, students emerged from a course in measure theory and measure-theoretic integration theory unable to see an immediate connection with their calculus studies of the integral. A standard anecdote of the era describes graduate students frozen in fear on an oral exam that asked

"What is the value of the Lebesgue integral $\int_0^1 x^2\,dx$?"

Reform school. Efforts to reform were misguided. At the first appearance of the Henstock and Kurzweil formulation of the correct integral, many people, impressed with the natural simplicity of the concepts, viewed this as a way to change the calculus curriculum. They proposed a substitution of this integral for the Cauchy-Riemann theory. But the intention still remained to avoid all measure theory, and the result was a clumsy mess.

The real lesson of the new approach, often misinterpreted, was not that there was a simple way to develop the program of the calculus, but that now there was a *correct* way. The measure theory was part of the program. Measure theory can be developed in the constructive manner that Lebesgue chose or in dozens of other ways, but it is an essential feature of the program and cannot be diluted.

Gauged and tagged students The presentation of the theory often assumed a peculiar form, that doubtless also hindered its acceptance. While the integral and derivative are captured effectively by the notion of a covering relation, one can more awkwardly capture these notions by a gauge (or gage), i.e., a positive function δ. Partitions are tagged partitions, "tags" might be anchored.

All definitions were given in terms of the gauge, no matter how clumsy, rather than covering relations. This focus on the gauge, with frequent reference to the tags, which to us may seem quaint and oddly old-fashioned, made many formulations cumbersome and obscured the nature of the processes. Often curious questions were asked of the gauges themselves, perhaps from some mysterious and naive belief that it was the gauge that was the key to the ideas.

This is odd in retrospect. No other notion that was local in character and that could have allowed a "gauge" treatment ever received one. For example functions were continuous or differentiable if they were continuous or differentiable (i.e., locally) at each point. No notion of gauge-continuity or gauge-differentiability was employed and yet gauge-integrability was commonly used as a term.

Other aspects of the language of the times added to the unpopularity. One reads of gage integrals, δ-fine divisions, generalized Riemann integrals, tagged partitions, locally small Riemann sums, and a host of conditions defined by the "gages." "Tags" might appear at the endpoints, in the interior, or even outside the interval depending on the intention of the writer. It is little wonder that the general mathematical community considered the topic to be some backwater devoted to the arcana of Denjoy and Perron type integrals and belonging to the same obscure tradition.

By the way . . . here's measure theory. Some proposed that this integral could offer an easier introduction to measure theory itself, avoiding what they saw as the long and tedious process of developing Lebesgue's measure. Indeed it does; the variational techniques provide the measure theory in a natural formal way, although the constructive techniques need to be learned in any case.

But they came up with the dreadful suggestion that the integral be fully developed with no measure theory at all and then, at a later stage, Lebesgue's measure be introduced as

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

for those sets E that would allow this. Thus the measure theory, which is fundamental and essential to the entire theory, would be pushed to a corner and a fragment of it re-injected as an afterthought to the integration theory, and studied in an artificial way.

Post apocalypse. None of this, however, had any impact on the calculus curriculum. Textbooks of the late 1990's continued to be recycled into newer and larger editions, some of them surviving hardly unchanged into the 2090's. It was only after the cataclysmic events of that decade, destroying most of the old materials and most of the old teachers, that a more modern version was able to emerge as the mathematicians then still alive were able to rethink the fundamentals of their subject.

The early writings of a small group that published, anonymously, under the strange and never explained pseudonym "N. Bourbaki Jr," set out all the material that you now encounter in a calculus course. Not having access to the older material, they developed the subject without preconceptions in the current fashion. We recognize this as significantly better than the misguided teachers of an earlier time had done. But we should, nonetheless, fear a return to rigidity and inflexibility in our teaching of basic mathematics concepts.

CHAPTER 1

Real Numbers

1.1. Real numbers

The set of real numbers is denoted $\mathbb{R}$. It includes, as subsets, the set of natural numbers $\mathbb{N}$ and the set of rational numbers $\mathbb{Q}$.

> *We take for granted all of the elementary arithmetic and order properties of the real numbers.*

Mostly familiar notations are employed. We use, for example, $[a]^+ = \max\{a, 0\}$ and $[a]^- = \max\{-a, 0\}$. Then, evidently, $a = [a]^+ - [a]^-$ and $|a| = [a]^+ + [a]^-$.

1.2. Bounds

Definition 1.1. A set of real numbers S is *bounded above* if there is a real number M so that $s \leq M$ for all $s \in S$. Such a number M, if it exists, is called *an upper bound* for S. (If there is one upper bound, then there are many.)

Similarly S is *bounded below* if there is a real number m so that $s \geq m$ for all $s \in S$ and m would be called *a lower bound* for S. A set bounded both above and below is said to be *bounded*.

1.2.1. Sups and infs. Among all upper bounds for a set there is always a minimal one. Among all lower bounds for a set there is always a maximal one.

> *We take for granted that for any nonempty set S that has an upper bound there is a real number M denoted as $\sup S$ that is an upper bound for S that is not larger than any other upper bound.*

If we use the device $-S = \{s : -s \in S\}$ we can prove the corresponding statement for lower bounds:

Theorem 1.2. For any nonempty set S that has a lower bound there is a real number m denoted as $\inf S$ that is a lower bound for S that is not smaller than any other lower bound.

1.2.2. The symbols ∞ and $-\infty$. The supremum and infimum are extended to all sets by employing the two symbols ∞ and $-\infty$. We write $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$. If a set S has *no* upper bounds (i.e., if it is not bounded above) then we write $\sup S = \infty$. Similarly if S has *no* lower bounds (i.e., if it is not bounded below) then we write $\inf S = -\infty$.

The symbols ∞ and $-\infty$ are not included in $\mathbb{R}$, are not considered to be real numbers, and should not (except with considerable caution) enter into arithmetic statements.

They can be used in order assertions. Thus

$$-\infty < a < b < \infty$$

would mean only that a and b are real numbers (finite real numbers) and that $a < b$. Also

$$-\infty \le a < b \le \infty$$

would mean that a and b are real numbers or possibly $a = -\infty$ and/or $b = \infty$.

1.2.3. Exercises. The exercises illustrate the sup/inf principle. In each case we are required to claim the existence of a real number with a certain property. Algebraic and order methods alone (without an appeal to the sup/inf principle) would be unable to show such existence.

EXERCISE 1.3. Every nonempty subset S of $\mathbb{N}$ has a first element.
Hint: Take $c = \inf S$ and verify that c is a real number. Show c is a natural number.

EXERCISE 1.4. For any positive number ϵ there is a natural number n for which $1/n < \epsilon$.
Hint: If not, then $1/\epsilon$ is an upper bound for the set of all natural numbers. Get a contradiction by taking the supremum (which would apparently have to be a real number).

Between any two rational numbers a and b must be another rational number: simply take the average $(a + b)/2$. Algebraic methods alone do not, however, work if a and b are arbitrary real numbers.

EXERCISE 1.5. Between any two real numbers $a < b$ there is a rational number $a < r < b$.
Hint: Find a natural number n with $1/n < b - a$. Consider all the rational numbers m/n.

1.3. Sequences of real numbers

A function whose domain is the set $\mathbb{N}$ of natural numbers with range in some specified set is called a *sequence*. In the calculus we use sequences of real numbers, sequences of sets, and sequences of functions.

If s is a sequence, accordingly, then an object $s(n)$ must be prescribed for every natural number n. Since the natural numbers form a list

$$1, 2, 3, 4, 5, \dots, n, \dots$$

it is convenient and natural to think of the sequence s as being itself the list

$$s(1), s(2), s(3), \dots, s(n), \dots.$$

In practise subscript notation, replacing $s(n)$ by s_n, is preferred so the sequence becomes

$$s_1, s_2, s_3, \dots, s_n, \dots.$$

The element s_n is said to be the *nth term of the sequence* and, rather than referring to the sequence as s, we more often use $\{s_n\}$ to indicate the sequence we have in mind.

1.4. Subsequences

A subsequence $\{s_{n_k}\}$ of a sequence $\{s_n\}$ is obtained by choosing first a sequence of natural numbers

$$n_1 < n_2 < n_3 < \cdots < n_k < \dots$$

and defining then the sequence whose kth term is s_{n_k}:

$$s_{n_1}, s_{n_2}, s_{n_3}, \dots, s_{n_k}, \dots.$$

Any sequence so obtained from $\{s_n\}$ is said to be a *subsequence* of that sequence.

1.5. Monotonic sequences

A sequence $\{s_n\}$ of real numbers does not decrease if $s_n \leq s_{n+1}$ for all natural numbers n. We say it is *nondecreasing*. Similarly a sequence does not increase if $s_n \geq s_{n+1}$ for all natural numbers n. We say it is *nonincreasing*. Should a sequence be either nondecreasing or nonincreasing it is said to be *monotonic*.

THEOREM 1.6 (Existence of monotonic subsequences). Every sequence of real numbers has monotonic subsequences.

Proof. First we look for a nonincreasing subsequence, if possible. Let us "highlight" the mth element s_m of the sequence $\{s_n\}$ if all later elements are less than or equal to it, i.e., in symbols if $s_m \geq s_n$ for all $n > m$. If there is an infinite subsequence of highlighted points s_{m_1}, s_{m_2}, s_{m_3}, s_{m_4}, ... then we have found our nonincreasing subsequence since

$$s_{m_1} \geq s_{m_2} \geq s_{m_3} \geq s_{m_4} \geq \dots.$$

This would not be possible if there are only finitely many highlighted points. So let us suppose that s_M is the last highlighted point; then any element s_n for $n > M$ is not highlighted, meaning that there must be an element further on in the sequence greater than it. Thus we can choose $s_{m_1} > s_{M+1}$ with $m_1 > M + 1$, then $s_{m_2} > s_{m_1}$ with $m_2 > m_1$, and then $s_{m_3} > s_{m_2}$ with $m_3 > m_2$, and so on to obtain an increasing subsequence

$$s_{M+1} < s_{m_1} < s_{m_2} < s_{m_3} < s_{m_4} < \dots$$

as required, completing the proof. ∎

1.6. Limits of monotonic sequences

DEFINITION 1.7. If $\{s_n\}$ is a nondecreasing sequence then we shall write

$$\lim_{n\to\infty} s_n = \sup\{s_n : n = 1, 2, 3, \dots\}.$$

Similarly if $\{s_n\}$ is a nonincreasing sequence then we write

$$\lim_{n\to\infty} s_n = \inf\{s_n : n = 1, 2, 3, \dots\}.$$

Note that a nondecreasing sequence either has a finite limit or is assigned ∞ as its limit. In the former case we say it is *convergent*. Similarly a nonincreasing sequence is convergent if it has a finite limit, otherwise the limit is assigned as $-\infty$.

1.7. Limits of arbitrary sequences

While every monotonic sequence can be assigned a limit (possibly infinite) only certain nonmonotonic sequences allow this.

DEFINITION 1.8. For a general sequence $\{s_n\}$ we write $\lim_{n\to\infty} s_n = c$ provided that $\lim_k s_{n_k} = c$ for every monotonic subsequence $\{s_{n_k}\}$.

The expression $\lim_{n\to\infty} s_n$ in Definition 1.8 may assume the values ∞ or $-\infty$ in accordance with our rules for sups and infs. When this limit exists and is finite the sequence is said to *converge*; the sequence is said to *diverge to* ∞ or $-\infty$ when $\lim_{n\to\infty} s_n$ assumes those values. There are sequences which do not converge and for which $\lim_{n\to\infty} s_n$ does not exist, even allowing for infinite values. Such sequences are sometimes called *oscillatory*.

EXERCISE 1.9. Show that every convergent sequence is bounded above and below.

EXERCISE 1.10. Show that every sequence that is bounded above and below has at least one subsequence that converges.

1.8. Limit superior and inferior

For a monotonic sequence $\lim_{n\to\infty} s_n$ has a definite meaning. For a general sequence this may not be the case: it is possible for two different monotonic subsequences $\{s_{n_k}\}$ and $\{s_{m_p}\}$ to exist which have different limits. To handle these requires the following definition.

DEFINITION 1.11. For a general sequence $\{s_n\}$ we write

$$\limsup_{n\to\infty} s_n = \lim_{n\to\infty} \left(\sup\{s_n, s_{n+1}, s_{n+2}, \dots\}\right)$$

and

$$\liminf_{n\to\infty} s_n = \lim_{n\to\infty} \left(\inf\{s_n, s_{n+1}, s_{n+2}, \dots\}\right).$$

Notice that

$$t_n = \sup\{s_n, s_{n+1}, s_{n+2}, \dots\}$$

defines a monotonic sequence (obtained by taking sups for all values in a particular subsequence of $\{s_n\}$) and consequently $\lim_{n\to\infty} t_n$ has a meaning. This justifies the definition of the limit superior; the limit inferior has a similar justification.

EXERCISE 1.12. [Properties of sequence limits] Let $\{s_n\}$ and $\{t_n\}$ be real sequences. Show that

(a) $\limsup_{n\to\infty} cs_n = c\,(\limsup_{n\to\infty} s_n)$ if $c > 0$ and we agree to interpret $c\infty = \infty$ and $c(-\infty) = -\infty$.
(b) $\limsup_{n\to\infty} s_n = -\liminf_{n\to\infty}(-s_n)$.
(c) $\limsup_{n\to\infty}(s_n + t_n) \le \limsup_{n\to\infty} s_n + \limsup_{n\to\infty} t_n$ provided these are finite numbers.
(d) $\liminf_{n\to\infty}(s_n + t_n) \ge \liminf_{n\to\infty} s_n + \liminf_{n\to\infty} t_n$ provided these are finite numbers.
(e) $\limsup_{n\to\infty} s_n = \liminf_{n\to\infty} s_n$ if and only if $\lim_{n\to\infty} s_n$ exists in which case all values are the same.

1.9. Metric properties of sequence limits

The sequence limits can be characterized, in the case where the limit is a finite number, by metric properties of the sequence. These characterizations are due to Cauchy who was the first to give such metric notions and to apply them to the study of limits and integrals.

The term "metric" refers to the notion of distance. For real numbers c and d the distance between them is interpreted as $|d-c|$. In the two Cauchy criteria notice that the concept of sequence limit, originally defined by means of sups and infs, now has a characterization by an expression using metric ideas, namely the expressions

$$|s_n - c| \text{ and } |s_n - s_m|.$$

The two Cauchy criteria appear in numerous other settings and it is essential to learn the proofs here in order to apply some or all of the ideas later on. In our presentation of material in the text here and in later chapters nearly all concepts are defined using sup/inf recipes; all of these have Cauchy type criteria expressed using the metric (distance) notions in $\mathbb{R}$. The proofs can be used as models for similar metric characterizations in other settings.

1.9.1. First Cauchy Criterion.

THEOREM 1.13 (Cauchy 1st Criterion). *Let $\{s_n\}$ be a sequence of real numbers. A necessary and sufficient condition in order that the limit $\lim_{n\to\infty} s_n = c$ exist for some finite number c is that for every $\epsilon > 0$ there is an integer m so that*

$$|s_n - c| < \epsilon \quad \text{for all } n \geq m.$$

Note that this criterion may also be written in the form

$$\lim_{m\to\infty} \sup_{n\geq m} |s_n - c| = 0.$$

Proof. Suppose that $\lim_{n\to\infty} s_n = c$ but that the condition does not hold. Then, for some ϵ there must be a subsequence $\{s_{n_k}\}$ of $\{s_n\}$ that has no terms in the interval $(c-\epsilon, c+\epsilon)$. Take a further monotone subsequence of $\{s_{n_k}\}$. Whatever that sequence has as its limit it is certainly not c. But this is a contradiction, hence the condition must hold.

For the converse direction observe that, if $|s_n - c| < \epsilon$ for all $n \geq m$, then

$$c - \epsilon < s_n < c + \epsilon.$$

Consider limits of monotone subsequences $\{s_{n_k}\}$. Such a sequence lies in the interval $(c-\epsilon, c+\epsilon)$ after $n_k \geq m$ and hence

$$c - \epsilon \leq \lim_k s_{n_k} \leq c + \epsilon.$$

Since this is true for all $\epsilon > 0$ we must conclude that $\lim_k s_{n_k} = c$ for every monotone subsequence. Thus, by definition, $\lim_{n\to\infty} s_n = c$. ■

1.9.2. Second Cauchy Criterion.

THEOREM 1.14 (Cauchy 2nd Criterion). *Let $\{s_n\}$ be a sequence of real numbers. A necessary and sufficient condition in order that the limit $\lim_{n\to\infty} s_n = c$ exist for some finite number c is that for every $\epsilon > 0$ there is an integer p so that*

$$|s_n - s_m| < \epsilon \quad \text{for all } n, m \geq p.$$

Note that this criterion may also be written in the form

$$\lim_{p} \sup_{n,m\geq p} |s_n - s_m| = 0.$$

Proof. We begin with the simpler proof of necessity and we use the first Cauchy criterion. Assuming that $\lim_{n\to\infty} s_n = c$, we can choose p so that

$$|s_n - c| < \epsilon/2$$

for all $n \geq p$. Hence, for all $n, m \geq p$,

$$|s_n - s_m| \leq |s_n - c| + |s_n - c| < \epsilon$$

as required.

To prove that the condition is sufficient let $\epsilon > 0$ and suppose that there is a p so that

$$|s_n - s_m| < \epsilon/2$$

for all $n, m \geq p$. Take any monotone subsequence $\{s_{n_k}\}$ and write $c = \lim_k s_{n_k}$. The inequality above shows that

$$s_n - \epsilon/2 \leq c \leq s_n + \epsilon/2$$

for all $n \geq p$. If we rewrite this in the usual form

$$|s_n - c| \leq \epsilon/2 < \epsilon$$

we recognize the first Cauchy criterion and this means that $\lim_{n\to\infty} s_n = c$ as required. ∎

1.10. Series

A special notation is used when the sequence is formed by summing the elements of a sequence $\{a_k\}$:

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=1}^{n} a_k.$$

The meaning assigned here depends on the limit having a meaning (possibly $\pm\infty$). If the terms are nonnegative then $\sum_{k=1}^{n} a_k$ forms a nondecreasing sequence as n increases. Thus such series always have a sum.

Definition 1.15. We use the following language of convergence:

- The series $\sum_{k=1}^{\infty} a_k$ is said to be *convergent* if the limit $\lim_{n\to\infty} \sum_{k=1}^{n} a_k$ exists (as a finite number).
- The series $\sum_{k=1}^{\infty} a_k$ is said to be *absolutely convergent* if it is convergent and $\sum_{k=1}^{\infty} |a_k| < \infty$.
- The series $\sum_{k=1}^{\infty} a_k$ is said to be *nonabsolutely convergent* if it is convergent and $\sum_{k=1}^{\infty} |a_k| = \infty$.

Exercise 1.16. Show that $\sum_{i=1}^{\infty} 2^{-i} = 1$.

1.10.1. Positive and negative parts of a series. By splitting the series terms into positive and negative parts, using the device

$$a_n = [a_n]^+ + [a_n]^-$$

we can often analyse a series and estimate its sum.

THEOREM 1.10.1. *If $\sum_{k=1}^{\infty} |a_k| < \infty$ then the series $\sum_{k=1}^{\infty} a_k$ is absolutely convergent and*

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} [a_k]^+ - \sum_{k=1}^{\infty} [a_k]^-.$$

Proof. This follows easily from the identity

$$\sum_{k=1}^{N} a_k = \sum_{k=1}^{N} [a_k]^+ - \sum_{k=1}^{N} [a_k]^-$$

and the convergence properties of sequences. ∎

THEOREM 1.10.2. *If $\sum_{k=1}^{\infty} |a_k| = \infty$ and the series $\sum_{k=1}^{\infty} a_k$ is convergent then it is necessarily nonabsolutely convergent and*

$$\sum_{k=1}^{\infty} [a_k]^+ = \sum_{k=1}^{\infty} [a_k]^- = \infty.$$

Proof. Suppose, for example, that

$$\sum_{k=1}^{\infty} [a_k]^+ < \infty.$$

Then

$$\sum_{k=1}^{\infty} [a_k]^- = \infty$$

and

$$\lim_{N\to\infty} \sum_{k=1}^{N} a_k = \lim_{N\to\infty} \left(\sum_{k=1}^{N} [a_k]^+ - \sum_{k=1}^{N} [a_k]^- \right) = -\infty$$

and the series cannot converge. ∎

1.10.2. Rearrangements. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijection, thus a rearrangement of the natural numbers into a different order. Then the sequences $\{a_n\}$ and $\{a_{\sigma(n)}\}$ have the same terms but in a different order. One might expect the sums to be the same.

THEOREM 1.10.3. *If $\sum_{k=1}^{\infty} |a_k| < \infty$ then the series*

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} a_{\sigma(k)}$$

for any rearrangement σ.

Proof. Under this hypothesis the series is absolutely convergent and can be treated by studying its positive and negative parts. Thus we may reduce the proof to handling just the case where $a_n \geq 0$ for all n.

Fix N. Then there must be an M sufficiently large that

$$\sum_{k=1}^{N} a_k \leq \sum_{k=1}^{M} a_{\sigma(k)}$$

simply by ensuring that all integers 1, 2, ..., N appear in the list

$$\{\sigma(1), \sigma(2), \ldots, \sigma(M)\}.$$

Consequently, for all N,

$$\sum_{k=1}^{N} a_k \leq \sum_{k=1}^{\infty} a_{\sigma(k)}$$

and so also

$$\sum_{k=1}^{\infty} a_k \leq \sum_{k=1}^{\infty} a_{\sigma(k)}.$$

This inequality is true for any two rearrangements of an absolutely convergent series and so the two sums must be in fact equal. ■

THEOREM 1.10.4. *If* $\sum_{k=1}^{\infty} a_k$ *is nonabsolutely convergent then there exists a rearrangement* σ *so that*

$$\sum_{k=1}^{\infty} a_k \neq \sum_{k=1}^{\infty} a_{\sigma(k)}$$

Proof. Since $\sum_{k=1}^{\infty} a_k$ is nonabsolutely convergent we can claim two facts that are needed in the construction. First, for the series of positive terms, that $\sum_{k=1}^{\infty} [a_n]^+ = \infty$. Second, that the sequence $\{a_n\}$ is bounded, say that $|a_n| < M$ for all n.

The strategy is simple: write successively all the nonnegative terms of the series (in the order in which they appear, for example) and then, rarely, insert the negative terms starting with the first of these. Specifically list nonnegative terms until the accumulated sum exceeds $2M$. Insert the first neglected negative term; the sum will be reduced but still exceeds $2M - M = M$. Continue with further nonnegative terms until the accumulated sum exceeds $3M$. Insert the next neglected negative term; the sum will be reduced but still exceeds $3M - M = 2M$. Continue this way inductively. The new series will be an rearrangement of the original series since all terms appear in some position. But the new series has sums that grow past any bound and so has a total sum ∞. ■

CHAPTER 2

Real Sets

Any analysis of the elements of the calculus will require attention to certain kinds of subsets of real numbers.

2.1. Displaying a real set as a sequence

If S is a set of real numbers it would be most convenient for many arguments to have all of its members displayed as the terms of a sequence. We would require that there be a sequence $\{s_n\}$ all of whose terms belong to S with the property that each $s \in S$ appears at least once in the sequence. A nonempty set with this property is said to be *countable*. (We also call the empty set $\emptyset$ countable.)

THEOREM 2.1 ($\mathbb{Q}$ is countable). There is a sequence consisting of rational numbers such that every rational number appears once and only once.

Proof. We need to write all numbers of the form m/n for m and n positive or negative integers and $n \neq 0$. Define $\|m/n\| = |m| + |n|$ and let

$$A_k = \{m/n : \|m/n\| = k\}$$

for $k = 1, 2, \dots$. Check that each A_k is finite. Check that every rational number appears in one of the sets A_k. Now construct the list by listing first all the members of A_1 (namely just the number $0/1$), then all the members of A_2 (namely the numbers $1/1$, $-1/1$), and so on. This list contains all rational numbers, but with repetition (e.g., $2/2$ appears in A_4 but has already appeared in A_2 as $1/1$). Simply delete any repetitions, starting at the beginning of the list. ■

In contrast to this, the set of real numbers itself is not countable. (See Exercise 2.4.)

EXERCISE 2.2. Let E_1, E_2, E_3, ... be a sequence of countable sets. Show that the union $\bigcup_{n=1}^{\infty} E_n$ is also countable.

2.2. Open intervals

DEFINITION 2.3 (Open interval). For any $-\infty \leq a < b \leq \infty$ the set

$$(a, b) = \{x \in \mathbb{R} : a < x < b\}$$

is said to be an *open interval*.

Open intervals may be bounded if a and b are finite; in that case we call a and b the *endpoints* of the interval. Open intervals may be unbounded, as for example $(-\infty, \infty)$, $(-\infty, b)$ and (a, ∞).

EXERCISE 2.4. [No open interval is countable] For any interval (a, b) and any real sequence $\{s_n\}$ there is an element $c \in (a, b)$ that does not appear in the sequence.

Hint: Choose $a < a_1 < b_1 < b$ so that s_1 does not lie between a_1 and b_1. Then choose $a_1 < a_2 < b_2 < b_1$ so that s_2 does not lie between a_2 and b_2. Continue in this way. The limiting property of sequences will provide you with a number c with this property.

EXERCISE 2.5. [Listing of open intervals] There is a sequence

$$(a_1, b_1),\ (a_2, b_2),\ (a_3, b_3), \ldots$$

so that every open interval with rational endpoints appears in the list.

2.2.1. Length of an open interval.

DEFINITION 2.6 (Length of an open interval). For any $-\infty \leq a < b \leq \infty$

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

is called the *length* or *measure* of the open interval (a, b).

Note that unbounded intervals have infinite length since we would interpret each of $\infty - (-\infty)$, $\infty - b$ and $a - (-\infty)$ as ∞.

2.3. Dense sets

A set is dense if it appears in every interval.

DEFINITION 2.7. A set E is *dense* if $E \cap (a, b) \neq \emptyset$ for every open interval (a, b).

DEFINITION 2.8. Let E and E_1 be real sets. We say that E_1 is *dense in E* provided that whenever (a, b) is an open interval for which $E \cap (a, b) \neq \emptyset$ then necessarily $E_1 \cap (a, b) \neq \emptyset$.

EXERCISE 2.9. Show that the set of all rationals and the set of all irrationals are dense.

2.4. Open sets

DEFINITION 2.10 (Open set). If a nonempty set G is the union of a finite or infinite sequence of open intervals

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

in such a way that

$$(a_i, b_i) \cap (a_j, b_j) = \emptyset \quad i \neq j$$

then G is said to be an *open set* and the intervals displayed are said to be its *component intervals*. [The empty set is considered to be an open set with *no* component intervals.]

2.4.1. Length of an open set. The measure of an open set is the sum of the measure of its component intervals.

DEFINITION 2.11 (Open set). Let G be an open set. Then the length or measure $\mathcal{L}(G)$ is defined as the sum of the lengths of the component intervals of G.

For the empty set we interpret this definition as yielding $\mathcal{L}(\emptyset) = 0$. If G has an infinite sequence of component intervals (a_i, b_i), $i = 1, 2, 3, ...$ then

$$\mathcal{L}(G) = \sum_{i=1}^{\infty}(b_i - a_i).$$

If there are only finitely many component intervals the series would be replaced by a finite sum.

If G has an unbounded component then certainly $\mathcal{L}(G) = \infty$. If G is bounded then $\mathcal{L}(G)$ is finite. If G is unbounded, then $\mathcal{L}(G)$ can be finite or infinite depending on the sum of the series.

2.5. Properties of open sets

The definition makes it clear what the structure of an open set is. Often the set we wish to study is not presented to us in a way that reveals the structure. How then might we verify that it is open? A point in an open set must belong to some component interval. That feature provides a local property that characterizes open sets.

LEMMA 2.12. A set G is open if and only if for every $x \in G$ there is an open interval (c, d) containing that point and itself entirely contained in G.

Proof. The condition is easily seen to be necessary. Suppose that, for every $x \in G$, we can choose an open interval I_x containing x and contained in G. Choose I_x to be maximal, i.e., as large as possible. This will require a step to show that there is a maximal interval: use the sup/inf property to define $I_x = (c, d)$ where

$$c = \inf\{s : s < x \text{ and } (s, x) \subset G\}$$

and

$$c = \sup\{t : x < t \text{ and } (x, t) \subset G\}.$$

The family $\{I_x : x \in G\}$ is the family of components of G. [Note that I_x would be the same as I_y for many different x and y.]

If the family is empty then $G = \emptyset$ and we have nothing to prove. If the family is nonempty we arrange its members into a sequence: first list all rational numbers $\{r_n\}$ then select $I_{r_{n_1}} = J_1$ as the first r_{n_1} we find in the list for which $r_{n_1} \in G$. Then select $I_{r_{n_2}} = J_2$ as the first r_{n_2} we find in the list for which $r_{n_1} \in G$ but $r_{n_2} \notin J_1$. Continue inductively and check that this process exhausts the family and so produces a sequence $\{J_i\}$ of component intervals of G. Thus G is open. ■

EXERCISE 2.13. Let G be an open set. Show that there is a sequence of open intervals with rational endpoints

$$(a_1, b_1),\ (a_2, b_2),\ (a_3, b_3), \ldots$$

so that

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

Hint: [This is not the list of components, nor will the intervals be disjoint.] Use Exercise 2.5 to produce a listing of *all* open intervals with rational endpoints (a_1, b_1), (a_2, b_3),(a_3, b_3), ... Now remove from the list any interval that is not contained in G. Check that if $x \in G$ then $x \in (a_i, b_i) \subset G$ for at least one member in the sequence.

2.5.1. Further properties of open sets.

EXERCISE 2.14. Suppose that G_1, G_2, G_3, ... is a sequence of open sets. Show that $\bigcap_{i=1}^n G_i$ is open (for all n) but that $\bigcap_{i=1}^\infty G_i$ need not be.

EXERCISE 2.15. Suppose that $\mathcal{G}$ is a family of open sets. Show that $\bigcup_{G\in\mathcal{G}} G$ is open.

EXERCISE 2.16. Let $\{x_n\}$ be a sequence of real numbers converging to a point x that belongs to an open set G. Show that there is some integer N so that $x_n \in G$ for all $n \geq N$.

2.5.2. Measure properties of open sets.

THEOREM 2.17 (Additivity). Let G_1 and G_2 be open sets with $G_1 \cap G_2 = \emptyset$. Then $G_1 \cup G_2$ is an open set for which

$$\mathcal{L}(G_1 \cup G_2) = \mathcal{L}(G_1) + \mathcal{L}(G_2).$$

Proof. If the two open sets G_1 and G_2 are disjoint then it is easy to describe the component intervals of the open set $G_1 \cup G_2$: these are the open intervals that are components of either G_1 or of G_2. Consequently the sum formula is obvious. ■

THEOREM 2.5.1 (Monotone property). *If G_1 and G_2 are open sets with $G_1 \subset G_2$ then $\mathcal{L}(G_1) \leq \mathcal{L}(G_2)$.*

Proof. Suppose that $G_2 = (a, b)$. Then all components of G_1 are contained in the interval (a, b). Thus, if

$$(a_1, b_1),\ (a_2, b_2),\ (a_3, b_3), \dots$$

is the sequence of components of G_1, then certainly

$$\sum_{i=1}^N (b_i - a_i) \leq (b - a).$$

From that it follows that

$$\mathcal{L}(G_1) = \sum_{i=1}^\infty (b_i - a_i) \leq (b - a) = \mathcal{L}(G_2).$$

If G_2 has more than one component then, the same argument applied to a finite number of its components will show that

$$\sum_{i=1}^N (b_i - a_i) \leq \mathcal{L}(G_2).$$

Once again, then,

$$\mathcal{L}(G_1) = \sum_{i=1}^\infty (b_i - a_i) \leq \mathcal{L}(G_2).$$

■

2.6. Measure of arbitrary sets

The measure theory extends from open sets to arbitrary sets by using open sets to approximate the measure.

DEFINITION 2.18. For any set $E \subset \mathbb{R}$ we define

$$\mathcal{L}(E) = \inf\{\mathcal{L}(G) : E \subset G \text{ and } G \text{ open}\}.$$

2.6.1. Measure properties of arbitrary sets. The properties of the measure for open sets transfer naturally to the measure of arbitrary sets.

THEOREM 2.6.1 (Monotone property). *If E_1 and E_2 are arbitrary real sets with $E_1 \subset E_2$ then $\mathcal{L}(E_1) \leq \mathcal{L}(E_2)$.*

Proof. Let G be any open set containing E_2. Then G contains E_1 as well so that

$$\mathcal{L}(E_1) \leq \mathcal{L}(G).$$

As this holds for all such G it follows that $\mathcal{L}(E_1) \leq \mathcal{L}(E_2)$. ■

2.6.2. Additive properties. Two sets E_1 and E_2 are said to be *separated by open sets* if there are open sets G_1 and G_2 with $E_1 \subset G_1$, $E_2 \subset G_2$, and $G_1 \cap G_2 = \emptyset$. Such sets would certainly have $E_1 \cap E_2 = \emptyset$ but this condition is not enough to ensure that E_1 and E_2 are separated by open sets.

THEOREM 2.6.2 (Additive property). *If E_1 and E_2 are separated by open sets then*

$$\mathcal{L}(E_1 \cup E_2) = \mathcal{L}(E_1) + \mathcal{L}(E_2).$$

Proof. We can assume that $\mathcal{L}(E_1) + \mathcal{L}(E_2) < \infty$. Let $\epsilon > 0$. Choose an open set G so that $E_1 \cup E_2 \subset G$ and

$$\mathcal{L}(G) < \mathcal{L}(E_1 \cup E_2) + \epsilon.$$

If G_1 and G_2 are disjoint open sets separating these two sets then

$$\mathcal{L}(E_1)+\mathcal{L}(E_2) \leq \mathcal{L}(G\cap G_1)+\mathcal{L}(G\cap G_2) = \mathcal{L}(G\cap(G_1\cap G_2)) \leq \mathcal{L}(G) < \mathcal{L}(E_1\cup E_2)+\epsilon.$$

Hence, since ϵ is arbitrary,

$$\mathcal{L}(E_1) + \mathcal{L}(E_2) \leq \mathcal{L}(E_1 \cup E_2) \leq \mathcal{L}(E_1) + \mathcal{L}(E_2).$$

■

2.7. Closed sets

DEFINITION 2.19 (Closed set). A set E is said to be a *closed set* provided that its complement $G = \mathbb{R} \setminus E$ is open. The component intervals of G are said to be the *complementary intervals* of E.

EXERCISE 2.20. Suppose that E_1, E_2, E_3, ... is a sequence of closed sets. Show that $\bigcup_{i=1}^n E_i$ is closed (for all n) but that $\bigcup_{i=1}^\infty E_i$ need not be.

Hint: Verify that

$$\mathbb{R} \setminus \bigcup_{i=1}^n E_i = \bigcap_{i=1}^n (\mathbb{R} \setminus E_i)$$

and then use properties of open sets.

EXERCISE 2.21. Suppose that E_1 and E_2 are disjoint closed sets. Show that E_1 and E_2 are separated by open sets.

Hint: For any $x_1 \in E_1$, x_1 must be in a component interval of $\mathbb{R} \setminus E_2$. Thus there is $\delta(x_1) > 0$ so that $(x_1 - \delta(x_1), x_1 + \delta(x_1))$ contains no points of E_2. Similarly for $x_2 \in E_2$, there is $\delta(x_2) > 0$ so that $(x_2 - \delta(x_2), x_2 + \delta(x_2))$ contains no points of E_1. Write

$$G_1 = \bigcup_{x_1 \in E_1} (x_1 - \delta(x_1)/2, x_1 + \delta(x_1)/2)$$

and

$$G_2 = \bigcup_{x_2 \in E_2} (x_2 - \delta(x_2)/2, x_2 + \delta(x_2)/2)$$

If G_1 and G_2 have a point z in common, then there must be points $x_1 \in E_1$ and $x_2 \in E_2$ for which

$$(x_1 - \delta(x_1)/2, x_1 + \delta(x_1)/2) \cap (x_2 - \delta(x_2)/2, x_2 + \delta(x_2)/2)$$

contains z. Why is this impossible?

EXERCISE 2.22. Let E be a closed set and $\{x_n\}$ a convergent sequence of points in E. Show that $x = \lim_{n\to\infty} x_n$ must also belong to E.

Hint: If x does not belong to E then it belongs to a component interval (a, b) of $\mathbb{R} \setminus E$ that contains no points of E.

EXERCISE 2.23. Suppose that a set E has the property that whenever $\{x_n\}$ is a convergent sequence of points in E then $x = \lim_{n\to\infty} x_n$ must also belong to E. Show that E must be closed.

Hint: If $E' = \mathbb{R} \setminus E$ is not open then some point x in E' does not belong to a component interval, i.e., there must be an increasing sequence or decreasing sequence of points of E that converge to it. Note that this means that Exercise 2.22 characterizes closed sets.

EXERCISE 2.24. Show that every closed set E can be written as $E = \bigcap_{m=1}^{\infty} G_m$ for some sequence $\{G_m\}$ of open sets.

Hint: Suppose, that $\{(a_n, b_n)$ is the sequence of complementary intervals to E. If a_n and b_n are both finite, write

$$I_{nm} = \left[a_n + \frac{b_n - a_n}{4m}, b_n - \frac{b_n - a_n}{4m}\right].$$

If $a_n = -\infty$ or $b_n = \infty$ write instead

$$I_{nm} = \left(-\infty, b_n - \frac{1}{m}\right] \quad \text{or} \quad I_{nm} = \left[a_n + \frac{1}{m}, \infty\right)$$

Then use

$$E_m = \bigcup_{n=1}^{m} I_{nm}, \quad \text{and} \quad G_m = \mathbb{R} \setminus E_m.$$

2.8. Compact sets

DEFINITION 2.25 (Compact set). A set K is said to be *compact* if it is both closed and bounded.

Every nonempty compact set K has endpoints, i.e., a minimum point and a maximum point. We can define these as

$$a = \inf K \quad \text{and} \quad b = \sup K.$$

Since K is closed the set $G = \mathbb{R} \setminus K$ is open. Since K is bounded G must have unbounded component intervals of the form $(-\infty, a)$ and (b, ∞). Thus a and b are the minimum and maximum points in K.

EXERCISE 2.26 (Additivity). Let K_1 and K_2 be compact sets with $K_1 \cap K_2 = \emptyset$. Show that $K_1 \cup K_2$ is a compact set for which

$$\mathcal{L}(K_1 \cup K_2) = \mathcal{L}(K_1) + \mathcal{L}(K_2).$$

2.9. Compact interval

DEFINITION 2.27 (Compact interval). If a and b are real numbers for which $a < b$ we write

$$[a, b] = \{x : a \le x \le b\}$$

and refer to this as a *compact interval*.

The numbers a and b are called the endpoints of the interval $[a, b]$. The complement of a compact interval is easy to determine since

$$\mathbb{R} \setminus [a, b] = (-\infty, a) \cup (b, \infty).$$

This illustrates that $[a, b]$ is closed and that these two open intervals are complementary to $[a, b]$. Since $[a, b]$ is also bounded ($b = \sup[a, b]$ is an upper bound and $a = \inf[a, b]$ is a lower bound) this shows that it is also compact, hence the name.

EXERCISE 2.28 (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 $\lim_{n\to\infty}(b_n - a_n) = 0$ then there is a unique point z that belongs to each of the intervals.
Hint: Use a sequential proof by defining $z = \lim_{n\to\infty} a_n$. Check that z belongs to each interval, and that no other point can. (cf. Theorem 3.25 for a different proof using covering arguments.)

2.9.1. Length of a compact interval.

THEOREM 2.9.1 (Length of a compact interval). *For any* $-\infty < a < b < \infty$

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

is the length of the compact interval $[a, b]$.

Proof. If G is any open set with $G \supset [a, b]$ then $[a, b]$ is in one of the components of G with length exceeding $b - a$, so $\mathcal{L}(G) \ge b - a$. Thus $\mathcal{L}([a, b]) \ge b - a$. In the other direction take $G = (a - t, b + t)$ for $t > 0$ and note that $[a, b] \subset G$, G is open, and that $\mathcal{L}(G) = (b - a) + 2t$. Thus for all $t > 0$,

$$b - a \le \mathcal{L}([a, b]) < b - a + 2t.$$

■

EXERCISE 2.29. Let K be a compact set with endpoints a and b. Show that $G = [a, b] \setminus K$ is open, that $[a, b] = G \cup K$, and that

$$\mathcal{L}([a, b]) = \mathcal{L}(G) + \mathcal{L}(K).$$

EXERCISE 2.30. Show that the measure of open sets may be approximated from inside by compact sets, i.e., show that for every open set G,

$$\mathcal{L}(G) = \sup\{\mathcal{L}(K) : K \text{ compact and } K \subset G\}.$$

2.9.2. Unions of compact intervals. The union of a finite collection of compact intervals will be a compact set. We need to address also arbitrary unions.

LEMMA 2.31 (Unions of compact intervals). Let $\mathcal{C}$ be any collection of compact intervals. Then the set

$$C = \bigcup\{I : I \in \mathcal{C}\}$$

is the union of an open set and a countable set.

Proof. Let

$$G = \bigcup\{(c,d) : [c,d] \in \mathcal{C}\}$$

and

$$E = \{x : \ x \notin G \text{ and } x = c \text{ or } x = d \text{ for at least one } [c,d] \in \mathcal{C}\ \}.$$

Check that G is open, E is countable, and $C = G \cup E$. To show that E is countable write, for $n = 1, 2, 3, \dots,$

$$E_n = \{x : \ x \notin G \text{ and } x = c \text{ for at least one } [c,d] \in \mathcal{C} \text{ with } d - c > 1/n\}.$$

$$E_n' = \{x : \ x \notin G \text{ and } x = d \text{ for at least one } [c,d] \in \mathcal{C} \text{ with } d - c > 1/n\}.$$

We easily show that that each E_n and E_n' is countable. [How many points are there in E_n within $1/n$ of each other?] It follows that the set $E = \bigcup_{n=1}^{\infty}(E_n \cup E_n')$ is countable. ∎

CHAPTER 3

The Fundamental Covering Lemma of the Calculus

The notion of a covering relation and the elementary covering lemma of Cousin are central to all of the theoretical development of the calculus. It is the single concept which ties closely together, the derivative, the integral, and the measure theory.

3.1. Covering Lemmas

A *covering relation* is a family of interval-point pairs

$$([a, b], c)$$

where $[a, b]$ is a compact interval and c is a point in $[a, b]$. A *covering lemma* is a statement that from some covering relation β a subset β_1 can be extracted with certain desired properties.

3.2. Full covers and Cousin covers

A cover is said to be full at a point x_0 if it contains all pairs $([u, v], x_0)$ with the interval $[u, v]$ small enough. Here is the formal definition:

Definition 3.1. A covering relation β is said to be *full at a point* x_0 if there is a $\delta > 0$ so that β contains all pairs $([u, v], x_0)$ for which

$$u \leq x_0 \leq v \text{ and } \ 0 < v - u < \delta.$$

Definition 3.2. A covering relation β is said to be a *full cover of a set* E provided that β is full at each point of E.

Definition 3.3. A covering relation β is said to be simply a *full cover* provided that β is full at each point of the real line.

Exercise 3.4. Show that if β_1 is a full cover and $\beta_2 \supset \beta_1$ then β_2 is also a full cover.

Exercise 3.5. Show that if β_1 and β_2 are both full covers then so too is $\beta_1 \cap \beta_2$.

3.2.1. Cousin covers. In applications of covering ideas to compact intervals or compact sets the "full" requirement just introduced can be relaxed. For example if a function $F : [a, b] \to \mathbb{R}$ is being studied then intervals $[u, v]$ that are not subintervals of $[a, b]$ would play no role. This slight adjustment we name after Pierre Cousin who appears to be the first person to prove the simple covering lemma that is the key to all of the theory.

Definition 3.6. Let K be a compact set with endpoints a and b. A covering relation β is said to be a *Cousin cover* of K provided for every $x \in K$ there is a $\delta > 0$ so that $([c, d], x) \in \beta$ whenever $x \in [c, d] \subset [a, b]$ and $d - c < \delta$. (Note that the choice of δ is allowed to depend on the point x.)

EXERCISE 3.7. If β is a full cover of a compact set K show that β is also a Cousin cover of K. How might the converse fail?

EXERCISE 3.8. Show that if β is a Cousin cover of K and K_1 is a compact subset of K then β is also a Cousin cover of K_1.

EXERCISE 3.9. Show that if β_1 is a Cousin cover of K and $\beta_2 \supset \beta_1$ then β_2 is also a Cousin cover of K.

EXERCISE 3.10. Show that if β_1 and β_2 are both Cousin covers of K then so too is $\beta_1 \cap \beta_2$.

3.2.2. How full covers and Cousin covers arise. The exercises illustrate ways in which a property of sets or functions can give rise to Cousin covers.

EXERCISE 3.11. We say that $f : \mathbb{R} \to \mathbb{R}$ is locally constant at x if there is a $\delta > 0$ so that $f(x) = f(y)$ for all $|y - x| < \delta$. Show that, if f is locally constant at every point then

$$\beta = \{([c,d], x) : f(c) = f(d),\ c \le x \le d\}$$

is a full cover.

EXERCISE 3.12. Let $\mathcal{G}$ be a family of open sets so that every point in a compact set K belongs to some member of $\mathcal{G}$. Show that

$$\beta = \{([c,d], x) : c \le x \le d \ \text{ and } [c,d] \subset G \text{ for some } G \in \mathcal{G}\ \}$$

is a Cousin cover of K.

Many of the concepts of the calculus can be described and handled by covering relations. Students who recall the definition of continuous function should attempt this exercise.

EXERCISE 3.13. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous at every point in a compact interval $[a, b]$. Let $\epsilon > 0$. Show that

$$\beta = \{([c,d], x) : c \le x \le d \text{ and } |f(s) - f(t)| < \epsilon \text{ for all } s,\, t \in [c,d]\}$$

is a Cousin cover of $[a, b]$.

Students who recall the definition of the derivative should attempt the following:

EXERCISE 3.14. Suppose that $f : \mathbb{R} \to \mathbb{R}$ has a zero derivative $f'(x)$ at every point in a set E. Let $\epsilon > 0$. Show that

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

is a full cover of E.

3.2.3. Prunings. In working with covering relations it is useful to have a language to express a way in which these relations can be tailored to suit some analysis.

Let β be a covering relation and $E \subset \mathbb{R}$. We write

$$\beta(E) = \{(I, x) \in \beta : I \subset E\}$$

and

$$\beta[E] = \{(I, x) \in \beta : x \in E\}.$$

These subsets of β are called *prunings* of the original covering relation.

EXERCISE 3.15. If β is a Cousin cover of a compact set K show that $\beta[K]$ is also a Cousin cover of K.

EXERCISE 3.16. If β is a Cousin cover of a compact set K and G is any open set containing K, show that $\beta(G)$ is also a Cousin cover of K.

3.3. Partitions and subpartitions

Our main covering theorem will assert that a Cousin cover contains partitions or subpartitions with strong properties.

Suppose that we are able to construct a subdivision of $[a, b]$,

$$a = a_0 < a_1 < a_2 < \cdots < a_{k-1} < a_k = b$$

and to select points x_1, x_2, ..., x_k so that each pair

$$([a_{i-1}, a_i], x_i) \quad (i = 1, 2, \dots, k)$$

plays an important role in some problem. Then the collection

$$\pi = \{([a_{i-1}, a_i], x_i) : i = 1, 2, \dots, k\}$$

is a covering relation that is called a *partition* of $[a, b]$. Any subset of a partition is called a *subpartition*.

3.4. Cousin covering lemma

LEMMA 3.17 (Cousin covering lemma). 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. 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 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 (Exercise 2.4) 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 $\mathcal{L}([a_n, b_n]) < \delta$. 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 $[a, b]$, and indeed a partition of any compact subinterval. ■

3.4.1. Cousin covering lemma for compact sets.

For compact sets that are not intervals a version of the Cousin lemma is available, replacing partitions by subpartitions.

LEMMA 3.18 (Cousin covering lemma). Let β be a Cousin cover of a compact set K with endpoints a and b. Then β contains a subpartition π with the property that

$$K \subset \bigcup_{(I,x)\in\pi} I \subset [a,b].$$

Proof. Let

$$\beta_1 = \{(I,x) : \; x \in I, \, I \text{ a compact subinterval of } [a,b], \text{ and } I \subset [a,b] \setminus K\}.$$

Then check that $\beta \cup \beta_1$ is a Cousin cover of $[a,b]$. By the previous lemma there is a partition π' of $[a,b]$ contained in $\beta \cup \beta_1$. To produce the subpartition π merely remove from π' any elements that do not belong to β. ■

3.4.2. Some illustrations of the Cousin covering lemma. Facility with the Cousin lemma is essential to an understanding of a calculus course. For each exercise construct a covering relation that reflects the geometry of the problem. Then verify that this is a Cousin cover, extract a partition of the interval from that cover and use it to solve the exercise.

EXERCISE 3.19. [Sup/inf property] The course material began with the assumption of the sup/inf principle (Section 1.1), that a bounded, nonempty set has a sup and an inf. Show that this principle could have been proved if we had assumed the covering lemma as our starting point.

Hint: If $\emptyset \neq S \subset [a,b]$ and we assume that there is no minimal upper bound for S then let

$$\beta = \{([c,d],x) : \; c \text{ is an upper bound of } S \text{ or } d \text{ is not an upper bound of } S\}.$$

Check that this is a Cousin cover of $[a,b]$. Get a contradiction from the promised partition of $[a,b]$.

EXERCISE 3.20. Show that a function that is locally constant at each point of the real line is constant.

Hint: Use Exercise 3.11.

EXERCISE 3.21. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous at every point in a compact interval $[a,b]$. Show that f is bounded on $[a,b]$.

Hint: Use Exercise 3.13.

EXERCISE 3.22. Suppose that $f : \mathbb{R} \to \mathbb{R}$ has a zero derivative $f'(x)$ at every point in a compact interval $[a,b]$. Show that f is constant on $[a,b]$.

Hint: Use Exercise 3.14.

3.5. Compactness arguments

Deeper properties of compact sets require what are known as compactness arguments. There are a number of equivalent compactness arguments on the real line, among which the Cousin covering lemma is the most prominent[1].

[1]For this course. . . . A.S.

3.5.1. Heine-Borel theorem.

THEOREM 3.23. [Heine-Borel] Let $\mathcal{G}$ be a family of open sets so that every point in a compact set K is contained in at least one member of the family. Then there is a finite collection $\mathcal{G}_0 \subset \mathcal{G}$ with the same property.

Proof. (cf. Exercise 3.12.) The geometry of the situation is expressed in the covering relation

$$\beta = \{(I, x) : x \in I \text{ and } I \subset G \text{ for some } G \in \mathcal{G}\}.$$

Just check that β is a Cousin cover of K and apply the covering lemma to obtain a subpartition from β that has the appropriate properties. ■

3.5.2. Bolzano-Weierstrass theorem.

THEOREM 3.24. [Bolzano-Weierstrass] Let E be an infinite bounded set. Then E must have a point of accumulation[2].

Proof. This situation is expressed in the covering relation

$$\beta = \{(I, x) : x \in I \text{ and } I \cap E \text{ is finite}\}.$$

Suppose, contrary to the theorem, that E has no point of accumulation then β must be a Cousin cover of any compact interval $[a, b]$ that contains E. It would follow that E cannot be infinite. ■

3.5.3. Nested interval property.

THEOREM 3.25. [Nested interval property] Let

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n] \ldots$$

be a sequence of compact intervals, nested in the sense of containment as indicated. Then there is a point c of the interval $[a, b]$ that is contained in all of them.

Proof. We have proved this earlier (Exercise 2.4) using a sequential limit argument. Indeed our proof of the Cousin covering lemma itself used the nested interval property. Thus the proof we now give, obtaining it from the Cousin lemma, shows that they are equivalent.

If the statement in the theorem is not true, then every point $z \in [a, b]$ fails to belong to at least one of the intervals $[a_k, b_k]$. Take the covering relation

$$\beta = \{([c, d], x) : \ [c, d] \subset [a, b] \text{ and } [c, d] \cap [a_k, b_k] = \emptyset \text{ for at least one } k\}.$$

Check that this is a Cousin cover of $[a, b]$. A contradiction follows from the promised partition of $[a, b]$. ■

EXERCISE 3.26. Suppose that

$$(a, b) \supset (a_1, b_1) \supset (a_2, b_2) \supset \cdots \supset (a_n, b_n) \ldots$$

is a sequence of open intervals, nested in the sense of containment as indicated. Then must there be a point that is contained in all of them?

---

[2]A point of accumulation of a set E is a point $x \in \mathbb{R}$ so that every interval $(x - t, x + t)$ for $t > 0$ contains infinitely many points of E.

3.6. Covering arguments and measure theory

Compactness arguments allow us now to prove some deeper properties of the measure.

3.6.1. The measure of compact sets. The measure of a set can be estimated directly from Cousin covers. A very general version of this will appear in later chapters, but for now we illustrate that the measure $\mathcal{L}(K)$ of any compact set K is captured by this technique. This is, perhaps, our first indication of the deep role covering relations will later have in the measure theory.

THEOREM 3.6.1. *Let K be a compact set. Then*

$$\mathcal{L}(K) = \inf_{\beta} \sup \left\{ \sum_{(I,x)\in\pi} \mathcal{L}(I) : \pi \subset \beta \right\}$$

where the infimum is taken with regard to all Cousin covers β of K and the supremum is with regard to all subpartitions π contained in β.

Proof. Let $\epsilon > 0$. Choose an open set $G \supset K$ for which $\mathcal{L}(G) < \mathcal{L}(K) + \epsilon$. Note that

$$\beta = \{(I, x) : x \in K,\ I \subset G\}$$

is a Cousin cover of K and that, for every subpartition $\pi \subset \beta$,

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) \leq \mathcal{L}(G) < \mathcal{L}(K) + \epsilon.$$

This proves that

$$\mathcal{L}(K) \geq \inf_{\beta} \sup \left\{ \sum_{(I,x)\in\pi} \mathcal{L}(I) : \pi \subset \beta \right\}.$$

In the other direction, for any Cousin cover β of K, the covering lemma assures us that there is at least one subpartition $\pi \subset \beta$ with the property that

$$K \subset \bigcup_{(I,x)\in\pi} I. \quad \text{and} \quad \mathcal{L}(K) \leq \sum_{(I,x)\in\pi} \mathcal{L}(I).$$

From this it follows that

$$\mathcal{L}(K) \leq \inf_{\beta} \sup \left\{ \sum_{(I,x)\in\pi} \mathcal{L}(I) : \pi \subset \beta \right\}.$$

■

3.6.2. Measures of sequences of sets. With a compactness argument we are able to obtain estimates of the measures for sequences of sets.

LEMMA 3.27 (Sequential subadditive property). Let G, G_1, G_2, G_3, ... be a sequence of open sets for which

$$G \subset \bigcup_{n=1}^{\infty} G_n.$$

Then

$$\mathcal{L}(G) \leq \sum_{n=1}^{\infty} \mathcal{L}(G_n).$$

Proof. Let $\{(a_i, b_i)\}$ be the component intervals of G, fix an integer N and consider the compact set

$$K = \bigcup_{k=1}^{N} [c_k, d_k] \subset \bigcup_{n=1}^{\infty} G_n$$

for choices $a_k < c_k < d_k < b_k$. The most natural covering relation describing our situation is

$$\beta = \{(I, x) : \ x \in I \text{ and } I \subset G_n \text{ for some } n\}.$$

Check that β is a Cousin cover of the compact set K. Thus we can extract a subpartition π from β for which

$$K \subset \bigcup_{(I,x)\in\pi} I.$$

Collect the intervals I with $(I, x) \in \pi$ that are subsets of a particular set G_n and observe that the total length of these intervals I cannot exceed $\mathcal{L}(G_n)$. There are only a finite number of pairs (I, x) in π to handle and so we see that

$$\sum_{k=1}^{N} (d_k - c_k) \leq \sum_{(I,x)\in\pi} \mathcal{L}(I) \leq \sum_{n=1}^{\infty} \mathcal{L}(G_n).$$

As this would be true for all such choices of $[c_k, d_k] \subset (a_k, b_k)$ we can conclude that

$$\sum_{k=1}^{N} (b_k - a_k) \leq \sum_{n=1}^{\infty} \mathcal{L}(G_n).$$

Consequently

$$\mathcal{L}(G) = \sum_{k=1}^{\infty} (b_k - a_k) \leq \sum_{n=1}^{\infty} \mathcal{L}(G_n).$$

proving the lemma. ■

Corollary 3.28 (Additivity over sequences). Suppose that G_1, G_2, G_3, ... is a disjointed sequence of open sets. Then

$$\mathcal{L}\left(\bigcup_{i=1}^{\infty} G_i\right) = \sum_{i=1}^{\infty} \mathcal{L}(G_i).$$

Proof. We already know that $\mathcal{L}\left(\bigcup_{i=1}^{N} G_i\right) = \sum_{i=1}^{N} \mathcal{L}(G_i)$ for $N = 2$. By induction this must hold for any integer N. The sequential subadditive property supplies the final step. ■

3.6.3. Subadditive property for arbitrary sets. The theorem just proved for open sets extends readily to arbitrary sequences of sets.

THEOREM 3.6.2 (Subadditive property). *If $E_1, E_2, E_3, \ldots$ is a sequence of real sets then*

$$\mathcal{L}\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathcal{L}(E_i).$$

Proof. If $\sum_{i=1}^{\infty} \mathcal{L}(E_i) = \infty$ then there is nothing to prove. Otherwise, let $\epsilon > 0$ and choose a sequence of open sets G_1, G_2, G_3, ... in such a way that $E_i \subset G_i$ and

$$\mathcal{L}(G_i) < \mathcal{L}(E_i) + \epsilon 2^{-i}$$

for each $i = 1, 2, 3, \ldots$. Then $G = \bigcup_{i=1}^{\infty} G_i$ is an open set that contains the set $E = \bigcup_{i=1}^{\infty} E_i$. Using the subadditivity property for open sets (Theorem 3.27) we obtain

$$\mathcal{L}(E) \leq \mathcal{L}(G) \leq \sum_{i=1}^{\infty} \mathcal{L}(G_i) \leq \sum_{i=1}^{\infty} \left[\mathcal{L}(E_i) + \epsilon 2^{-i}\right].$$

Since $\sum_{i=1}^{\infty} \epsilon 2^{-i} = \epsilon$ and ϵ is an arbitrary positive number, the inequality of the theorem follows. ■

COROLLARY 3.29 (Additive property for separated sets). If $E_1, E_2, E_3, \ldots$ is a sequence of real sets and each pair E_i and E_j for $i \neq j$ is separated by open sets then

$$\mathcal{L}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathcal{L}(E_i).$$

Proof. Using the theorem and the fact that the measure is additive over finite collections of separated sets we obtain, for all n, that

$$\sum_{i=1}^{n} \mathcal{L}(E_i) = \mathcal{L}\left(\bigcup_{i=1}^{n} E_i\right) \leq \mathcal{L}\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathcal{L}(E_i).$$

From this the corollary now follows. ■

COROLLARY 3.30 (Additive property for closed sets). Suppose that E_1, E_2, E_3, ... is a disjointed sequence of closed sets. Then

$$\mathcal{L}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathcal{L}(E_i).$$

3.7. Null sets

There are several classes of sets that play a key role in the calculus. The class we introduce here play the role of the "negligible" sets, the sets that can frequently be discounted in a discussion.

DEFINITION 3.31. A set $N \subset \mathbb{R}$ is said to be a null set [or a set of Lebesgue measure zero] provided only that

$$\mathcal{L}(N) = 0.$$

THEOREM 3.32. Let $N_1,\ N_2,\ N_3,\ \dots$ be a sequence of null sets and suppose that

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

Then E is also a null set.

Proof. This follows immediately from the subadditive property of the measure. ■

3.8. Portions

If E is a closed set and (a, b) an open interval then

$$E \cap (a, b)$$

is called a *portion* of E provided only that $E \cap (a, b) \neq \emptyset$. It is possible that a portion could be trivial in that $E \cap (a, b)$ might contain only a single point of E; such a point is said to be an *isolated point* of E and we should be alert to the possibility that a portion might merely contain such a point.

3.8.1. Baire-Osgood Theorem. Our interest is in situations where E, E_1, E_2, E_3, … is a sequence of closed sets and we wish to be assured that one of the sets E_n contains a portion of E. This requires a compactness argument; the nested interval property is particularly suited to this problem.

LEMMA 3.33. Suppose that E and E_1 are nonempty closed sets and that E_1 contains no portion of E. Then there must exist a portion

$$E \cap (a, b)$$

so that $E_1 \cap (a, b) = \emptyset$.

Proof. Choose $z \in E$ but not in E_1. Consider the intervals

$$I_n = (z - 1/n, z + 1/n).$$

If $E_1 \cap I_n \neq \emptyset$ for all n then from Exercise 2.22 we can deduce that z would have to belong to the closed set E_1. ■

EXERCISE 3.34. Suppose that E, E_1, E_2, …, E_n are nonempty closed sets and that

$$E \subset \bigcup_{k=1}^{n} E_k.$$

Show that at least one of the sets E_k must contain a portion of E.

Hint: If $n = 1$ this is obvious. Use induction on n.

3.8.2. Baire-Osgood theorem. The Baire-Osgood theorem, one of the basic tools in our analysis later on, takes this exercise and extends the result to infinite sequences of closed sets.

THEOREM 3.35 (Baire-Osgood). Suppose that E, E_1, E_2, …, E_n, … are nonempty closed sets and that

$$E \subset \bigcup_{k=1}^{\infty} E_k.$$

Then at least one of the sets E_k must contain a portion of E.

Proof. Suppose not, i.e., suppose that none of the sets E_k contain a portion of E. Then, using Exercise 3.34, select a portion $E \cap (a_1, b_1)$ so that $E \cap (a_1, b_1) = \emptyset$ and pass to a closed subinterval $[c_1, d_1] \subset (a_1, b_1)$ for which $E \cap [c_1, d_1] \neq \emptyset$. Continue inductively, choosing portions

$$E \cap (a_n, b_n) \subset [c_{n-1}, d_{n-1}]$$

and closed subintervals $[c_n, d_n] \subset (a_n, b_n)$ for which $E \cap [c_n, d_n] \neq \emptyset$.

The nested sequence of intervals $\{[c_n, d_n]\}$ all must contain a point z of E in common. But this point z cannot belong to any of the sets E_k which is in contradiction to the hypothesis that $E \subset \bigcup_{k=1}^{\infty} E_k$. ∎

3.8.3. Some other versions of the Baire-Osgood theorem.

EXERCISE 3.36. Later on we will need this theorem without having to assume that E is closed. Show that theorem remains true if $E = \bigcap_{j=1}^{\infty} G_j$ where $\{G_j\}$ is some sequence of open sets.

Hint: To adjust the proof, at the nth stage of the induction select the interval $(a_n, b_n) \subset G_n$. The point z you will find must belong to each of the G_n and, consequently, to E. Sets of the form $E = \bigcap_{j=1}^{\infty} G_j$ for some sequence $\{G_j\}$ of open sets are said to be sets of type $\mathcal{G}_\delta$.

EXERCISE 3.37. If the closed set E is contained in a sequence of sets $\{E_n\}$ but we cannot be assured that they are closed sets then a simple device is to replace them by their closures. [The closure of a set E is the set $\overline{E}$ defined as the smallest closed set containing E.] If we do this show that the conclusion of the theorem would have to be, not that some set E_n contains a portion of E, but that some set E_n is dense in a portion of E.

3.9. Language of meager/residual subsets

The exploitation of the Osgood-Baire theorem can often be clarified by using the language of meager and residual subsets. If E is a closed set[3] of real numbers then a meager subset is one that represents a "small," insubstantial part of E; what remains after a meager subset is removed would be called a residual subset. It would be considered a "large" subset since only an insubstantial part has been removed.

DEFINITION 3.38. Let E be a closed set. A subset A of E is said to be *a meager subset of* E provided that there exists a sequence of closed sets $\{E_n\}$ none of which contains a portion of E so that

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

EXERCISE 3.39. Let E be a compact interval or $\mathbb{R}$. Show that a subset A of E is a meager subset of E provided that there exists a sequence of compact sets $\{E_n\}$ none of which contains a subinterval and so that

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

[3] In this section the language is restricted to subsets of closed sets. In view of Exercise 3.36 all of this would apply equally well to subsets of $\mathcal{G}_\delta$ sets, that is sets that are intersections of some sequence of open sets.

DEFINITION 3.40. Let E be a closed set. A subset A of E is said to be *a residual subset of* E provided that the complementary subset $E \setminus A$ is a meager subset of E.

THEOREM 3.41. Let E be a closed set. If A is a residual subset of E then A is dense in E.

Proof. This is a consequence of the Baire-Osgood theorem (Theorem 3.35). ■

3.9.1. The main tools. The main tools in using this language are the following easily proved lemmas.

LEMMA 3.42. Let $\{A_k\}$ be a sequence of meager subsets of a closed set E. Then $\bigcup_{k=1}^{\infty} A_k$ is also a meager subset of E.

LEMMA 3.43. Let $\{A_k\}$ be a sequence of residual subsets of a closed set E. Then $\bigcap_{k=1}^{\infty} A_k$ is also a residual subset of E.

While colorful and suggestive mathematical language plays a key role in much mathematical thinking, there is always the danger of relying too much on metaphors and too little on precise knowledge. For example, if a set is a meager subset of the interval $[0, 1]$, then surely is it not so insubstantial, so small, that it must have measure zero?

EXERCISE 3.44. Show that there is a subset A of the interval $[0, 1]$ for which $\mathcal{L}(A) = 1$ and yet A is a meager subset of $[0, 1]$.

CHAPTER 4

Real Functions

A real-valued function whose domain is some nonempty set of real numbers is called a *real function*. In the calculus we are mainly interested in functions

$$f : \mathbb{R} \to \mathbb{R}$$

and

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

In fact we could restrict our attention to one of these: if $f : [a, b] \to \mathbb{R}$ then extend by writing $f(x) = f(a)$ for $x < a$ and $f(x) = f(b)$ for $x > b$. If $f : \mathbb{R} \to \mathbb{R}$ then we can ignore values outside of $[a, b]$ if needed.

Most of our definitions are formulated in the language of covering relations. For functions $f : \mathbb{R} \to \mathbb{R}$ we use full covers, while for functions $f : [a, b] \to \mathbb{R}$ defined only on a compact interval $[a, b]$ we use Cousin covers of $[a, b]$. Often both versions will not need to be stated and the reader will adjust to whichever version is appropriate. For example the definition of "continuous function" is formally the same for functions $F : \mathbb{R} \to \mathbb{R}$ and for functions $F : [a, b] \to \mathbb{R}$ but slightly different interpretations and details would arise in the two situations.

4.1. Functions, increments, oscillations

The calculus is most often focussed on the amount at which a function changes in an interval. These can be expressed using the following interval functions.

DEFINITION 4.1. Suppose $f : \mathbb{R} \to \mathbb{R}$ and let $[x, y]$ be a compact interval. Then

(increment): $\Delta f([x, y]) = f(y) - f(x)$.
(oscillation): $\omega f([x, y]) = \sup\{|f(c) - f(d)| : x \leq c < d \leq y\}$.

4.1.1. Additivity of interval functions.

The interval function Δf is *additive* while ωf is *subadditive* in the sense that now follows:

EXERCISE 4.2 (Additivity). Let $f : [a, b] \to \mathbb{R}$ and suppose that π is a partition of $[a, b]$. Show that

$$\Delta f([a, b]) = \sum_{(I,x)\in\pi} \Delta f(I).$$

Hint: Clarify the notation by writing $a = a_0 < a_1 < a_2 < \dots a_{n-1} < a_n = b$ and $\pi = \{[a_{i-1}, a_i], x_i)\}$. Show that

$$\Delta f([a, b]) = \sum_{i=1}^{n} \Delta f([a_{i-1}, a_i]) = \sum_{i=1}^{n} [f(a_i) - f(a_{i-1})].$$

EXERCISE 4.3 (Subadditivity). Let $f : [a, b] \to \mathbb{R}$ and suppose that π is a partition of $[a, b]$. Show that

$$\omega f([a, b]) \leq \sum_{(I,x)\in\pi} \omega f(I).$$

4.2. Growth of a function

We will require a measure of how much a function $F : \mathbb{R} \to \mathbb{R}$ might grow on a set E. This is one of the central concerns of the calculus. We estimate this by measuring the size of the expression

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

over partitions and subpartitions π.

DEFINITION 4.4. Let $F : \mathbb{R} \to \mathbb{R}$, let E be a set, let π be a subpartition and let β be a covering relation. Then we write

(a)
$$V(\Delta F, \pi) = \sum_{([u,v],w)\in\pi} |\Delta F([u, v])|.$$
.

(b)
$$V(\Delta F, \beta) = \sup_{\pi\subset\beta} V(\Delta F, \pi)$$
where the supremum is taken over all subpartitions π that are contained in β.

(c)
$$V^*(\Delta F, E) = \inf_{\beta \text{ full}} V(\Delta F, \beta[E])$$
where the infimum is taken over all full covers β [or, equivalently, over all full covers of E.]

Note: Our understanding, as usual, is that if $F : [a, b] \to \mathbb{R}$ for some compact interval $[a, b]$ then in (c) we would use Cousin covers of $[a, b]$ in place of full covers.

We refer to all of these expressions as the *variation*. The first $V(\Delta F, \pi)$ is the variation over a partition, the second $V(\Delta F, \beta)$ is the variation over a covering relation, and the final measure $V^*(\Delta F, E)$ is the variation over a set. Our concerns in the sequel shall be

(a) Investigate the case of zero variation [functions that do not grow on a set].
(b) Investigate the case of finite variation [functions whose growth is bounded on a set].
(c) Obtain estimates on the variation or growth of a function on a set from certain local computations [Lipschitz numbers, derivatives].

EXERCISE 4.5 (Jordan Variation). Show that if $F : [a, b] \to \mathbb{R}$ then

$$V^*(\Delta F, [a, b]) = \sup \sum_{([u,v],w)\in\pi} |\Delta F([u, v])| = \sup V(\Delta F, \pi)$$

where the supremum is taken over all partitions π of the interval $[a, b]$. [Thus the variation of a function defined on a compact *interval* does not need a description in the language of Cousin covers.]

Exercise 4.6 (Jordan Variation). Show that if $F : [a, b] \to \mathbb{R}$ and F is nondecreasing then

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

4.3. Full characterization of Lebesgue's measure

The special case of the variation where $F(x) = x$ gives rise to a measure that is identical with the Lebesgue measure. This allows many relationships in the calculus between measure properties and other analytic properties (derivatives, etc.) to be determined by using a full covering argument. Historically such relationships were less transparent since the constructive version $\mathcal{L}$ of Lebesgue's measure was used exclusively.

Definition 4.7. For any set of real numbers E denote

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

Theorem 4.8. $\mathcal{L} = \mathcal{L}^*$.

Proof. We first prove that $\mathcal{L}^* \leq \mathcal{L}$. For any set E and any open set $G \supset E$ note that

$$\beta = \{([u, v], w) : [u, v] \subset G\}$$

is a full cover of E. Consequently

$$\mathcal{L}^*(E) \leq V(\mathcal{L}, \beta) \leq \mathcal{L}(G).$$

Since this is true for all open sets G that contain E it follows that $\mathcal{L}^*(E) \leq \mathcal{L}(E)$.

In the other direction suppose that $\mathcal{L}^*(E) < \infty$ (or there is nothing to prove), and choose any full cover β of E. For each w in the set E choose a positive number $\delta(w)$ so that $([u, v], w) \in \beta$ whenever $v - u < 2\delta(w)$. Define the open set

$$G = \bigcup_{w \in E} (w - \delta(w), w + \delta(w)).$$

Write $G = \bigcup_{i=1}^{\infty} (a_i, b_i)$ for the open component intervals of G. Let $t < \mathcal{L}(G)$ and select intervals

$$[c_1, d_1], [c_2, d_2], [c_3, d_3], \dots [c_m, d_m]$$

with $[c_i, d_i] \subset (a_i, b_i)$ in such a way that

$$\sum_{i=1}^{m} (d_i - c_i) > t$$

Fix i. The Heine-Borel compactness argument allows us to select a finite numbers of points $w_1, w_2, \dots, w_p$ from E in such a way that

$$[c_i, d_i] \subset \bigcup_{k=1}^{p} (w_k - \delta(w_k), w_k + \delta(w_k)).$$

We remove any w_k from this list that is redundant and then relabel the points so that $w_1 < w_2 < \cdots < w_p$.

We construct a partition π of the interval $[c_i, d_i]$ by a simple device now: start with $([c_i, w_1], w_1)$, then pick a point t_1 in between w_1 and w_2 so that both

$$([w_1, t_1], w_1) \text{ and } ([t_1, w_2], w_2)$$

belong to β. Continue by selecting a point t_2 in between w_1 and w_2 so that both

$$([w_2, t_2], w_2) \text{ and } ([t_2, w_3], w_3)$$

belong to β. Continue, finishing with $([w_p, d_i], w_p)$. This partition π is evidently a subset of β. Certainly

$$V(\mathcal{L}, \pi) = d_i - c_i.$$

We can do this for all $i = 1, 2, 3, \dots, m$ and, in this way discover that

$$V(\mathcal{L}, \beta) \geq \sum_{i=1}^{m} (d_i - c_i) > t.$$

Since this holds for such t we now know that

$$V(\mathcal{L}, \beta) \geq \mathcal{L}(G) \geq \mathcal{L}(E)$$

and hence that

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

From this we see that, in general, $\mathcal{L} \leq \mathcal{L}^*$ completing the proof. ■

4.4. Properties of the variation

THEOREM 4.9. Suppose that $F : \mathbb{R} \to \mathbb{R}$ and that $[a, b]$ is a compact interval. Then

$$|F(b) - F(a)| \leq V^*(\Delta F, [a, b]).$$

Proof. If β is any full cover of $[a, b]$ then β contains a partition π of the interval $[a, b]$. Observe that

$$|F(b) - F(a)| = \left| \sum_{([u,v],w)\in\pi} \Delta F([u, v]) \right| \leq \sum_{([u,v],w)\in\pi} |\Delta F([u, v])| = V(\Delta F, \pi).$$

The theorem follows from this. ■

THEOREM 4.10. Suppose that $F : \mathbb{R} \to \mathbb{R}$ and that E, E_1, E_2, ... is a sequence of sets for which

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

Then

$$V^*(\Delta F, E) \leq \sum_{n=1}^{\infty} V^*(\Delta F, E_n).$$

Proof. If

$$\sum_{n=1}^{\infty} V^*(\Delta F, E_n) = \infty$$

there is nothing to prove. Otherwise, let $\epsilon > 0$ and choose full covers β_i of E_i so that

$$V(\Delta F, \beta_i) < V^*(\Delta F, E_i) + \epsilon 2^{-i}.$$

Check that $\beta = \bigcup_{i=1}^{\infty} \beta_i[E_i]$ is a full cover of E. Consequently

$$V^*(\Delta F, E) < V(\Delta F, \beta) \leq \sum_{i=1}^{\infty} V(\Delta F, \beta_i) \leq \sum_{i=1}^{\infty} V^*(\Delta F, E_i) + \epsilon.$$

■

4.5. Functions that do not grow on a set

DEFINITION 4.11. Let $F : \mathbb{R} \to \mathbb{R}$ and let E be a set of real numbers. We say that F *does not grow* on the set E provided that for every $\epsilon > 0$ there is a full cover β so that

$$\sum_{([u,v],w)\in\pi} |\Delta F([u,v])| < \epsilon \tag{4.1}$$

whenever π is a subpartition, $\pi \subset \beta[E]$.

Note: If $F : [a,b] \to \mathbb{R}$ then the definition should be altered to require a Cousin cover β of $[a,b]$.

We should recognize immediately that a function does not grow on a set E if and only if it has zero variation there, i.e., if $V^*(\Delta F, E) = 0$. Thus "does not grow" and "zero variation" will be taken as synonymous.

EXERCISE 4.12. Show that N is a null set if and only if the function $F(x) = x$ does not grow on N.

EXERCISE 4.13. Show that in the definition the requirement that (4.1) holds can be replaced by the (apparently stronger) requirement:

$$\sum_{([u,v],w)\in\pi} \omega F([u,v]) < \epsilon \tag{4.2}$$

4.6. Some properties of growth

LEMMA 4.14. Let $F : [a,b] \to \mathbb{R}$ and suppose that F does not grow on $[a,b]$. Then F is constant.

Proof. This follows from the Cousin covering lemma. ■

LEMMA 4.15. Let $F : \mathbb{R} \to \mathbb{R}$, let E_1, E_2, E_3, ... be a sequence of sets and suppose that F does not grow on E_i $(i = 1, 2, 3, \dots)$. Then F does not grow on any subset of the union $\bigcup_{i=1}^{\infty} E_i$.

Proof. This follows from Theorem 4.10. ■

4.7. Continuity and absolute continuity

The notions of continuity and absolute continuity arise from the focus on functions that do not grow on small sets.

DEFINITION 4.16. Let $F : \mathbb{R} \to \mathbb{R}$.

(a) We say that F is *continuous at a point* x_0 provided that F does not grow on the singleton set $E = \{x_0\}$.
(b) We say that F is *continuous* provided that F does not grow on any countable set.
(c) We say that F is *absolutely continuous* provided that F does not grow on any null set.

An opposite kind of notion is described by the notion of singularity. It is possible for a function to grow *only* on a null set.

DEFINITION 4.17. Let $F : \mathbb{R} \to \mathbb{R}$. We say that F is *singular* provided that there is a null set N and F does not grow on $\mathbb{R} \setminus N$.

Note 1. If $F : [a, b] \to \mathbb{R}$ then the definitions of continuity and absolute continuity would not need to be altered, but we must then recall what it means for such a function not to grow on a set.

4.7.1. Properties of continuous and absolutely continuous functions. A simple ϵ, δ characterization of continuity at a point follows easily from the nature of full covers.

LEMMA 4.18. Let $F : \mathbb{R} \to \mathbb{R}$ and let x_0 be a real number. Then F is continuous at x_0 if and only if for every $\epsilon > 0$ there exists a $\delta > 0$ so that

$$|F(x) - F(x_0)| < \epsilon$$

for all $|x - x_0| < \delta$.

THEOREM 4.19. Let $F : \mathbb{R} \to \mathbb{R}$. Then F is continuous if and only if F is continuous at every point in $\mathbb{R}$.

Proof. Use Lemma 4.15. If F is continuous at each point of a countable set

$$C = \{c_1, c_2, c_3, \dots\}$$

then it cannot grow on the set

$$C = \bigcup_{i=1}^{\infty} \{c_i\}.$$

■

THEOREM 4.20. If a function F is absolutely continuous it is also necessarily continuous.

THEOREM 4.21. If $F : \mathbb{R} \to \mathbb{R}$ is both absolutely continuous and singular then F is constant.

EXERCISE 4.22. Let $F : \mathbb{R} \to \mathbb{R}$ and let x_0 be a real number. Show that F is continuous at x_0 if and only if for every $\epsilon > 0$ there exists a $\delta > 0$ so that

$$\omega F(I) < \epsilon$$

for all intervals I that contain x_0 and have length smaller than δ.

EXERCISE 4.23. Suppose that each of F, F_1 $F_2 : [a, b] \to \mathbb{R}$ is continuous at a point x_0. Show that $|F|$ and any linear combination $rF_1 + rF_2$ are also continuous at x_0.

4.8. Functions uniformly continuous throughout an interval

In order for a function $F : [a, b] \to \mathbb{R}$ to be continuous we recognize immediately (Lemma 4.18) that this would be equivalent to requiring that for every point x_0 and for every $\epsilon > 0$ there exists a $\delta > 0$ so that

$$|F(x) - F(x_0)| < \epsilon$$

for all $|x - x_0| < \delta$, $x \in [a, b]$.

It is easy to misread this assertion; indeed Cauchy, who was one of the originators of these ideas, did get confused over a very subtle point here. It seems that continuity merely expects the increment

$$\Delta F([c,d]) = F(d) - F(c)$$

to be small if the interval $[c,d]$ is small. But that is a formally stronger requirement which we capture using this language:

DEFINITION 4.24. Let $F : [a,b] \to \mathbb{R}$. Then F is said to be *uniformly continuous throughout* $[a,b]$ provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that for every subinterval $[c,d] \subset [a,b]$ if $\mathcal{L}([c,d]) < \delta$ then

$$|\Delta F([c,d])| < \epsilon.$$

The theorem connecting continuity and uniform continuity is remarkable, but is simply proved by a covering argument.

THEOREM 4.25. Let $F : [a,b] \to \mathbb{R}$. Then F is continuous if and only if F is uniformly continuous throughout the interval $[a,b]$.

Proof. If F is uniformly continuous throughout $[a,b]$ then it is clear that it is continuous at each point, thus continuous. Conversely, suppose that F is continuous.

Let

$$\beta = \{(I,x) : \ x \in I \text{ and } \omega f(I) < \epsilon/2\}.$$

Check that β is a Cousin cover of $[a,b]$. There is, hence, a partition π of $[a,b]$ contained in β. Let

$$\delta = \frac{1}{2}\max\{\mathcal{L}(I) : (I,x) \in \pi\}.$$

Suppose that J is any compact subinterval of $[a,b]$ with length less than δ. Such a J can overlap with no more than two intervals from the partition. Thus, using the subadditivity of the interval function ωf,

$$\omega f(J) \leq \sum_{(I,x)\in\pi} \omega(J \cap I)$$

and only two terms appear on the right-hand side, each of them smaller than $\epsilon/2$. This implies that $|\Delta f(J)| \leq \omega f(J) < \epsilon$ for any such J, proving that f is uniformly continuous throughout $[a,b]$. ■

EXERCISE 4.26. Show that Theorem 4.25 is not valid for continuous functions $F : \mathbb{R} \to \mathbb{R}$. Let $F(x) = x^2$. Check that F is continuous but that there is no number $\delta > 0$ with the property that, for every interval $[c,d]$ if $\mathcal{L}([c,d]) < \delta$ then $|\Delta F([c,d])| < \epsilon$.

4.9. Vitali's condition

We have just seen in Theorem 4.25 that, for functions $F : [a,b] \to \mathbb{R}$, continuity can be expressed by a single ϵ, δ-statement. There is a similar, but weaker, ϵ, δ-assertion for absolute continuity. The concept that expresses this notion is due to Vitali.

DEFINITION 4.27. Let $F : [a,b] \to \mathbb{R}$. We say that F is *absolutely continuous in the sense of Vitali* on $[a,b]$ provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever π is a subpartition from $[a,b]$ for which

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) < \delta$$

then

$$\sum_{(I,x)\in\pi} |\Delta F(I)| < \epsilon.$$

THEOREM 4.28. Let $F : [a,b] \to \mathbb{R}$ and suppose that F is absolutely continuous in the sense of Vitali. Then F is absolutely continuous.

EXERCISE 4.29. Show that the converse of Theorem 4.28 is invalid.

Hint: Define $F : [0,1] \to \mathbb{R}$ by $F(0) = F(1/(2n-1)) = 0$ and $F(1/2n) = 1/n$ for all $n = 1, 2, 3, \ldots$. Extend F to be linear on each of the intervals contiguous to these points where it has so far been defined. Show that F is absolutely continuous but that Vitali's condition does not hold.

4.10. Continuous functions map compact intervals to compact intervals

THEOREM 4.30. Let $F : [a,b] \to \mathbb{R}$ be continuous, but not constant. Then the image of $[a,b]$ under f is a compact interval $[A,B]$.

Proof. We prove first that the image of F is bounded using a simple covering argument. Take

$$\beta = \{(I,x) : \omega F(I) \leq 1\}.$$

Check that is a Cousin cover of $[a,b]$. By the Cousin covering lemma there is a partition π of $[a,b]$ contained in β. Let K be the number of elements of π and check that

$$|F(x) - F(a)| \leq K$$

for all $x \in [a,b]$ and hence that

$$|F(x)| \leq |f(a)| + K.$$

Now, since the range of F is a bounded set, we may write A and B for the inf and sup of this range. So the range of F is contained in $[A,B]$ although we do not yet know that this interval is the image set. We shall establish that F takes on every value in $[A,B]$ using a covering argument that may not be obvious at first sight.

Let $A \leq p \leq B$ and assume, contrary to what we wish to prove, that $F(x)$ is never equal to p. Let $F_1(x) = f(x) - p$ so F_1 is also continuous (merely because $\Delta F = \Delta F_1$) and $F_1(x)$ is never 0. Let

$$\beta = \{([x,y],z) : z \in [x,y] \text{ and } |f(y) - f(x)| < |F_1(z)|/2.\ \}.$$

It is easy to check that this is a Cousin cover of $[a,b]$. Thus there must be a partition π of $[a,b]$ contained in β. Let

$$\epsilon = \min\{|F_1(z)|/2 : (I,z) \in \pi\}.$$

By the nature of the construction we can check that all values $F_1(z)$ for which $(I,z) \in \pi$ are of the same sign. Just take two contiguous intervals, say $[r,s]$ and $[s,t]$ with $([r,s],z_1)$, $([s,t],z_2) \in \pi$. Observe in particular that

$$|F_1(s) - F_1(z_1)| < |F_1(z_1)|/2$$

and

$$|F_1(s) - F_1(z_2)| < |F_1(z_2)|/2.$$

Hence if, say, (contrary to our claim) $F_1(z_2) < 0 < F_1(z_1)$ then

$$F_1(s) > F_1(z_2)/2 > 0 > F_1(z_2)/2 > F_1(s)$$

which is impossible. Thus, arguing one-by-one along the intervals of the partition it must be true that that all values $F_1(z)$, for which $(I, z) \in \pi$ are of the same sign.

If all these values are positive then $F_1(x) > \epsilon$ for all $x \in [a, b]$ which means

$$F(x) > p + \epsilon > A$$

violating the definition of A. If all are negative then

$$F_1(x) < -\epsilon$$

for all $x \in [a, b]$ which means

$$F(x) < p - \epsilon < B$$

violating the definition of B. Both are impossible thus F assumes the value p. ■

EXERCISE 4.31 (Existence of extreme values). Let $F : [a, b] \to \mathbb{R}$ be continuous. Show that there is a maximum value and a minimum value of the function assumed at points of $[a, b]$.

EXERCISE 4.32 (Darboux property). A function $F : [a, b] \to \mathbb{R}$ is said to have the *Darboux property* if for any points x, y from $[a, b]$ for which $F(x) \neq F(y)$ and any number C between $F(x)$ and $F(y)$ there is a number c between x and y for which $F(c) = C$. Show that if $F : [a, b] \to \mathbb{R}$ is continuous then f has the Darboux property on that interval.

In Exercise 5.7.1 we will find that every derivative also has the Darboux property, although derivatives need not be continuous. Thus this property, which may seem similar to continuity, is distinct from it.

4.11. Discontinuities

A local measure of continuity can be defined as follows.

DEFINITION 4.33. Let $F : \mathbb{R} \to \mathbb{R}$ and let x_0 be a real number. Then we write

$$\omega_F(x_0) = \inf_\beta \left(\sup\{|\Delta F([u, v])| : ([u, v], x_0) \in \beta\}\right)$$

where the infimum is taken over all full covers β. The expression $\omega_F(x_0)$ is called the *oscillation* of the function F at the point x_0.

It is easy to check that F is continuous at x_0 if and only if $\omega_F(x_0) = 0$. Thus, if this number is positive it is a measure of how continuity might fail at the point x_0. If $F : \mathbb{R} \to \mathbb{R}$ fails to be continuous at a point x_0 then $\omega_F(x_0) > 0$ and we call x a point of *discontinuity*.

THEOREM 4.34. Let $F : \mathbb{R} \to \mathbb{R}$. The set D_f of points at which F is discontinuous can be displayed as the union of an increasing sequence of closed sets:

$$D_f = \bigcup_{n=1}^{\infty} \{x : \omega_F(x) \geq 1/n\}.$$

Proof. The only part not obvious is that each set of the form

$$E = \{x : \omega_F(x) \geq c\}$$

is closed. Equivalent would be to show that the complementary set

$$E' = \{x : \omega_F(x) < c\}$$

is open. Let us consider the latter.

Take any $x \in E'$. Since $\omega_F(x) < c$ there must be an open interval (a, b) containing x so that $\omega F([a, b]) < c$. But then $\omega_F(x') \leq \omega F([a, b]) < c$ for all $x' \in (a, b)$. Consequently $(a, b) \subset E'$ and, since this can be done for any $x \in E'$ we see that E' is open, and that E is closed. ∎

EXERCISE 4.35. Let $F : \mathbb{R} \to \mathbb{R}$ and let x_0 be a real number. Show that

$$\omega_F(x_0) = \inf_{\delta > 0} \sup\{|F(x) - F(x_0)| : |x - x_0| < \delta\}.$$

EXERCISE 4.36. Show that if $\{E_n\}$ is an increasing sequence of closed sets then there is a function $f : \mathbb{R} \to \mathbb{R}$ whose set of points of discontinuity is exactly $\bigcup_{n=1}^{\infty} E_n$.

4.12. Extension of continuous functions

DEFINITION 4.37. Let K be a compact set and suppose that $f : K \to \mathbb{R}$. Then f is said to be *uniformly continuous relative to* K provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that $|\Delta f([c, d])| < \epsilon$ whenever $[c, d]$ is an interval having both endpoints c and d in K and $\mathcal{L}([c, d]) < \delta$.

When a function is continuous throughout a compact set K in this way it is easy to extend it to a continuous function.

THEOREM 4.38. Let K be a compact set and suppose that the function $f : K \to \mathbb{R}$ is uniformly continuous relative to K. Then there is a continuous function $g : \mathbb{R} \to \mathbb{R}$ so that $f(x) = g(x)$ for all $x \in K$.

Proof. First let us display the complementary intervals $\{(a_i, b_i)\}$ to K. Two of these intervals are infinite, say $(-\infty, a)$ and (b, ∞).

The function g is to be defined on each of these complementary intervals in the simplest way that preserves continuity at each point. The function g is first defined so that $g(x) = f(x)$ at each point $x \in K$. Then set $g(x) = f(a)$ for all $x < a$ and $g(x) = f(b)$ for all $x > b$. Let (c, d) be any one of the bounded intervals complementary to K. Then both c and d belong to K. Set

$$g(c + t(d - c)) = f(c) + t(f(d) - f(c)) \quad (0 < t < 1).$$

Such a function is continuous at each point (as the student will check) so provides a continuous function that agrees with f on K. ∎

4.13. Convergent sequences map to convergent sequences

THEOREM 4.39. Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose that $\{x_n\}$ is a convergent sequence of real numbers from $[a, b]$. Then

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

For continuity at a single point there remains at least a partial conclusion about convergent sequences.

THEOREM 4.40. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous at a point x. Then if $\{x_n\}$ is a convergent sequence of real numbers and convergent to x, the sequence $\{f(x_n)\}$ is also convergent and

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

EXERCISE 4.41. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that f maps compact sets to compact sets but need not map closed sets to closed sets.

Hint: Use Exercise 2.23.

4.14. Limits of continuous functions

Suppose that $\{f_n\}$ is a sequence of functions defined on an interval $[a, b]$ and that

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

for every $x \in [a, b]$. We would reasonably expect f should inherit whatever properties are possessed by members of the sequence.

Specifically, if $f_n : [a, b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ is a sequence of functions each of which is continuous and $f_n(x)$ converges to $f(x)$ at each point x, does it follow that f too must be continuous? It is not hard to come up with an example showing that f may be discontinuous.

The key to allowing such a conclusion is to ensure that the conditions defining the continuity or the conditions defining the convergence are "uniform." That is we require that the continuity of f_n is attained in the same way for all n, or alternatively that the convergence of the sequence $\{f_n(x)\}$ is attained in the same way for all x.

DEFINITION 4.42. A sequence of functions $f_n : [a, b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ is said to be *equicontinuous* provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that $|\Delta f_n([c, d])| < \epsilon$ whenever $\mathcal{L}([c, d]) < \delta$ and n is any natural number.

DEFINITION 4.43. A sequence of functions $f_n : [a, b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ is said to be is said to be *uniformly convergent* to a function f on $[a, b]$ provided that for every $\epsilon > 0$ there is an integer $N > 0$ so that

$$|f_n(x) - f(x)| < \epsilon$$

whenever $n \geq N$ and x is any point in $[a, b]$.

THEOREM 4.44 (Continuity of limits). Suppose that $f_n : [a, b] \to \mathbb{R}$ for $n = 1, 2, 3, \dots$ is a sequence of continuous functions and that the sequence of real numbers $\{f_n(x)\}$ converges to $f(x)$ for every $x \in [a, b]$. Then $f : [a, b] \to \mathbb{R}$ is continuous if one at least of the following conditions holds:

(a) $\{f_n\}$ is uniformly convergent to $f(x)$ on $[a, b]$, or
(b) $\{f_n\}$ is equicontinuous.

Proof. Use the inequality

$$|f(x) - f(y)| \leq |f_n(x) - f(x)| + |f_n(y) - f(y)| + |f_n(x) - f_n(y)|.$$

Under assumption (a) we can choose m so that $|f_n(t) - f(t)| < \epsilon/3$ for all $t \in [a,b]$ and all $n \geq m$. That allows us to claim, in particular that

$$|f(x) - f(y)| \leq +|f_m(x) - f_m(y)| + 2\epsilon/3. \tag{4.3}$$

Similarly under assumption (b) we can choose $\delta > 0$ so that $|f_n(x) - f_n(y)| < \epsilon/3$ for all n and all $|x - y| < \delta$. That allows us to claim, in particular that

$$|f(x) - f(y)| \leq |f_n(x) - f(x)| + |f_n(y) - f(y)| + \epsilon/3. \tag{4.4}$$

The remainder of the proof just uses the continuity assumptions on the function f_m in the inequality (4.3) and the convergence assumptions of the sequence of functions at the two points x and y in (4.4). ■

EXERCISE 4.45. Suppose that $f_n : \mathbb{R} \to \mathbb{R}$ for $n = 1 = 2, 3, \dots$ is a sequence of functions, each of which is continuous relative to a compact set K. Suppose that $\{f_n\}$ is uniformly convergent to $f(x)$ on K. Show that f is continuous relative to K.

4.15. Local Continuity of limits

Without the uniformity assumptions in Theorem 4.44, the function f that is a limit of a sequence of continuous functions need not be continuous. We now show that the set of points of continuity is dense: every subinterval must contain a point at which f is continuous.

THEOREM 4.46. Suppose that $f_n : [a,b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ is a sequence of continuous functions and that the sequence of real numbers $\{f_n(x)\}$ converges to $f(x)$ for every $x \in [a,b]$. Then every subinterval $[c,d]$ of $[a,b]$ contains a point of continuity of the function f.

Proof. We are looking to find somewhere in a given interval $[c,d]$ a point x where $\omega_f(x) = 0$. Let $\epsilon > 0$.

Step 1. First of all check, for any integers n, k, m, and all $\epsilon > 0$, that each set $\{x \in [a,b] : |f_n(x) - f_k(x)| \leq \epsilon\}$ is closed (because the functions in the sequence are all continuous). Let

$$E_{n\epsilon} = \bigcap_{i=n}^{\infty} \bigcap_{k=n}^{\infty} \{x \in [a,b] : |f_i(x) - f_k(x)| \leq \epsilon\}$$

and observe that these sets, too, are closed. For any $x \in [a,b]$ the sequence of $\{f_n(x)\}$ converges and it follows from the Cauchy criterion that x must belong to one of the sets $\{E_{n\epsilon}\}$. Thus

$$[a,b] \subset \bigcup_{k=1}^{\infty} E_{k\epsilon}.$$

Step 2. We recall from Theorem 4.34 in Section 4.11 that the set of points where f is discontinuous can be expressed as a union of a sequence of closed sets, $\bigcup_{n=1}^{\infty} A_n$, where

$$A_n = \{x : \omega_f(x) \geq 1/n\}.$$

Step 3. With these preliminary steps out of the way we now prove the theorem. Suppose contrary to the statement of the theorem that F is discontinuous at every point in $[c,d]$. This means, for every point $x \in [c,d]$, that $\omega_f(x) > 0$. In particular then $[c,d] \subset \bigcup_{n=1}^{\infty} A_n$.

An application of the Baire-Osgood theorem shows that there must be a subinterval $[c', d'] \subset [c, d]$ and a set A_N so that $[c', d'] \subset A_N$. Choose $\epsilon < 1/(2N)$ and note that

$$[c', d'] \subset \bigcup_{n=1}^{\infty} E_{k\epsilon}.$$

Another application of the Baire-Osgood theorem (3.35) shows that there must be a subinterval $[c'', d''] \subset [c', d'] \subset [c, d]$ and a set $E_{p\epsilon}$ so that $[c'', d''] \subset E_{p\epsilon}$. Using the fact that this function f_p is continuous we may pass to a further subinterval $[c_0, d_0] \subset [c'', d'']$ so that $\Delta f_p([c_0, d_0]) < \epsilon$.

We now show that this is impossible. For if x is any point from this interval, then the fact that it is also in $E_{p\epsilon}$ means that

$$|f_p(x) - f_i(x)| \leq \epsilon$$

for all $i \geq p$. Consequently (taking limits on i) we also have

$$|f_p(x) - f(x)| \leq \epsilon$$

for all points $x \in [c_0, d_0]$. For this inequality to hold at all such points x must require that the

$$\Delta f([c_0, d_0]) < 2\epsilon < 1/N.$$

In particular for every point $x \in (c_0, d_0)$ we see that $\omega_f(x) < 1/N$. This contradicts the fact that all such points x must belong to A_N and satisfy instead $\omega_f(x) \geq 1/N$. ■

4.16. Bernstein polynomials

Continuous functions can be approximated by polynomials on any compact interval. The polynomial approximation can be effected with a natural kind of computation. We work only in the interval $[0, 1]$ but the results extend easily to any compact interval $[a, b]$ by finding an affine transformation $\phi : [a, b] \to [0, 1]$ and using ϕ and ϕ^{-1} to translate back and forth between functions on the two compact intervals $[a, b]$ and $[0, 1]$.

Use the binomial theorem to check that

$$1 = [(1-x) + x]^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \tag{4.5}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This identity and probabilistic intuition suggests the following definition.

Definition 4.47. Let $f : [0, 1] \to \mathbb{R}$ be a continuous function. Then the polynomial

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

is called the Bernstein polynomial of degree n associated with f.

THEOREM 4.16.1. *Let $f : [0,1] \to \mathbb{R}$ be a continuous function and let*

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k}.$$

Then B_n converges uniformly to f on $[0,1]$.

Proof. Let $\epsilon > 0$. Choose $M > |f(x)|$ for all x in $[0,1]$ and choose $\delta > 0$ so that the oscillation of f is smaller than $\epsilon/2$ on any interval shorter in length than δ. Finally choose N so that $1/N < \delta$ and $N > M/(4\epsilon\delta^2)$.

Suppose now that $n \geq N$ and fix $x \in [0,1]$. Then

$$|B_n(x) - f(x)| = \left|\sum_{k=0}^{n}\left[f\left(\frac{k}{n}\right) - f(x)\right]\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k}\right| \leq$$

$$\sum_{k=0}^{n}\left|f\left(\frac{k}{n}\right) - f(x)\right|\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k}.$$

Let

$$A = \{k \in \{0,1,2,\ldots,n\} : |k/n - x| < \delta\}$$

and

$$B = \{k \in \{0,1,2,\ldots,n\} : |k/n - x| \geq \delta\}$$

The computation

$$\sum_{k\in A}\left|f\left(\frac{k}{n}\right) - f(x)\right|\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k} < \epsilon/2$$

follows easily from the identity (4.5) and the fact that

$$\left|f\left(\frac{k}{n}\right) - f(x)\right| < \epsilon/2$$

for each $k \in A$. For $k \in B$ note that

$$\frac{(k-nx)^2}{n^2\delta^2} \geq 1.$$

Thus, using using Exercise 4.48, we can obtain that

$$\sum_{k\in B}\left|f\left(\frac{k}{n}\right) - f(x)\right|\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k} \leq$$

$$\frac{2M}{n^2\delta^2}\sum_{k=0}^{n}(k-nx)^2\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k} = nx(1-x) \leq \frac{M}{4n\delta^2} < \epsilon/2.$$

Putting these together gives

$$|B_n(x) - f(x)| < \epsilon$$

for all $n \geq N$ and all x in the interval $[a,b]$. ∎

EXERCISE 4.48. Show, for all x, that

$$\sum_{k=0}^{n}(k-nx)^2\left(\begin{array}{c} n \\ k \end{array}\right) x^k(1-x)^{n-k} = nx(1-x) \leq n/4.$$

Hint: Start with $z = x/(1-x)$ and the identity

$$(1+z)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) z^k.$$

By differentiating the identity with respect to z and multiplying by z (do it twice), obtain

$$n^2x^2 \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^k(1-x)^{n-k} = n^2x^2 = (-2nx)$$

$$\sum_{k=0}^{n} k \left(\begin{array}{c} n \\ k \end{array} \right) x^k(1-x)^{n-k} = (-2nx)(nx) = -2n^2x^2$$

and

$$\sum_{k=0}^{n} k^2 \left(\begin{array}{c} n \\ k \end{array} \right) x^k(1-x)^{n-k} = nx(1-x+nx).$$

CHAPTER 5

The Derivative

The central concerns of the calculus are the derivative, the integral, and the measure. We address the derivative in this chapter and the integral in the next.

5.1. Growth of a function

The increment
$$\Delta f([a,b]) = f(b) - f(a)$$
represents the amount by which the function f has grown in the interval $[a,b]$. A useful estimate is to ask for the "average growth" over the interval by dividing by the length of the interval:
$$\frac{\Delta f([a,b])}{\mathcal{L}([a,b])} = \frac{f(b) - f(a)}{b-a}.$$
Elementary students will recognize this as the "slope" of the line joining the two points
$$(a, f(a)) \quad \text{and} \quad (b, f(b))$$
in the plane.

This expression carries no information about the behaviour of the function f on the entire interval, only about the relation between the values at the two endpoints a and b. We shall study a number of variants on this theme all related to the problem of estimating the growth properties of a function.

5.1.1. Lipschitz functions.

DEFINITION 5.1. Let $f : [a,b] \to \mathbb{R}$. Then we write
$$Lip_f([a,b]) = \sup\left\{ \left| \frac{\Delta f([c,d])}{\mathcal{L}([c,d])} \right| : [c,d] \subset [a,b] \right\}$$
and call this the *Lipschitz constant* for f on the interval $[a,b]$.

The Lipschitz constant is either ∞ or it offers a useful estimate of the maximum rate at which f might rise or fall within the interval. If $Lip_f([a,b]) < \infty$ then f is said to be a *Lipschitz function* on the interval $[a,b]$.

EXERCISE 5.2. Suppose that $f : [a,b] \to \mathbb{R}$ is a Lipschitz function. Show that f is absolutely continuous [in fact, absolutely continuous in the sense of Vitali].

EXERCISE 5.3. Suppose that $f : [a,b] \to \mathbb{R}$ is a Lipschitz function. Show that $Lip_f([a,b]) = 0$ if and only if f is constant on $[a,b]$.

5.2. Lipschitz number at a point

A local version of Lipschitz constants is available too. As usual we define this in terms of Cousin covers and full covers.

Definition 5.4. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $x_0 \in \mathbb{R}$. Then

$$lip_f(x_0) = \inf_{\beta} \sup \left\{ \left| \frac{\Delta f([c,d])}{\mathcal{L}([c,d])} \right| : ([c,d], x_0) \in \beta \right\},$$

where the infimum is taken over all full covers β, is called the *Lipschitz number* of f at x_0.

Note: If $f : [a,b] \to \mathbb{R}$ replace the full covers by Cousin covers of $[a,b]$ in the definition. This is needed so that the estimates of $lip_f(a)$ and $lip_f(b)$ make sense.

The Lipschitz numbers are directly related to growth.

Lemma 5.5. Let $f : [a,b] \to \mathbb{R}$ and suppose that $lip_f(x) \leq M$ for all x in $[a,b]$. Then f is Lipschitz on $[a,b]$ and $Lip_f([a,b]) \leq M$.

Proof. Let $t > M$ and define the covering relation

$$\beta = \{(I,x) : x \in I \text{ and } |\Delta f(I)| < t\mathcal{L}(I)\}.$$

It is easy to check that β is a Cousin cover of $[a,b]$. Take any partition π of a subinterval $[c,d]$ from β and check that

$$|\Delta f([c,d])| \leq \sum_{(I,x)\in\pi} |\Delta f(I)| \leq \sum_{(I,x)\in\pi} t\mathcal{L}(I) = t(d-c).$$

Hence $Lip_f([c,d]) \leq t$ and the conclusion follows. ∎

With a very similar proof we have the following variant on this which is more general since it relates to growth on an arbitrary set.

Lemma 5.6. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that

$$lip_f(x) \leq M$$

for all x in a set E. Then

$$V^*(\Delta F, E) \leq M\mathcal{L}^*(E).$$

Exercise 5.7. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $x_0 \in \mathbb{R}$. Show that

$$lip_f(x_0) = \inf_{\delta>0} \sup \left\{ \left| \frac{f(x)-f(x_0)}{x-x_0} \right| : 0 < |x-x_0| < \delta \right\}.$$

Exercise 5.8. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $lip_f(x_0) < \infty$. Show that f is continuous at x_0.

Exercise 5.9. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $lip_f(x) < \infty$ at each point x of a null set N. Show that f does not grow on N.

5.3. The Derivative

The Lipschitz constant and the Lipschitz numbers relate to the growth of a function, but do not pay attention whether that growth is positive or negative. The derivative does. The definition is nearly identical with the definition for Lipschitz numbers; because the values considered may be positive and negative we need both upper and lower limits.

Definition 5.10. Let $f : \mathbb{R} \to \mathbb{R}$ and let $x_0 \in \mathbb{R}$. Then

$$\overline{D}f(x_0) = \inf_{\beta} \sup \left\{ \frac{\Delta f([c,d])}{\mathcal{L}([c,d])} : ([c,d], x_0) \in \beta \right\},$$

where the infimum is taken over all full covers β, is called the *upper derivate* of f at x_0. Similarly

$$\underline{D}f(x_0) = \sup_{\beta} \inf \left\{ \frac{\Delta f([c,d])}{\mathcal{L}([c,d])} : ([c,d], x_0) \in \beta \right\}$$

is called the *lower derivate* of f at x.

If these are equal then the common value would be written as $Df(x_0)$ or more frequently as $f'(x_0)$ and called the *derivative*. If the derivative exists and is also finite then f is said to be *differentiable*.

Exercise 5.11. Let $f : \mathbb{R} \to \mathbb{R}$ and let $x_0 \in \mathbb{R}$. Show that

$$\overline{D}f(x_0) = \inf_{\delta > 0} \sup \left\{ \frac{f(x) - f(x_0)}{x - x_0} : 0 < |x - x_0| < \delta \right\}.$$

Exercise 5.12. Show that (assuming all these numbers are finite)

$$\underline{D}[f(x) + g(x)] \geq \underline{D}[f(x)] + \underline{D}[g(x)] \quad \text{and} \quad \overline{D}[f(x) + g(x)] \leq \overline{D}[f(x)] + \overline{D}[g(x)].$$

Exercise 5.13. Show that

$$-lip_f(x) \leq \underline{D}f(x) \leq \overline{D}f(x) \leq lip_f(x) \quad \text{and} \quad lip_f(x) = \max\{|\underline{D}f(x)|, |\overline{D}f(x)|\}.$$

Exercise 5.14. If f is differentiable at x show that $lip_f(x) = |f'(x)|$.

5.4. Growth on a set

The growth properties from derivative statements are similar to those for the Lipschitz numbers (cf. Lemma 5.6) and similarly proved.

Lemma 5.15. Let $F : \mathbb{R} \to \mathbb{R}$ and suppose that

$$-M < \underline{D}F(x) \leq \underline{D}F(x) < M$$

for all x in a set E. Then

$$V^*(\Delta F, E) \leq M\mathcal{L}^*(E).$$

Corollary 5.16. If $F : \mathbb{R} \to \mathbb{R}$ has a zero derivative everywhere on a set E then F does not grow on E.

Corollary 5.17. If $F : \mathbb{R} \to \mathbb{R}$ has a finite derivative everywhere on a set E then F does not grow on any null subset of E. In particular F is continuous at every point of E.

Corollary 5.18. If $F : \mathbb{R} \to \mathbb{R}$ has a finite derivative everywhere excepting a null set and if F does not grow on that null set, then F is absolutely continuous.

Corollary 5.19. If $F : \mathbb{R} \to \mathbb{R}$ has a bounded derivative everywhere on a set E then F has finite variation on E.

5.5. Growth on an interval

The derivative monitors the growth of a function f on an interval $[a,b]$. Most introductory calculus courses will discuss a result similar to this:

> *Suppose that the function $f : [a,b] \to \mathbb{R}$ is differentiable and that, at every point x in $[a,b]$,*
> $$r < f'(x) < s.$$
> *Then*
> $$r(b-a) < f(b) - f(a) < s(b-a).$$

This follows from the mean-value theorem (see Section 5.7 below) and is a typical application of that theorem. It is typical because the demand that the derivative be known at *every* point is an unfortunate feature of the mean-value theorem.

In our setting the more appropriate version would be as follows; it cannot use the mean-value theorem and is proved using our standard covering arguments. We shall study and prove several variants of this statement.

> *Suppose that the function $f : [a,b] \to \mathbb{R}$ satisfies the following inequalities at every point x in $[a,b]$:*
> $$r < \underline{D}f(x) \leq \overline{D}f(x) < s.$$
> *Then*
> $$r(b-a) < f(b) - f(a) < s(b-a).$$

In the variants we shall permit an exceptional set of points to appear at which we do not have a derivative estimate. We must then add an hypothesis to the effect that the function F can not grow on such a set.

5.6. Growth using upper or lower derivates

THEOREM 5.20. Let $f : [a,b] \to \mathbb{R}$ and suppose that $\underline{D}f(x) > m$ for each x in a compact interval $[a,b]$. Then,

$$f(b) - f(a) > m(b-a).$$

Proof. The proof is an easy application of our covering lemmas (e.g., see Lemma 5.5). Simply consider the covering relation

$$\beta = \{(I,x) : \Delta f(I) > m\mathcal{L}(I)\}.$$

■

5.6.1. Derivate assumptions with exceptional sets.

A basic principle in calculus regarding the growth of functions is this: a set of points E where the derivative or derivate is unknown can be introduced into a growth theorem provided that the function does not grow on E.

The following theorem is typical and can be made the model for many others of this type. The proof is a straightforward application of our covering arguments.

THEOREM 5.21. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $\underline{D}f(x) \geq m$ for each x in a compact interval $[c,d]$ with the exception of points x in a set E. Assume that f does not grow on E. Then

$$f(d) - f(c) \geq m(d-c).$$

EXERCISE 5.22. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and suppose that $\underline{D}f(x) \geq m$ for each x in a compact interval $[c, d]$ with countably many exceptions. Then $f(d) - f(c) \geq m(d - c)$.

EXERCISE 5.23. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $\underline{D}f(x) \geq m$ for every x in $[a, b]$ excepting possibly a null set, but that $\underline{D}f(x) > -\infty$. for every point x. Show that

$$f(b) - f(a) \geq m(b - a).$$

EXERCISE 5.24. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Suppose that $\underline{D}f(x) \geq m$ for every x in $[a, b]$ excepting possibly a null set, but that $\underline{D}f(x) > -\infty$ for every x there with at most countably many exceptions. Show that

$$f(b) - f(a) \geq m(b - a).$$

5.7. Mean values

Some of the connections between the average growth of a function and its derivative are expressed in the following theorems. The hypothesis, that f is differentiable everywhere in (a, b), is too severe for most situations and we make only occasional use of these ideas.

THEOREM 5.25. Let $f : [a, b] \to \mathbb{R}$ be continuous and also differentiable at each point inside the open interval (a, b). Suppose that $\Delta f([a, b]) = 0$. Then there is at least one point c in (a, b) at which $f'(c) = 0$.

Proof. We can assume that $f(a) = f(b) = 0$. If f is constant then certainly any point c in (a, b) would satisfy $f'(c) = 0$. If f is not constant then, Theorem 4.30 asserts that f maps $[a, b]$ into some compact interval $[A, B]$. If $A < 0$ then there is a point $c \in (a, b)$ for which $f(c) = A$ and all values are larger or equal to A. At such a point c we can have neither $f'(c) > 0$ nor $f'(c) < 0$ which only allows that $f'(c) = 0$. If $A = 0$ then certainly $B > 0$ and a similar argument must give a point $c \in (a, b)$ for which $f(c) = B$ and all values are smaller or equal to A. Again this gives us $f'(c) = 0$. ■

THEOREM 5.26 (Mean-value theorem). Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable at each point inside the open interval (a, b). Then there is at least one point c in (a, b) at which

$$f'(c) = \frac{\Delta f([a, b])}{\mathcal{L}([a, b])}.$$

Proof. Let

$$k = \frac{\Delta f([a, b])}{\mathcal{L}([a, b])}$$

and define a new function

$$F(x) = f(x) - f(a) - k(x - a).$$

Note that $F(a) = F(b) = 0$. Also $\Delta F = \Delta f - k\mathcal{L}$, from which it is easy to deduce that the continuity of f implies the continuity of F. This function F satisfies the hypotheses of (5.25) so that we are assured the existence of a point c in (a, b) at which $F'(c) = 0$. A direct computation shows that this point has $f'(c) = k$ as desired. ■

5.7.1. Darboux property of derivatives.

THEOREM 5.7.1 (Darboux property of derivatives). *Suppose that* $f : \mathbb{R} \to \mathbb{R}$ *is differentiable at every point in an open interval* (a, b) *and that* $[c, d] \subset (a, b)$*. If* T *lies between* $f'(c)$ *and* $f'(d)$ *then there is necessarily a point* $t \in (c, d)$ *for which* $f'(t) = T$.

Proof. Let $g(x) = f(x) - Tx$ and suppose that g' is never zero. Then Theorem 5.25 shows that every level set

$$L_z = \{x \in [c, d] : g(x) = g(z)\}$$

contains only the one point z. The Darboux property of the continuous function g will supply a contradiction. ∎

CHAPTER 6

The Integral

The original version of the integral, as conceived by Newton, was as an antiderivative. If $F : [a, b] \to \mathbb{R}$ is differentiable at every point of a compact interval $[a, b]$ and $F'(x) = f(x)$ then we could write

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

Such an integral would be merely descriptive. The class of functions f that would have such a property is mysterious and hard to describe in any other way. The integral is also too limited if applied only to exact derivatives.

We begin with a Newton type definition of the integral and then find a better way to express this integral, all in the language of covering relations.

6.1. Descriptive characterization of the integral

That integral that we do seek can be described by Newton's procedure using the language developed so far.

Definition 6.1. Suppose that F, $f : [a, b] \to \mathbb{R}$ are functions with the following properties:

(a) there is a null set N,
(b) $F'(x) = f(x)$ for all x in $[a, b]$ with the exception possibly of points x in N, and
(c) F does not grow on N.

Then F is said to be an *indefinite integral* for f on $[a, b]$ and we write

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

In view of Corollary 5.18 this is equivalent to the following version which explicitly uses the class of absolutely continuous functions for indefinite integrals, a fact only implied by the first version.

Definition 6.2 (Alternative). Suppose that F, $f : [a, b] \to \mathbb{R}$ are functions with the following properties:

(a) there is a null set N,
(b) $F'(x) = f(x)$ for all x in $[a, b]$ with the exception possibly of points x in N, and
(c) F is absolutely continuous.

Then F is said to be an *indefinite integral* for f on $[a, b]$ and we write

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

The only justification needed for either definition is that if two functions F, $G : [a,b] \to \mathbb{R}$ both satisfy the requirements to be an indefinite integral of f then necessarily

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

The methods of the previous chapter applied to the function $H(x) = F(x) - G(x)$ easily supply this.

The rest of the chapter is devoted to obtaining a more constructive definition for this integral that the Newton process merely describes. The problem with this description is that it is difficult to work with. For example what functions f will have an indefinite integral? This is not easy to answer yet, other than to say "those functions that are derivatives of some other function."

6.2. The relation between a function and its indefinite integral

The relation between a function F and its derivative can be expressed as a covering relation as follows:

Lemma 6.3. Let $F : [a,b] \to \mathbb{R}$ be a differentiable function with $F'(x) = f(x)$ everywhere. Then, for every $\epsilon > 0$ there is a Cousin cover β of $[a,b]$ such that, for every partition π of $[a,b]$ contained in β,

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - [F(b) - F(a)] \right| < \epsilon. \tag{6.1}$$

Lemma 6.4 (Henstock criterion). Let $F : [a,b] \to \mathbb{R}$ be a differentiable function with $F'(x) = f(x)$ everywhere. Then, for every $\epsilon > 0$ there is a Cousin cover β of $[a,b]$ such that, for every partition π of $[a,b]$ contained in β,

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon. \tag{6.2}$$

The proofs of these two lemmas are completely trivial. Simply construct the covering relation

$$\beta = \left\{ (I,x) : \frac{|\Delta F(I) - f(x)\mathcal{L}(I)|}{\mathcal{L}(I)} < \frac{\epsilon}{b-a} \right\}.$$

These lemmas give us a serious clue how to design an integral that inverts all derivatives as Newton intended. The first inequality (6.1) suggests that estimates of the form

$$\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I)$$

taken over partitions from Cousin covers of $[a,b]$ will recover $F(b) - F(a)$ from f. The second inequality (6.2) will reappear in the guise of an integrability criterion in Section 9.5.

6.3. Upper and lower integrals

The discussion above motivates the following definition of what is now considered the *correct* calculus integral.

DEFINITION 6.5. For a function $f : [a, b] \to \mathbb{R}$ we define an upper integral by

$$\overline{\int_a^b} f(x)\,dx = \inf_{\beta} \sup_{\pi \subset \beta} \sum_{(I,z)\in\pi} f(z)\mathcal{L}(I)$$

where the supremum is taken over all partitions π of $[a, b]$ contained in β, and the infima over all Cousin covers β of $[a, b]$.

Similarly we define a lower integral, either by writing

$$\underline{\int_a^b} f(x)\,dx = -\overline{\int_a^b} [-f(x)]\,dx,$$

or directly as

$$\underline{\int_a^b} f(x)\,dx = \sup_{\beta} \inf_{\pi \subset \beta} \sum_{(I,z)\in\pi} f(z)\mathcal{L}(I)$$

where, again, π is a partition of $[a, b]$ and β is a Cousin cover of $[a, b]$.

EXERCISE 6.6. Let $f : \mathbb{R} \to \mathbb{R}$. Show that

$$\underline{\int_a^b} f(x)\,dx \le \overline{\int_a^b} f(x)\,dx.$$

Hint: Make use in your proof of the fact that the intersection of two Cousin covers, is again a Cousin cover.

EXERCISE 6.7. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $a < b < c$. Show that

$$\overline{\int_a^c} f(x)\,dx = \overline{\int_a^b} f(x)\,dx + \overline{\int_b^c} f(x)\,dx,$$

assuming the sum makes sense.

Hint: Infinite values are allowed but we would have to avoid $\infty + (-\infty)$ or $-\infty + \infty$. This is simpler if you first check that a single value $f(b)$ is irrelevant to the computations so that you may assume that $f(b) = 0$. Then ensure that any partition π contained in your choice of β of the interval $[a, b]$, $[a, c]$ or $[b, c]$ would have to contain an element (I, b).

6.3.1. The integral and integrable functions. If the upper and lower integrals are identical we write the common value as

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

allowing finite or infinite values. If this value is also finite then we say f is *integrable* and

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

is called simply the *integral*.

EXERCISE 6.8. Let $f : [a, b] \to \mathbb{R}$ show that a sufficient condition for f to be integrable on $[a, b]$ with $c = \int_a^b f(x)\,dx$ is that for every $\epsilon > 0$ there is a Cousin cover β of $[a, b]$ so that

$$\left| c - \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \right| < \epsilon$$

for every partition π of $[a,b]$ contained in β.

Hint: Use β to find estimates for the upper and lower integrals. (Later we will show that this condition is, in fact, both necessary and sufficient.)

6.4. Integration of derivatives

The descriptive version of the integral given above in Section 6.1 serves exactly to characterize the integral. We can, at this early stage prove this only in one direction.

THEOREM 6.9. Suppose that F, $f : [a,b] \to \mathbb{R}$ and that F is an indefinite integral for f on $[a,b]$. Then f is integrable on $[a,b]$ and

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

Note that as a result of this theorem the integral inverts derivatives in each of the following cases:

(a) F is continuous and $F'(x) = f(x)$ at every point in (a,b).
(b) F is continuous and $F'(x) = f(x)$ with countably many exceptions.
(c) F is Lipschitz and $F'(x) = f(x)$ with the exception only of a null set.
(d) F is absolutely continuous and $F'(x) = f(x)$ with the exception only of a null set.
(e) $F'(x) = f(x)$ with the exception only of a null set N and F does not grow on N.

Proof. Write N for the null set of points x at which the identity $F'(x) = f(x)$ is not given. We are assuming that F is absolutely continuous [or equivalently we are assuming that F does not grow on this set N]. We will prove that f must be integrable and that the value of the integral is $F(b) - F(a)$. Define first a new function $g : [a,b] \to \mathbb{R}$ by setting $g(x) = f(x)$ for $x \in [a,b] \setminus N$ and $g(x) = 0$ for $x \in N$.

Let $\epsilon > 0$. Since F does not grow on N there is a Cousin cover β_1 (Definition 4.11) so that

$$\sum_{([u,v],w)\in\pi} |\Delta F([u,v])| < \epsilon/2 \tag{6.3}$$

whenever π is a subpartition, $\pi \subset \beta_1[N]$.

We know, too, that $F'(x) = f(x) = g(x)$ for every x in $[a,b]$ excepting possibly at points in N. Write $E = [a,b] \setminus N$. Thus there is a Cousin cover β_2 of $[a,b]$ that includes all of the following collection:

$$\{(I,x) : x \in E,\ |\Delta F(I) - f(x)\mathcal{L}(I)| \leq \epsilon\mathcal{L}(I)/(2[b-a]).\} \tag{6.4}$$

Check that

$$\beta = \beta_1[E] \cup \beta_2[N]$$

is a Cousin cover of $[a,b]$.

Take any partition π from β. Note that π decomposes into the two subpartitions $\pi[N]$ and $\pi[E]$. Check that

$$\sum_{(I,x)\in\pi[N]} |\Delta F(I)| < \epsilon/2.$$

This allows the computation,

$$\left|[F(b)-F(a)]-\sum_{(I,x)\in\pi} g(x)\mathcal{L}(I)\right| \leq \sum_{(I,x)\in\pi} |\Delta F(I)-g(x)\mathcal{L}(I)|$$

$$\leq \sum_{(I,x)\in\pi[N]} |\Delta F(I)| + \sum_{(I,x)\in\pi[E]} |\Delta F(I)-g(x)\mathcal{L}(I)| < \epsilon.$$

Consequently g is integrable on $[a,b]$ and

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

The remainder of the proof requires us to establish that if g is integrable then so too is f and that the integrals coincide. This is proved later on in Section 6.9.1 and can be left to the reader to assemble here. ■

6.5. Estimates from Cauchy sums

The integral $\int_a^b f(x)\,dx$ of a function f is defined directly as a kind of limiting process that exploits the computation

$$\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I)$$

of the Cauchy sums for a function f taken over partitions π of the interval $[a,b]$. But the definition must be considered purely formal since, in general, the Cousin covers arising in the definition will not be constructible. Can these Cauchy sums, nonetheless, be used to estimate the integral? The answer is yes in very limited situations; for monotonic functions or continuous functions we can illustrate how these might be used.

Lemma 6.10. Let $f:[a,b]\to\mathbb{R}$ be an integrable function and let π be any partition of $[a,b]$. Then

$$\left|\int_a^b f(x)\,dx - \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I)\right| \leq \sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I).$$

Proof. Let $\epsilon>0$ and, using Exercise 6.8, choose a Cousin cover β of $[a,b]$ so that

$$\left|\int_a^b f(x)\,dx - \sum_{(J,t)\in\pi'} f(t)\mathcal{L}(J)\right| < \epsilon$$

for every partition π' of $[a,b]$ taken from β. Choose any fixed $(I,x)\in\pi$ and any partition π' of I. Note that

$$\left|f(x)\mathcal{L}(I) - \sum_{(J,t)\in\pi'} f(t)\mathcal{L}(J)\right| \leq \omega f(I)\mathcal{L}(I).$$

Thus if we choose a partition π' of $[a,b]$ from β in such a way that π' contains a partition of each interval I for which $(I,x) \in \pi$, we must have

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \sum_{(J,t)\in\pi'} f(t)\mathcal{L}(J) \right| \leq \sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I).$$

Combining this inequality with what we know for any such partition π' chosen from β, we obtain

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \int_a^b f(x)\,dx \right| \leq \epsilon + \sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I).$$

As ϵ is an arbitrary positive number the inequality of the lemma follows. ■

Lemma 6.10 is limited in usefulness to bounded functions. Here is a variant that might be used if the function is unbounded at one of the endpoints of the interval.

Exercise 6.11. Let $f : [a,b] \to \mathbb{R}$ be an integrable function and let $\{x_n\}$ be a decreasing sequence of real numbers with $x_1 = b$ and $\lim_{n\to\infty} x_n = a$. Suppose that

$$\sum_{n=1}^{\infty} \omega f([x_{n+1}, x_n])\mathcal{L}([x_{n+1}, x_n]) < \infty.$$

Show that the series

$$\sum_{n=1}^{\infty} f(x_{n+1})\mathcal{L}([x_{n+1}, x_n])$$

converges and that

$$\left| \int_a^b f(x)\,dx - \sum_{n=1}^{\infty} f(x_{n+1})\mathcal{L}([x_{n+1}, x_n]) \right| \leq \sum_{n=1}^{\infty} \omega f([x_{n+1}, x_n])\mathcal{L}([x_{n+1}, x_n]).$$

6.6. Estimates of integrals from derivates

Lemma 6.12. Let F, $f : [a,b] \to \mathbb{R}$. If F is continuous at a and b and

$$\underline{D}F(x) \geq f(x)$$

at every point of (a,b), then

$$\overline{\int_a^b} f(x)\,dx \leq F(b) - F(a).$$

Proof. Let $\epsilon > 0$. Take the covering relation

$$\beta_1 = \{(I,x) : \Delta F(I) \geq (f(x) - \epsilon)\mathcal{L}(I)\}$$

and

$$\beta_2 = \{(I,x) : x = a \text{ or } b,\ x \in I \text{ and } |\Delta F(I)| + |f(x)|\mathcal{L}(I) < \epsilon\}.$$

Check that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a,b]$. At the endpoints a or b the continuity of F needs to be used in the verification, while at the points in (a,b) the inequality $\underline{D}F(x) \geq f(x)$ is used.

Any partition $\pi \subset \beta$ of the interval $[a,b]$ will satisfy

$$\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \leq \sum_{(I,x)\in\pi} [\Delta F(I) + \epsilon\mathcal{L}(I)] + 2\epsilon = F(b) - F(a) + \epsilon(2 + b - a).$$

This is true for all partitions π from this β and all $\epsilon > 0$ and so the conclusion that

$$\underline{\int_a^b} f(x)\,dx \leq F(b) - F(a)$$

now follows. ∎

LEMMA 6.13. Let F, $f : [a,b] \to \mathbb{R}$. If F is continuous at a and b and

$$\overline{D}F(x) \leq f(x)$$

at every point of (a,b), then

$$\underline{\int_a^b} f(x)\,dx \geq F(b) - F(a).$$

Proof. Apply Lemma 6.12 to the functions $-F$ and $-f$. ∎

6.7. Elementary properties of the integral

6.7.1. Integration and order.

THEOREM 6.14. Suppose that f, $g : [a,b] \to \mathbb{R}$ are both integrable and that $f(x) \leq g(x)$ for each x in that interval. Then

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

Proof. This follows easily from the inequality

$$\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \leq \sum_{(I,x)\in\pi} g(x)\mathcal{L}(I)$$

which would be true for any partition π of the interval $[a,b]$. ∎

6.7.2. Integration of linear combinations.

THEOREM 6.15. Suppose that f, $g : [a,b] \to \mathbb{R}$ are both integrable. Then so too is any linear combination $rf + sg$ and

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

Proof. Use Exercise 6.8 and some simple algebra. ∎

6.7.3. The integral as an additive interval function.

THEOREM 6.16. If $f : [a,c] \to \mathbb{R}$ is integrable on each of the intervals $[a,b]$, $[b,c]$, and $[a,c]$ then

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

Proof. This follows from Exercise 6.7. ∎

6.7.4. Change of variable. Let $\phi : [a,b] \to \mathbb{R}$ be a strictly increasing differentiable function. We would expect from elementary formulas of the calculus that

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\phi'(t)\,dt.$$

If f is itself everywhere a derivative then this could be justified. If f is assumed only to be integrable then a different proof, using ϕ to map Cousin covers and partitions, is needed.

THEOREM 6.17 (Change of variable). *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing, differentiable function. If $f : \mathbb{R} \to \mathbb{R}$ is integrable on $[\phi(a), \phi(b)]$ then*

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\phi'(t)\,dt.$$

Proof. Let $\epsilon > 0$ and define the covering relation

$$\beta_1 = \left\{ (I,t) : \left| \frac{\Delta\phi(I)}{\mathcal{L}(I)} - \phi'(t) \right| < \frac{\epsilon}{2(b-a)|(1+|f(\phi(t))|)} \right\}.$$

Since ϕ is everywhere differentiable on $[a,b]$ this is a Cousin cover of $[a,b]$. Note that we can write $\Delta\phi(I)$ also as $\mathcal{L}(J)$ where $J = \phi(I)$ is just the compact interval that ϕ maps I onto.

Write

$$\beta_1' = \{(\phi(I), \phi(x)) : (I,x) \in \beta_1\}$$

and check that β_1' is a Cousin cover of the compact interval $\phi([a,b]) = [\phi(a), \phi(b)]$. Observe that elements $(J,x) = (\phi(I), \phi(t))$ of β_1' must satisfy

$$|f(\phi(x))\mathcal{L}(\phi(I)) - f(\phi(x))\phi'(x)\mathcal{L}(I)| < \epsilon\mathcal{L}(I)/2(b-a).$$

The expression $f(\phi(t))\mathcal{L}(\phi(I))$ here is better viewed as $f(x)\mathcal{L}(J)$.

Choose a Cousin cover β_2' of $[\phi(a), \phi(b)]$ so that

$$\left| \int_{\phi(a)}^{\phi(b)} f(x)\,dx - \sum_{(J,x)\in\pi'} f(x)\mathcal{L}(J) \right| < \epsilon/2$$

for all partitions $\pi' \subset \beta_2'$ of the interval $[\phi(a), \phi(b)]$. Write β_2 for the collection of all (I,x) for which $(I,x) = (\phi(J), \phi(t))$ for some $(J,t) \in \beta_2'$. This is a Cousin cover of $[a,b]$.

Write $\beta = \beta_1 \cap \beta_2$. Check that β is a Cousin cover of $[a,b]$ and check that

$$\left| \int_{\phi(a)}^{\phi(b)} f(x)\,dx - \sum_{(I,x)\in\pi} f(\phi(x))\phi'(x)\mathcal{L}(I) \right| < \epsilon$$

for all partitions $\pi \subset \beta$ of the interval $[a,b]$. ■

6.8. Summing inside the integral

For a great many calculus applications we would 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 many simple hypotheses.

6.8.1. Two lemmas.

LEMMA 6.18. Suppose that f, f_1, f_2, ... is a sequence of nonnegative functions defined on a compact interval $[a, b]$. If, for each x

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

then

$$\underline{\int_a^b} f(x)\,dx \geq \sum_{n=1}^{\infty} \left(\underline{\int_a^b} f_n(x)\,dx \right). \tag{6.5}$$

Proof. Let $\epsilon > 0$. Take any integer N and choose Cousin covers β_n of the interval $[a, b]$ ($n = 1, 2, \dots, N$) so that all sums

$$\sum_{\pi} f_n(x)\mathcal{L}(I) \geq \underline{\int_a^b} f_n(x)\,dx - \epsilon 2^{-n}$$

whenever $\pi \subset \beta_n$ is a partition of $[a, b]$. (If the integrals here are not finite then there is nothing to prove, since both sides of the inequality (6.5) will be infinite.)

Let

$$\beta = \bigcap_{n=1}^{N} \beta_n.$$

By Exercise 3.5 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

$$\sum_{\pi} f(x)\mathcal{L}(I) \geq \sum_{\pi} \left(\sum_{n=1}^{N} f_n(x)\mathcal{L}(I) \right) = \sum_{n=1}^{N} \left(\sum_{\pi} f_n(x)\mathcal{L}(I) \right) \geq$$
$$\sum_{n=1}^{N} \left(\underline{\int_a^b} f_n(x)\,dx - \epsilon 2^{-n} \right).$$

This gives a lower bound for all Cauchy sums and hence, since ϵ is arbitrary, shows that

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

As this is true for all N the inequality (6.5) must follow. ∎

LEMMA 6.19. Suppose that f, f_1, f_2, ... is a sequence of nonnegative functions defined on a compact interval $[a, b]$. If, for each x

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

then

$$\overline{\int_a^b} f(x)\,dx \leq \sum_{n=1}^{\infty} \left(\overline{\int_a^b} f_n(x)\,dx \right). \tag{6.6}$$

Proof. This lemma is similar to the preceding one, but requires a bit of bookkeeping and a new technique with the covers. Let $t < 1$ and choose for each $x \in [a, b]$ the first integer $N(x)$ so that

$$tf(x) \leq \sum_{n=1}^{N(x)} f_n(x).$$

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

$$\sum_{\pi} f_n(x)\mathcal{L}(I) \leq \overline{\int_a^b} f_n(x)\,dx + \epsilon 2^{-n}$$

whenever $\pi \subset \beta_n$ is a partition of $[a, b]$. (Again, if the integrals here are not finite then there is nothing to prove, since the larger side of the inequality (6.6) will be infinite.)

Let

$$E_n = \{x \in [a, b] : N(x) = n\}.$$

We use these sets to carve up the covering relations. Write

$$\beta_n[E_n] = \{(I, x) \in \beta_n : x \in E_n\}$$

and

$$\beta = \bigcup_{n=1}^{\infty} \beta_n[E_n].$$

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 $j = 1, 2, 3, \dots, N$,

$$\pi_j = \{(I, x) \in \pi : x \in E_j\}$$

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

$$\pi = \pi_1 \cup \pi_2 \cup \dots \cup \pi_N.$$

Check the following computations, making sure to use the fact that for $x \in E_i$,

$$tf(x) \leq f_1(x) + f_2(x) + \dots + f_i(x).$$

$$\sum_{\pi} tf(x)\mathcal{L}(I) = \sum_{i=1}^{N}\sum_{\pi_i} tf(x)\mathcal{L}(I) \leq \sum_{i=1}^{N}\sum_{\pi_i} \left(f_1(x) + f_2(x) + \dots + f_i(x)\right)\mathcal{L}(I)$$

$$= \sum_{j=1}^{N}\left(\sum_{\sigma_j} f_j(x)\mathcal{L}(I)\right) \leq$$

$$\sum_{j=1}^{N}\left(\overline{\int_a^b} f_j(x)\,dx + \epsilon 2^{-j}\right) \leq \sum_{j=1}^{\infty}\left(\overline{\int_a^b} f_j(x)\,dx\right) + \epsilon.$$

This gives an upper bound for all Cauchy sums and hence, since ϵ is arbitrary, shows that

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

As this is true for all $t < 1$ the inequality (6.6) must follow too. ■

6.8.2. Summing inside the integral.

THEOREM 6.20. Let $f_n : [a,b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ be a sequence of nonnegative integrable functions and suppose that

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

for every x. Then

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

Proof. This follows from Lemmas 6.18 and 6.19. ■

6.8.3. Monotone convergence theorem.

THEOREM 6.21 (Monotone convergence theorem). Let $f_n : [a,b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ be a nondecreasing sequence of integrable functions and suppose that

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

for every x. Then

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

Proof. This follows directly from Theorem 6.20 and the identity

$$f(x) = f_1(x) + \sum_{n=1}^{\infty} \left(f_n(x) - f_{n-1}(x) \right).$$

■

6.9. Integration of Null functions

A function $f : [a,b] \to \mathbb{R}$ is a null function if it assumes the value zero everwhere with the possible exception of a null set. The null functions play a special role in integration theory since they always integrate to zero.

THEOREM 6.22. Let $f : [a,b] \to \mathbb{R}$ be a null function. Then f is integrable on $[a,b]$ and

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

Proof. Suppose that f is bounded, say $|f(x)| < M$ for all x and that $f(x) = 0$ for all x not in an open set G. It is easy to check that

$$\overline{\int_a^b} |f(x)|\,dx < M\mathcal{L}(G).$$

From that inequality the student can easily deduce that all bounded, null functions are integrable and with a zero integral.

The extension to unbounded functions uses an important device. For integers $m = 1, 2, 3, \dots$, write

$$E_m = \{x \in [a,b] : m - 1 < |f(x)| \leq m\}.$$

Each set E_m is a null set and f is bounded on each E_m. An application of Lemma 6.18 reveals that

$$\overline{\int_a^b} |f(x)|\,dx \leq \sum_{i=1}^{\infty} \overline{\int_a^b} |\chi_{E_i}(x) f(x)|\,dx. \tag{6.7}$$

As each integrand in the sum on the right is a bounded, null function it follows that each term in the sum is zero. From this we easily obtain that f is integrable on $[a,b]$ with a zero integral, completing the proof. ■

6.9.1. Negligible sets. Suppose that $f_1,\ f_2 : [a,b] \to \mathbb{R}$. Under what circumstances can we claim that f_1 and f_2 have the same integrability behaviour and that

$$\int_c^d f_1(x)\,dx = \int_c^d f_1(x)\,dx$$

for all compact subintervals $[c,d]$ of $[a,b]$?

If $f_1 - f_2$ is a null function then, certainly, Theorem 6.22 can be used to verify that f_1 is integrable if and only if f_2 is integrable and that, in that case

$$\int_c^d f_1(x)\,dx = \int_c^d f_1(x)\,dx.$$

This suggests that the following definition will be useful in integration applications:

Definition 6.23. Two functions $f_1,\ f_2 : [a,b] \to \mathbb{R}$ are *Lebesgue equivalent* on $[a,b]$ if $f_1 - f_2$ is a null function, i.e., if $f_1(x) = f_2(x)$ for almost every x in $[a,b]$.

Essentially we may ignore a null set in discussing the integrability of a function or its integral. For example this would allow us to ask about the integrability of a function such as

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

on the interval $[0,1]$ even though $f(x)$ fails to be defined at $x = 0$. We would simply use any other function g that is defined everywhere and agrees with f almost everywhere.

CHAPTER 7

Littlewood's Three Principles

The famous British mathematician J. E. Littlewood once remarked that the techniques of measure theory as they are used in a theoretical calculus development are really very simple:

(a) All sets that appear are almost closed.
(b) All functions that appear are almost continuous.
(c) All convergence of sequences or series of functions is almost uniform.

In this chapter we discuss these principles of Littlewood.

As the calculus program developed in the 19th century it became clear that a special analytic role was played by the continuous functions, the closed sets, and by uniform convergence. The introduction of the measure theory allowed the same arguments to be applied in an approximate version: one removes an open set G for which the measure $\mathcal{L}(G)$ is small.

We shall say some property or concept *almost* holds if we have our property or concept after the exclusion of an open set G which has arbitrarily small measure. To exploit this properly requires just one device:

Let $G_1, G_2, G_3, \ldots$ be a sequence of open sets with $\mathcal{L}(G_k) < \epsilon 2^{-k}$.
Then $G = \bigcup_{k=1}^{\infty} G_k$ is an open set for which $\mathcal{L}(G) < \epsilon$.

This allows a sequence of almost-arguments (even an infinite sequence) to culminate in one single one.

7.1. Null sets

A set is a *null set* if it is almost empty.

Definition 7.1. A set $N \subset \mathbb{R}$ is said to be a *null set* if for every $\epsilon > 0$ there is an open set G for which $\mathcal{L}(G) < \epsilon$ and $E \setminus G = \emptyset$ (i.e., $E \subset G$).

Theorem 7.1.1. *A set E is a null set if and only if $\mathcal{L}(E) = 0$.*

Proof. Certainly it is clear from the definition that if N is null then $\mathcal{L}(N) < \epsilon$ for every $\epsilon > 0$. Consequently $\mathcal{L}(N) = 0$. But conversely, if $\mathcal{L}(N) = 0$ then by definition

$$\inf\{\mathcal{L}(G) : \ G \text{ open}, N \subset G\} = \mathcal{L}(N) = 0$$

which proves that N is almost empty. ∎

The following properties are obvious consequences of the properties of the measure $\mathcal{L}$.

Theorem 7.2. The class of null sets has the following properties:

(a) $\emptyset$ is null.
(b) If $E_1 \subset E$ and E is null so too is E_1.

(c) If E_1, E_2, E_3, ... is a sequence of null sets then $E = \bigcup_{n=1}^{\infty} E_n$ is also null.

Exercise 7.3. Show that every countable set is a null set.

Exercise 7.4. Show that an open set G is a null set if and only if $G = \emptyset$.

7.2. Null functions

A *null function* is a function that is almost the zero function.

Definition 7.5. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be a *null function* if for every $\epsilon > 0$ there is an open set G with $\mathcal{L}(G) < \epsilon$ so that $f(x) = 0$ for all $x \notin G$.

Exercise 7.6. Show that $f : \mathbb{R} \to \mathbb{R}$ is a null function if and only if the set $\{x : f(x) \neq 0\}$ is a null set.

7.3. Almost everywhere

A property is said to hold *everywhere* on a set E if that property holds at every point $x \in E$. The *almost everywhere* version is this:

Definition 7.7. A property is said to hold *almost everywhere* on a set E if for every $\epsilon > 0$ there is an open set G for which $\mathcal{L}(G) < \epsilon$ and that property holds at every point $x \in E \setminus G$.

Theorem 7.8. The following are equivalent:

(a) A property holds almost everywhere on a set E.
(b) The set of points in E at which that property does not hold is a null set.

Proof. Suppose that this property holds almost everywhere on a set E. For each integer n choose an open set G_n so that the property holds for all x in $E \setminus G_n$ where $\mathcal{L}(G_n) < 1/n$. Let

$$N = \bigcap_{n=1}^{\infty} G_n.$$

Observe that the property holds for every x in E that is not also in N. But N is a null set since, given any $\epsilon > 0$ there is an integer n with $\mathcal{L}(G_n) < \epsilon$ and $N \subset G_n$.

Conversely suppose that the set of points N in E at which that property does not hold is a null set. Then for any $\epsilon > 0$ there is an open set G with $\mathcal{L}(G) < \epsilon$ and $N \subset G$. The property holds, by assumption, on $E \setminus N$ and consequently also holds on $E \setminus G$. This proves that our property holds almost everywhere on E. ■

7.4. Almost closed sets

A set is *almost closed* if removing a small open set makes it closed. All sets that appear in the calculus will prove to be almost closed.

Definition 7.9. A set E is said to be *almost closed* if for every $\epsilon > 0$ there is an open set G with $\mathcal{L}(G) < \epsilon$ so that $E \setminus G$ is closed.

Exercise 7.10. Show that every null set is almost closed.

Exercise 7.11. Show that every open set is almost closed.

Exercise 7.12. Show that E is almost closed if and only if, for every $\epsilon > 0$, there is an open set G with $\mathcal{L}(G) < \epsilon$ so that $E \setminus G$ is *almost* closed.

7.5. Properties of almost closed sets

The closed sets have a number of properties of interest and importance in calculus arguments. Let us recall that

(a) If E_1, E_2, E_3, is a sequence of closed sets then the intersection $\bigcap_{n=1}^{\infty} E_n$ is closed.
(b) If E_1, E_2, E_3, $\ldots E_N$ is a finite collection of closed sets then the union $\bigcup_{n=1}^{N} E_n$ is closed.
(c) The complement of a closed set is open.

The almost closed sets have stronger properties than these.

THEOREM 7.13. The family of all almost closed sets has the following properties:

(a) Every null set is almost closed.
(b) Every closed set is almost closed.
(c) If E_1, E_2, E_3, is a sequence of almost closed sets then the union $\bigcup_{n=1}^{\infty} E_n$ is also almost closed.
(d) If E_1, E_2, E_3, is a sequence of almost closed sets then the intersection $\bigcap_{n=1}^{\infty} E_n$ is also almost closed.
(e) If E is almost closed then the complement $\mathbb{R} \setminus E$ is also almost closed.

Proof. Items (a) and (b) are easy. Let us prove (e) first. Let E be almost closed 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.

Use Exercise 7.11 to find an open set G_2 with $\mathcal{L}(G_2) < \epsilon/2$ so that $O \setminus G_2$ is closed. 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 almost closed.

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 almost closed by (f) and hence B is almost closed by (d). The complement of B, namely the set A, is almost closed by (f) again. ■

7.6. Measure properties of derivates

There are many connections between the measure theory and the derivatives. We will need to know that certain sets arising in the study of derivatives are almost closed. That will allow us to make use of them in proving some almost everywhere statements later on in the text.

LEMMA 7.14. For any function $f : \mathbb{R} \to \mathbb{R}$ and any real number r the sets

$$\{x : \overline{D}f(x) < r\} \text{ and } \{x : \underline{D}f(x) < r\}$$

are almost closed.

Proof. Let us address the set $E = \{x : \overline{D}f(x) > r\}$ for any r. Let m, n be integers and define $\mathcal{C}_{mn}$ to be the collection of all compact intervals I for which $\mathcal{L}(I) < 1/m$ and $\Delta f(I) \geq r + 1/n$. Write $E_{mn} = \bigcup\{I : I \in \mathcal{C}_{mn}\}$.

We already know (because of Lemma 2.31) that E_{mn} is an almost closed set. To complete the proof we check that

$$E = \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} E_{mn}. \tag{7.1}$$

To verify this we use the fact that any point x for which $\overline{D}f(x) > r$ there must be at least one integer n with $\overline{D}f(x) \geq r + 1/n$; and, moreover, for every integer m there would have to be at least one compact interval I containing x with length less than $1/m$ so that $\Delta f(I) \geq r + 1/n$. Conversely, should these statements hold, it would follow that $\overline{D}f(x) > r$ and so $x \in E$.

The identity (7.1) exhibits E as a combination of sequences of almost closed sets and so E too is an almost closed set because of Theorem 7.13. ■

EXERCISE 7.15. Let $f : \mathbb{R} \to \mathbb{R}$. Show that the set of points where f is differentiable is an almost closed set.

Hint: Let A be the set of *all* points where f has a finite derivative. We know that each set of the form

$$\{x : \overline{D}f(x) > c\}$$

and

$$\{x : \underline{D}f(x) < c\}$$

is almost closed. Thus the set A' of points at which f does not have a derivative (finite or finite) can be expressed as the union of the family of sets

$$A_{pq} = \{x : \underline{D}f(x) < p < q < \overline{D}f(x)\}$$

for rational numbers p and q. It follows that A' is also almost closed. Again the set A'' of points where $f'(x) = \pm\infty$ can be written as

$$A'' = \bigcap_{n=1}^{\infty} \{x : \overline{D}f(x) < -n\} \cup \bigcap_{n=1}^{\infty} \{x : \underline{D}f(x) > n\}.$$

Thus this set is almost closed. But $A = A' \setminus A''$.

7.7. Measure computations with almost closed sets

If $\{E_n\}$ is a disjointed sequence of closed sets then we know that

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathcal{L}(E_n)$$

because we can separate pairs of disjoint closed sets by open sets. It is natural to expect that this same assertion would be true if the sets were merely almost closed.

THEOREM 7.16. Let $\{E_n\}$ be a disjointed sequence of almost closed sets and A an arbitrary real set. Then

$$\mathcal{L}\left(A \cap \bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathcal{L}(A \cap E_n).$$

Proof. Let $\epsilon > 0$ and choose open sets G_n with $\mathcal{L}(G_n) < \epsilon 2^{-n}$ so that $A_n = E_n \setminus G_n$ is closed. Observe that the sequence of closed sets $\{A_n\}$ is disjointed so that the members of the sequence $\{A \cap A_n\}$ are separated by open sets. Fix any integer N. Then

$$\sum_{n=1}^{N} \mathcal{L}(A \cap E_n) \leq \sum_{n=1}^{N} [\mathcal{L}(A \cap A_n) + \mathcal{L}(G_n)] \leq \mathcal{L}\left(A \cap \bigcup_{n=1}^{N} A_n\right) + \epsilon$$

$$\leq \mathcal{L}\left(\bigcup_{n=1}^{\infty} A \cap E_n\right) + \epsilon \leq \sum_{n=1}^{\infty} \mathcal{L}(A \cap E_n) + \epsilon.$$

■

EXERCISE 7.17. Suppose that $\{E_n\}$ is an increasing sequence of almost closed sets. Show that

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{L}(E_n).$$

Hint: Write $A_0 = \emptyset$ and $A_n = E_n \setminus E_{n-1}$ for each $n = 1, 2, 3, \dots$.

EXERCISE 7.18. Suppose that $\{E_n\}$ is a decreasing sequence of almost closed sets contained in a compact interval $[a, b]$. Show that

$$\mathcal{L}\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{L}(E_n).$$

Hint: Write $A_n = [a, b] \setminus E_n$ for for each $n = 1, 2, 3, \dots$ and note that $\{A_n\}$ must be an increasing sequence of sets.

7.7.1. Increasing sequences of sets. In Exercise 7.17 we have just checked that if

$$E_1 \subset E_2 \subset E_3 \subset \dots$$

is an increasing sequence of sets then

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{L}(E_n)$$

but our methods required us to assume that the sets are almost closed. By a further "almost" argument we can drop this assumption.

THEOREM 7.19. Suppose that $\{E_n\}$ is an increasing sequence of sets. Then

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{L}(E_n).$$

Proof. The proof will use Exercise 7.17. Observe first that

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} E_n\right) \geq \lim_{m\to\infty} \mathcal{L}(E_m)$$

merely because each set E_m is contained in this union.

To prove the opposite inequality, begin by choosing almost closed sets $H_n \supset E_n$ with the same measures, i.e., so that $\mathcal{L}(E_n) = \mathcal{L}(H_n)$. (For example, start with a sequence of open sets G_{nm} containing E_n with $\mathcal{L}(E_n) \leq \mathcal{L}(G_{nm}) \leq \mathcal{L}(E_n) + 1/n$ and take $H_n = \bigcap_{m=1}^{\infty} G_{nm}$.)

Write $V_m = \bigcap_{k=m}^{\infty} H_k$ and $V = \bigcup_{m=1}^{\infty} V_m$. These sets are all almost closed because we choose the $\{H_k\}$ to be almost closed. Applying Exercise 7.17, we obtain

$$\mathcal{L}(V) = \lim_{m\to\infty} \mathcal{L}(V_m).$$

But $E_m \subset V_m \subset H_m$ so that $V \supset E$ and $\mathcal{L}(E_m) = \mathcal{L}(V_m) = \mathcal{L}(H_m)$. Consequently

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \mathcal{L}(V) = \lim_{m\to\infty} \mathcal{L}(V_m) = \lim_{m\to\infty} \mathcal{L}(E_m).$$

This completes the proof. ∎

7.8. Testing for null

A set is a null set if and only if all its subsets are null sets. In particular all compact subsets K of a null set N must be null and so must satisfy $\mathcal{L}(K) = 0$.

Is the converse true? That is, if we are able to check that $\mathcal{L}(K) = 0$ for every compact $K \subset N$, then can we conclude that N is a null set? This is an important way of verifying that a set is null because we can often make estimates on the measure of its compact subsets. The test is valid only for almost closed sets.

THEOREM 7.8.1. *In order for an almost closed set E to be a null set it is necessary and sufficient that $\mathcal{L}(K) = 0$ for every compact set $K \subset E$.*

Proof. The necessity is obvious. For the sufficiency suppose that E is almost closed and satisfies the condition that all its compact subsets are null. Let $\epsilon > 0$ and choose an open set G_0 so that $\mathcal{L}(G_0) < \epsilon 2^{-n}$ and $E \setminus G_0$ is closed. For each integer n the set $[-n, n] \cap (E \setminus G_0)$ is compact and so, by hypothesis, is a null set. Thus we may choose open sets G_n so that $\mathcal{L}(G_n) < \epsilon 2^{-n-1}$ for which

$$[-n, n] \cap (E \setminus G_0) \subset G_n.$$

Let

$$G = \bigcup_{n=0}^{\infty} G_n$$

and observe that $\mathcal{L}(G) < \epsilon$ and that $E \subset G$. This verifies that E is null. ∎

7.9. Measure estimates for almost closed sets

For an arbitrary set E we have defined the measure $\mathcal{L}(E)$ using an estimate for $\mathcal{L}(E)$ from outside by measures of open sets that contain E. Could we instead estimate the measure from inside using compact sets? For almost closed sets the answer is yes.

THEOREM 7.9.1. *Let E be an almost closed set. Then*

$$\mathcal{L}(E) = \sup\{\mathcal{L}(K) : K \subset E, K \text{ compact}\}.$$

Proof. Let $\epsilon > 0$ and determine an open set G so that $\mathcal{L}(G) < \epsilon$ and $E \setminus G$ is closed. If E is bounded then take $K = E \setminus G$, note that K is compact, $K \subset E$, and $\mathcal{L}(E) \leq \mathcal{L}(K) + \mathcal{L}(G)$. Thus $\mathcal{L}(K) > \mathcal{L}(E) - \epsilon$. This would complete the proof.

If E is not bounded then write $K_n = [n, n+1] \cap (E \setminus G)$ for $n = 0, \pm 1, \pm 2, \dots$. Check that

$$\bigcup_{n=-\infty}^{\infty} K_n = (E \setminus G)$$

and that

$$\sum_{n=-\infty}^{\infty} \mathcal{L}(K_n) = \mathcal{L}(E \setminus G).$$

In particular if $t < \mathcal{L}(E)$ then

$$t < \mathcal{L}(E \setminus G) + \epsilon$$

and there must be an integer N for which

$$\sum_{n=-N}^{N} \mathcal{L}(K_n) + \epsilon > t.$$

Set

$$K' = \bigcup_{n=-N}^{N} K_n,$$

observe that K' is compact, $K' \subset E$, and

$$\mathcal{L}(K') > t - \epsilon.$$

Since t and ϵ are arbitrary we have a compact set K' inside E whose measure approximates $\mathcal{L}(E)$ as close as we please. ■

Exercise 7.20. Derive Theorem 7.8.1 from Theorem 7.9.1.

Exercise 7.21 (Continuous functions). Show that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only for every real number r the sets

$$\{x : f(x) < r\} \quad \text{and} \quad \{x : f(x) > r\}$$

are open.

Exercise 7.22 (Almost continuous functions). Show that a function $f : \mathbb{R} \to \mathbb{R}$ is almost continuous if and only for every real number r the set

$$\{x : f(x) < r\}$$

is almost closed.

7.10. Almost continuous functions

A function is almost continuous if it almost agrees with some continuous function. All functions that appear in the calculus will prove to be almost continuous.

Definition 7.23. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *almost continuous* if for every $\epsilon > 0$ there is an open set G with $\mathcal{L}(G) < \epsilon$ so that for some continuous function $g : \mathbb{R} \to \mathbb{R}$ the identity $f(x) = g(x)$ holds for all $x \notin G$.

Exercise 7.24. Suppose that the functions f, f_1, $f_2 : \mathbb{R} \to \mathbb{R}$ are almost continuous. Show that so too are the functions $|f|$, $\max\{f_1, f_2\}$, and $rf_1 + sf_2$ for all r and s real.

Exercise 7.25. Show that every null function $f : \mathbb{R} \to \mathbb{R}$ is almost continuous.

Exercise 7.26. Let E be an almost closed set and suppose that $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = 1$ for all $x \in E$ and $f(x) = 0$ for all $x \notin E$. Show that f is almost continuous.

Exercise 7.27. Give an example of an almost continuous function $f : \mathbb{R} \to \mathbb{R}$ that is not itself continuous at any point.

7.10.1. Associated sets for almost continuous functions. We recall that, if $f : \mathbb{R} \to \mathbb{R}$ is continuous, then for every real number r the sets

$$\{x : f(x) \leq r\} \quad \text{and} \quad \{x : f(x) \geq r\}$$

are closed. The analogous statement is true for almost continuous functions. (See Theorem 7.41 below where it is proved that this is a characterization.)

THEOREM 7.28. If the function $f : \mathbb{R} \to \mathbb{R}$ is almost continuous, then for every real number r the sets

$$\{x : f(x) \leq r\} \quad \text{and} \quad \{x : f(x) \geq r\}$$

are almost closed.

Proof. It is enough to show that the set

$$E = \{x : f(x) \leq r\}$$

is almost closed if f is almost continuous. Let $\epsilon > 0$ and choose an open set G so that $\mathcal{L}(G) < \epsilon$ and a continuous function $g : \mathbb{R} \to \mathbb{R}$ so that $f(x) = g(x)$ for all $x \notin G$. Write

$$E' = \{x : g(x) \leq r\}.$$

This is a closed set and $E \setminus G = E' \setminus G$. Thus E is an almost closed set. ■

7.10.2. Almost everywhere continuous functions. A function $f : \mathbb{R} \to \mathbb{R}$ is continuous if it is continuous at *every* point. We say then that a function is almost everywhere continuous if it is continuous at almost every point. The phrases "almost continuous" and "almost everywhere continuous" are perilously close, but have very different meanings.

EXERCISE 7.29. Show that any function that is almost everywhere continuous is also almost continuous, but not conversely.

7.11. Egorov-Taylor Theorem

Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of functions converging at each point x in a compact interval $[a, b]$ to a function f: i.e.,

$$\lim_{n\to\infty} f_n(x) = f(x) \quad (a \leq x \leq b).$$

In general there is not much more that can be said. But if the functions are continuous or almost continuous then stronger conclusions can be drawn in this situation. We can use a number of almost arguments to show that the convergence is "almost" uniform and, moreover, that the convergence is almost faster than it appears. Indeed, we can find a sequence of positive numbers δ_n decreasing to zero so that for any $\epsilon > 0$ an open set G can be selected for which $\mathcal{L}(G) < \epsilon$ and so that

$$\lim_{n\to\infty} \frac{f_n(x) - f(x)}{\delta_n} = 0$$

holds uniformly on $[a, b] \setminus G$. This is the content of the Egorov-Taylor theorem.

7.11.1. Almost uniform convergence. Uniform convergence is one of the standard tools of analysis. We have seen already that a *uniform* limit of continuous functions is continuous. The almost version allows that tool to be used in very weak situations.

DEFINITION 7.30. Let $f_n : \mathbb{R} \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ be a sequence of functions and K a compact set. We say that $\{f_n\}$ converges uniformly on K if for all $\eta > 0$ there is an integer N so that

$$|f_n(x) - f_m(x)| < \eta$$

for all n, $m \geq N$ and all $x \in K$.

DEFINITION 7.31. Let $f_n : [a, b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ be a sequence of functions. We say that f_n converges almost uniformly on $[a, b]$ if for all $\epsilon > 0$ there is an open set G so that $\mathcal{L}(G) < \epsilon$ and $\{f_n\}$ converges uniformly on the compact set $[a, b] \setminus G$.

It is transparent that almost uniform convergence implies convergence almost everywhere, since uniform convergence on a set $[a, b] \setminus G$ implies convergence at each point of that set.

EXERCISE 7.32. Let $f_n : [a, b] \to \mathbb{R}$ $(n = 1, 2, 3, \dots)$ be a sequence of functions that converges almost uniformly on $[a, b]$. Then there is a function $f : [a, b] \to \mathbb{R}$ so that

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

at almost every point x in $[a, b]$.

7.11.2. Pointwise convergence of almost continuous functions.

LEMMA 7.33. Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of almost continuous functions so that the sequence of real numbers $\{f_n(x)\}$ converges for each point x in the compact interval $[a, b]$. Write for any $\eta > 0$ and integer m,

$$E_{m\eta} = \bigcap_{n=m}^{\infty} \bigcap_{p=m}^{\infty} \{x \in [a, b] : |f_n(x) - f_p(x)| < \eta\}$$

Then,

$$\bigcup_{m=1}^{\infty} E_{m\eta} = [a, b]$$

and

$$\lim_{m\to\infty} \mathcal{L}([a, b] \setminus E_{m\eta}) = 0.$$

Proof. Since the functions $|f_n - f_p|$ here are all almost continuous we see that each set $E_{m\eta}$ is almost closed. By our assumptions on the convergence of the sequence of numbers $\{f_n(x)\}$ we can check that, holding η fixed, the sequence $E_{1\eta}$, $E_{2\eta}$, $E_{3\eta}$, $\dots$ is an increasing sequence of almost closed sets whose union is all of $[a, b]$. That means that the sequence

$$[a, b] \setminus E_{1\eta},\ [a, b] \setminus E_{2\eta},\ [a, b] \setminus E_{3\eta},\ \dots$$

is a decreasing sequence of almost closed sets whose intersection is empty. As a consequence,

$$\lim_{m\to\infty} \mathcal{L}\left([a, b] \setminus E_{m\eta}\right) = 0.$$

■

THEOREM 7.34 (Egorov). Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of almost continuous functions converging at each point x in a compact interval $[a, b]$ to a real-valued function f: i.e.,

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

Then $\{f_n\}$ converges almost uniformly on $[a, b]$ and the function f is almost continuous.

Proof. Let $\epsilon > 0$. Using the lemma and the fact that the sets $\{E_{m\eta}\}$ in the lemma are almost closed we may select, for each integer m, an integer N_m and an open set G_m so that $\mathcal{L}(G_m) < \epsilon 2^{-m}$ with the property that

$$|f_n(x) - f_p(x)| < 2^{-m}$$

for all n, $p \geq N_m$ and for every x in $[a, b]$ that does not belong to G_m.

We write $G = \bigcup_{m=1}^{\infty} G_m$. Then G is an open set for which

$$\mathcal{L}(G) \leq \sum_{m=1}^{\infty} \mathcal{L}(G_m) < \epsilon.$$

Just check now that $\{f_n\}$ converges uniformly on the compact set $[a, b] \setminus G$.

Finally, to complete the proof, it remains only to show that f is almost continuous. Since each function f_n is almost continuous, we may select open sets U_n for which $\mathcal{L}(U_n) < \epsilon 2^{-n}$ and continuous functions $g_n : \mathbb{R} \to \mathbb{R}$ so that $f_n(x) = g_n(x)$ for all $x \in [a, b] \setminus U_n$. Set

$$U = G \cup \bigcup_{n=1}^{\infty} U_n.$$

Note that $\mathcal{L}(U) < 2\epsilon$ and that $\{f_n\}$, and hence also $\{g_n\}$, is uniformly convergent on the compact set $K = [a, b] \setminus U$. Write $g(x) = \lim_{n\to\infty} g_n(x)$ for all $x \in K$. The function g, as a uniform limit of continuous functions, is continuous relative to the set K and so we may extend to a continuous function $g : \mathbb{R} \to \mathbb{R}$. As $f(x) = g(x)$ for all $x \in K$ we have that f is almost equal to a continuous function. Thus we have exhibited exactly the statement required to allow us to assert that f is almost continuous. ■

EXERCISE 7.35. Let $f_n(x) = x^n$ for each x in the compact interval $[0, 1]$. Show that $\{f_n(x)\}$ converges at each point x of $[0, 1]$ but that $\{f_n\}$ is not uniformly convergent on $[0, 1]$.

7.11.3. Strengthening the convergence.

LEMMA 7.36. Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of almost continuous functions converging at every point x in a compact interval to a real-valued function f. Then there is a nonincreasing sequence of positive real numbers δ_n converging to zero so that, for almost every x in $[a, b]$,

$$\lim_{n\to\infty} \frac{f_n(x) - f(x)}{\delta_n} = 0.$$

Proof. Choose a decreasing sequence of open sets $\{G_\mu\}$ so that $\mathcal{L}(G_\mu) < 2^{-\mu}$ and so that $f_n \to f$ uniformly on the compact set $[a,b] \setminus G_\mu$.

Let $N = \bigcap_{\mu=1}^\infty G_\mu$ and note that N is a null set. We already know that

$$\lim_{n\to\infty} f_n(x) - f(x) = 0$$

for each $x \in [a,b] \setminus N$. Now we wish to find a sequence $\{\delta_n\}$ decreasing to zero so that in fact

$$\lim_{n\to\infty} \frac{f_n(x) - f(x)}{\delta_n} = 0$$

for each $x \in [a,b] \setminus N$.

Choose, for each pair of integers μ and r, an integer $n_{\mu r}$ to be the least integer with the property that, for all $x \in [a,b] \setminus G_\mu$ and all $n \geq n_{\mu r}$,

$$|f_n(x) - f(x)| < \frac{1}{r+1}.$$

Note that $\{n_{\mu r}\}$ forms, for fixed μ, a nonincreasing sequence of integers $n_{\mu r}$ ($r = 1, 2, 3, \dots$) as well as, for fixed r, a nonincreasing sequence of integers $n_{\mu r}$ ($\mu = 1, 2, 3, \dots$). An examination of the double sequence shows that we are able to select an increasing sequence of integers n_1, n_2, n_3, $\dots$ growing rapidly enough so that, for each fixed μ,

$$\lim_{t\to\infty} \frac{n_t}{n_{\mu t}} = \infty.$$

Define δ_n as follows: to start the sequence let $\delta_n = 1$ for all $1 \leq n \leq n_1$. Then, for $n_{t-1} + 1 \leq n \leq n_t$ and $t = 2, 3, 4, \dots$ we let

$$\delta_n = \frac{1}{\sqrt{t}}.$$

This defines a nonincreasing sequence $\{\delta_n\}$ of positive numbers whose limit is evidently zero.

Take any point $x \in [a,b] \setminus N$. Select μ so that $x \notin G_\mu$ and then an integer t_0 so that

$$n_t \geq n_{\mu t}$$

whenever $t \geq t_0$. Check that when $n \geq n_{t_0}$

$$|f_n(x) - f(x)| \leq \delta_n^2.$$

In particular, we can conclude that

$$\lim_{n\to\infty} \frac{f_n(x) - f(x)}{\delta_n} = 0$$

for every point in $[a,b] \setminus N$. ■

The lemma cannot be improved to drop the reference to a set of measure zero.

Exercise 7.37. Let $[a,b]$ be a compact interval. Show that there is a sequence of null functions $\{f_n\}$ defined on $[a,b]$ so that $\lim_{n\to\infty} f_n(x) = 0$ for each x in $[a,b]$ but that, given any nonincreasing sequence of positive real numbers $\{\delta_n\}$ converging to zero, there must exist at least one point x_0 in $[a,b]$ at which,

$$\lim_{n\to\infty} \frac{f_n(x_0)}{\delta_n} = \infty.$$

Hint: Select a compact null subset K of $[a,b]$ that is not countable. You may assume that there is a correspondence identifying each nonincreasing sequence of real numbers $\{\delta_n\}$ converging to zero with a unique point in K. Then define $f_n(x)$ to be zero for all n and all x except when x has been identified with such a sequence $\{\delta_n\}$ converging to zero. In that case define $f_n(x) = \sqrt{\delta_n}$.

7.11.4. Egorov-Taylor theorem.

THEOREM 7.38 (Egorov-Taylor). Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of almost continuous functions converging at almost every point x in a compact interval to a real-valued function f. Then there is a nonincreasing sequence of real numbers δ_n converging to zero so that the sequence of functions

$$\frac{f_n(x) - f(x)}{\delta_n}$$

converges almost uniformly to zero on $[a,b]$.

Proof. The "almost every point" statement here can be ignored and replaced by "every point" if we define $f_n(x)$ and $f(x)$ to be zero at every point where we do not have $\lim_{n\to\infty} f_n(x) = f(x)$. That reduces the situation to the one we have been handling, with convergence at every point.

In Lemma 7.36, we have already proved that

$$\lim_{n\to\infty} \frac{f_n(x) - f(x)}{\delta_n} = 0$$

at almost every point of $[a,b]$. Again we may assume that this holds at every point. Finally, then, an application of Theorem 7.34 shows that this convergence is almost uniform on $[a,b]$. ■

7.12. Characterizations of almost continuity

There are a number of characterizations of the concept of almost continuity that are useful.

7.12.1. Limits of continuous functions.

THEOREM 7.39 (Limits of continuous functions). A function $f : \mathbb{R} \to \mathbb{R}$ is almost continuous if and only if it is the limit almost everywhere of a sequence of continuous functions.

Proof. Egorov's theorem supplies one direction. The other direction follows from the definition. Choose open sets G_n and continuous functions $g_n : \mathbb{R} \to \mathbb{R}$ so that

$$\mathcal{L}(G_n) < 2^{-n}$$

and $f(x) = g_n(x)$ for all $x \notin G_n$. Write

$$N = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} G_n.$$

Check that N is a null set and that

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

for all $x \notin N$. ■

7.12.2. Associated sets. Continuity and almost continuity have very similar characterizations in terms of closed sets and almost closed sets. In the first theorem continuity is characterized by the fact of certain associated sets being closed; in the second almost continuity is characterized by these sets being almost closed.

THEOREM 7.40 (Continuous functions). A function $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only for every real number r the sets

$$\{x : f(x) \leq r\} \quad \text{and} \quad \{x : f(x) \geq r\}$$

are closed.

Proof. If f is continuous we have already verified that these sets must be closed. Conversely suppose that such sets are always closed and fix a point x_0 and $\epsilon > 0$. The set

$$G = \{x : f(x_0) - \epsilon < f(x) < f(x_0) + \epsilon\}$$

must contain x_0 and must be open since it is the complement of the closed set

$$\{x : f(x) \leq f(x_0) - \epsilon\} \cup \{x : f(x) \geq f(x_0) + \epsilon\}.$$

If $[c, d]$ is a compact interval contained in G and containing x_0 then certainly $\omega f([c, d]) < 2\epsilon$. From that we can deduce that $\omega_f(x_0) < 2\epsilon$. This is true for every point x_0 and every $\epsilon > 0$ and so f is continuous. ∎

THEOREM 7.41 (Almost continuous functions). A function $f : \mathbb{R} \to \mathbb{R}$ is almost continuous if and only for every real number r the sets

$$\{x : f(x) \leq r\} \quad \text{and} \quad \{x : f(x) \geq r\}$$

are almost closed.

Proof. In Theorem 7.28 we proved the first direction. In the converse direction suppose that the sets of the theorem are almost closed for every r. Note that

$$\{x : f(x) > r\} = \bigcap_{m=1}^{\infty} \{x : f(x) \geq r - 1/m\}$$

so that sets of the form

$$\{x : f(x) < r\} \quad \text{and} \quad \{x : f(x) > r\}$$

are also almost closed.

Let us assume for the moment that f is nonnegative. Fix n, and for each $k = 1, 2, 3, \ldots, n2^n$, define $f_{kn}(x) = (k - 1)2^{-n}$ if

$$(k - 1)2^{-n} \leq f(x) < k2^{-n}$$

and 0 otherwise. Let $g_n(x) = n$ if $f(x) > n$ and 0 otherwise. Argue that each such function f_k, g_n is almost continuous because the sets

$$\{x : (k - 1)2^{-n} \leq f(x) < k2^{-n}\}$$

and

$$\{x : f(x) > n\}$$

are almost closed. Each of these functions assumes a single value on an almost closed set. Verify that such functions must be almost continuous. The sum of these functions must also be almost continuous (because sums of continuous functions are continuous).

Check the identity

$$f(x) = \lim_{n\to\infty} \sum_{k=1}^{n2^n} f_{kn}(x) + g_n(x).$$

This expresses f as the limit everywhere of a sequence of almost continuous functions. Thus f is almost continuous too.

In the general case $f(x) = [f(x)]^+ - [f(x)]^-$ expresses f as the sum of two nonnegative functions which also satisfy the hypotheses of the theorem. Thus f is almost continuous. ■

CHAPTER 8

Measure and Integral: Lebesgue's Program

The integral

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

cannot be considered to have been "constructed" by either of the definitions we have given. For the Newton version of the integral this is obvious: we have no general method for finding an indefinite integral of a given function, nor even of knowing in advance whether there is an indefinite integral. For the integral defined by Cousin covers, we must acknowledge the fact that, except in special cases, the Cousin covers themselves are not constructible. In that sense our definitions so far are purely formal.

In 1901 the program for the construction of the integral by means of the measure theory was undertaken successfully by Lebesgue. As he did not have the correct and simple definition for the integral, he took this program itself as the definition of the integral. Such a program is limited by two features. The first is that such a measure construction for the integral $\int_a^b f(x)\,dx$ is not possible for all integrable functions f.

The second unpleasant feature is that using this program to *define* the integral requires many technical details to be worked through before developing many of the properties of the integral. This feature caused much resistance among many of his contemporaries, who felt that the integration theory he had presented was difficult and nonintuitive. Even a century later, by which time Lebesgue's program had been fully accepted by professional mathematicians, this view caused many to avoid teaching the subject except to advanced students.

The start of the program is the theorem asserting that for any almost closed subset E of a compact interval $[a,b]$,

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

The function χ_E, here, is called the *indicator function* or *characteristic function* of the set E; it is defined as $\chi_E(x) = 1$ if $x \in E$ and as $\chi_E(x) = 0$ if $x \notin E$. We begin with compact sets.

8.1. Measure of compact sets

LEMMA 8.1. 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

$$\sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi[K]} \mathcal{L}(I) \geq \mathcal{L}(K).$$

This gives a lower bound for the lower integral:

$$\mathcal{L}(K) \leq \underline{\int_a^b} \chi_K(x)\,dx. \tag{8.1}$$

In the other direction take any open set $G_2 \supset K$. and 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

$$\sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi[K]} \mathcal{L}(I) \leq \mathcal{L}(G).$$

This gives a upper bound of $\mathcal{L}(G)$ for the upper integral. Since G_2 is an arbitrary open set containing K, we deduce that

$$\mathcal{L}(K) \geq \overline{\int_a^b} \chi_K(x)\,dx. \tag{8.2}$$

Together, (8.1) and (8.2) complete the proof. ∎

8.2. Measure of almost closed sets

LEMMA 8.2. Let E be an almost closed 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. First we show that

$$\overline{\int_a^b} \chi_E(x)\,dx \leq \mathcal{L}(E). \tag{8.3}$$

For this take any open set G containing E and define

$$\beta = \{(I,x) : \text{for } x \in G \text{ and } x \in I \subset G \cap [a,b] \text{ or for } x \notin G \text{ and any } x \in I \subset [a,b]\}.$$

This is a Cousin cover of $[a,b]$. For any partition $\pi \subset \beta$ of $[a,b]$ it is easy to check that

$$\sum_{(I,x)\in\pi} \chi_E(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi[E]} \mathcal{L}(I) \leq \mathcal{L}(G).$$

From this deduce that

$$\overline{\int_a^b} \chi_E(x)\,dx \leq \mathcal{L}(G).$$

Finally, as this is true for all open sets G containing E, the inequality (8.3) follows.

Suppose now that K is a compact subset of E. Then, from the preceding lemma and the simple inequality $\chi_K(x) \leq \chi_E(x)$, we obtain

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

Now, recall from Theorem 7.9.1 that $\mathcal{L}(E)$ may be estimated from below by compact sets whenever E is almost closed. From that we deduce that,

$$\mathcal{L}(E) \leq \underline{\int_a^b} \chi_E(x)\,dx \leq \overline{\int_a^b} \chi_E(x)\,dx \leq \mathcal{L}(E)$$

which yields the lemma. ■

8.3. Integral of simple functions

Lemma 8.3. Let $f : [a,b] \to \mathbb{R}$ be a simple function expressible as

$$f(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)$$

for a finite collection of real numbers $\{a_i\}$ and almost closed subsets $\{E_i\}$ of $[a,b]$. Then f is integrable on $[a,b]$ and

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} a_i \mathcal{L}(E_i).$$

Proof. We know that each function χ_{E_i} is integrable and so too, then, is any linear combination. The formula is now obvious. ■

8.4. Integral of nonnegative almost continuous functions

Lemma 8.4. Let $f : [a,b] \to \mathbb{R}$ be a nonnegative, almost continuous function. Then f is expressible 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 almost closed subsets $\{A_i\}$ of $[a,b]$ and

$$\int_a^b f(x)\,dx = \sum_{i=1}^{\infty} a_i \mathcal{L}(A_i). \tag{8.4}$$

This is finite if and only if f is integrable.

Proof. 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

$$A_k = \left\{ x \in [a,b] : f(x) \geq r_k + \sum_{j<k} r_j \chi A_j(x) \right\}$$

inductively, starting with $A_0 = \emptyset$. Then check that

$$f(x) = \sum_{k=1}^{\infty} r_k \chi A_k(x)$$

at every $x \in [a, b]$. The identity in (8.4) then follows directly from Lemma 8.3 and Theorem 6.20. Note that the series may have an infinite sum; when the sum is finite the function f will be integrable. ∎

8.5. Limitations of the Lebesgue program

How far can the program be pushed? If $f : [a, b] \to \mathbb{R}$ is an almost continuous function then f can be displayed in a familiar way as the difference of two nonnegative, almost continuous functions

$$f = [f]^+ - [f]^-.$$

In that case we can claim

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

precisely in the case where one of these two expressions is finite. If both are finite then we know further that f is integrable on $[a, b]$.

Thus, in this case, the integral has been constructed directly from the measure theory. The remaining case where f is integrable but

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

will be, by analogy with the situation for series, considered as the "nonabsolutely integrable" case. That case is the only situation that eludes the constructive program of Lebesgue.

THEOREM 8.5. Let $f : [a, b] \to \mathbb{R}$ be an almost continuous function for which

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

Then f must be integrable.

COROLLARY 8.6. A bounded, almost continuous function is integrable on every compact interval.

8.6. Derivatives of monotonic functions

Within Lebesgue's program it is by no means easy to handle derivatives. In the natural integration theory presented earlier this is immediate, but it is also trivial and formal. One of the early successes of Lebesgue's methods was this theorem asserting the integrability of derivatives almost everywhere of monotonic functions.

THEOREM 8.7. Suppose that $f : [a, b] \to \mathbb{R}$ is the derivative almost everywhere in a compact interval $[a, b]$ of a monotonic nondecreasing function F. Then f is integrable on $[a, b]$ and

$$0 \le \int_a^b f(x)\,dx \le F(b) - F(a). \tag{8.5}$$

Proof. Let E be the set of points x in $[a,b]$ where $F'(x) = f(x)$. Define $f_1(x) = 0$ for $x \notin E$ and $f_1(x) = f(x)$ for $x \in E$. Note that

$$f_1(x) = \lim_{n\to\infty} n[F(x+1/n) - F(x)]$$

for almost every x in $[a,b]$. We can check that F is almost continuous and then deduce that f_1 is almost continuous too.

Thus f_1 is a nonnegative, almost continuous function on $[a,b]$. Consequently, by the material of the sections above, the integrals

$$\int_a^b f(x)\,dx = \int_a^b f_1(x)\,dx$$

exist, but may be infinite. It remains to establish the inequality of the theorem and deduce that the integrals are finite, implying that f and f_1 are integrable.

Note that

$$\underline{D}F(x) \geq f_1(x)$$

at every point x of $[a,b]$. From this it follows, using standard derivate estimates (cf. Lemma 6.12), that

$$0 \leq \int_a^b f_1(x)\,dx = \overline{\int_a^b} f_1(x)\,dx \leq F(b) - F(a).$$

This establishes the inequality (8.5) and completes the proof. ∎

8.7. The Lebesgue "integral"

Thus far we have seen that Lebesgue's construction will yield the integral

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

precisely in the case the function $f : [a,b] \to \mathbb{R}$ is almost continuous. This would be finite only if f is also integrable.

In fact the Lebesgue construction is really a construction of a measure, not an integral. The integral is constructed as part of a larger goal. The goal is to construct a set function

$$E \to \int_E f(x)\,dx$$

and then the "integral" is the special case

$$\int_{[a,b]} f(x)\,dx.$$

In Lebesgue's theory this is defined in an entirely constructive manner provided only that $f : \mathbb{R} \to \mathbb{R}$ is a nonnegative, almost continuous function and E is an almost closed set.

We give a full covering version of Lebesgue's "integral" by a very familiar method.

Definition 8.8. Let $f : \mathbb{R} \to \mathbb{R}$ be an arbitrary nonnegative function and E an arbitrary set of real numbers. Then define

$$V^*(f\mathcal{L}, E) = \inf_{\beta} \sup_{\pi\subset\beta} \left(\sum_{([u,v],w)\in\pi} f(w)(v-w) \right),$$

where the infimum is with regard to all full covers β of the set E and the supremum with regard to all subpartitions π that are contained in β.

The "integral" notation will prove useful:

$$\int_E f(x)\,dx = V^*(f\mathcal{L}, E)$$

with the understanding that f is always taken as nonnegative. In the event that f is almost continuous and E almost closed this is equivalent to Lebesgue's version of this notion (proved by repeating the methods of this chapter). Since we do not have to make any such restrictions we have a simpler theory. Moreover, since the concept is defined by full covers there is an immediate connection between this measure and the integral.

It is not difficult to show, for any nonnegative function f, that

$$\int_{[a,b]} f(x)\,dx = \overline{\int_a^b} f(x)\,dx.$$

This leads immediately to the following theorem expressing the integral, in special cases, as a measure. [It will be proved later that all integrable functions are almost continuous; thus this characterization is already available to us from the material above.]

THEOREM 8.9 (Measure characterization). Suppose that $f : [a,b] \to \mathbb{R}$ and that both f and $|f|$ are integrable. Then

$$\int_a^b f(x)\,dx = \int_{[a,b]} f^+(x)\,dx - \int_{[a,b]} f^-(x)\,dx,$$

and

$$\int_a^b |f(x)|\,dx = \int_{[a,b]} f^+(x)\,dx + \int_{[a,b]} f^-(x)\,dx.$$

8.8. Some measure estimates for Lebesgue's "integral"

LEMMA 8.10. Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that $0 \le r < f(x) < s$ for all x in a set E. Then

$$r\mathcal{L}^*(E) \le \int_E f(x)\,dx \le s\mathcal{L}^*(E).$$

Proof. For any subpartition $\pi = \pi[E]$ note that $rV(\mathcal{L}, \pi) \le V(f\mathcal{L}, \pi) \le sV\mathcal{L}, \pi)$. From this deduce that

$$r\mathcal{L}^*(E) \le \int_E f(x)\,dx \le s\mathcal{L}^*(E).$$

■

LEMMA 8.11. Suppose that $F : \mathbb{R} \to \mathbb{R}$ and that F has a finite derivative $F'(x)$ at every point x in a set E. Then

$$V^*(\Delta F, E) = \int_E |F'(x)|\,dx.$$

Proof. Let $\epsilon > 0$ and suppose that E has finite Lebesgue measure. Consider the covering relation

$$\beta = \{([u,v],w) : w \in E,\ |F(v) - F(u) - F'(w)(v-u)| \leq \epsilon(v-u)\}.$$

This is evidently a full cover of E and one can use this to deduce that

$$V^*(\Delta F - F'\mathcal{L}, E) \leq \epsilon V^*(\mathcal{L}, E) = \epsilon \mathcal{L}^*(E).$$

Thus

$$V^*(\Delta F - F'\mathcal{L}, E) = 0$$

from which

$$V^*(\Delta F, E) = V^*(F'\mathcal{L}, E)$$

easily follows. If E has infinite measure, simply write it as a union of sets of finite measure and apply the same argument to each piece to deduce, once again, that

$$V^*(\Delta F - F'\mathcal{L}, E) = 0.$$

■

Exercise 8.12 (Construction of the Lebesgue "integral"). Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that f is a nonnegative, almost continuous function. Let $r > 1$ and write

$$A_{kr} = \{x : r^{k-1} < f(x) \leq r^k\}.$$

Show that, for any set E,

$$\int_E f(x)\,dx \leq \sum_{k=-\infty}^{\infty} r^k \mathcal{L}(E \cap A_{kr}) \leq r \int_E f(x)\,dx.$$

8.9. Absolute continuity property of Lebesgue's "integral"

It is easy to prove that the integral $\int_E f(x)\,dx$ has a simple absolute continuity property.

Lemma 8.13. Suppose that $f : \mathbb{R} \to \mathbb{R}$, that f is nonnegative, and that N is a null set. Then

$$\int_N f(x)\,dx = 0.$$

There is an ϵ, δ-version of this property similar to Vitali's notion of absolute continuity: we show that the Lebesgue integral is small over small sets under natural hypotheses.

Theorem 8.14. Suppose that $f : \mathbb{R} \to \mathbb{R}$, that f is nonnegative and almost continuous, and that

$$\int_E f(x)\,dx < \infty.$$

Then for every $\epsilon > 0$ there is a $\delta > 0$ so that if G is an open set with $\mathcal{L}(G) < \delta$ then

$$\int_{E \cap G} f(x)\,dx < \epsilon.$$

Proof. For illustrative purposes only we begin the proof with the bounded case. Suppose that $f(x) < N$ for all $x \in E$. Choose $\delta = \epsilon/N$ and observe that, if $\mathcal{L}(G) < \delta$ then the inequalities in the measure estimates of Lemma 8.10 provide

$$\int_{E\cap G} f(x)\,dx \leq N\mathcal{L}(G) < \epsilon.$$

Thus the proof in the bounded case is trivial and does not make use of the fact that f is almost continuous.

This argument suggests how to proceed. Let

$$A_n = \{x : n-1 \leq f(x) < n\}.$$

From the fact that f is almost continuous we can deduce that A_n is almost closed. Thus we can select an open set G_n for which $B_n = A_n \setminus G_n$ is closed and $\mathcal{L}(G_n) < \epsilon 2^{-n}n^{-1}$. That also requires

$$\begin{aligned}\int_{E\cap A_n} f(x)\,dx &\leq \int_{E\cap B_n} f(x)\,dx + \int_{E\cap A_n\cap G_n} f(x)\,dx\\ &\leq \int_{E\cap B_n} f(x)\,dx + \epsilon 2^{-n}.\end{aligned}$$

Note that $\{B_n\}$ is a disjointed sequence of closed sets whose union B can be handled by the usual additive properties of measures over such sets. Thus

$$\begin{aligned}\sum_{n=1}^{\infty}\int_{E\cap A_n} f(x)\,dx &\leq \sum_{n=1}^{\infty}\int_{E\cap B_n} f(x)\,dx + \epsilon\\ &= \int_{E\cap B} f(x)\,dx + \epsilon \leq \int_E f(x)\,dx + \epsilon < \infty.\end{aligned}$$

In particular there must be an integer N sufficiently large that

$$\sum_{n=N+1}^{\infty}\int_{E\cap A_n} f(x)\,dx < \epsilon/2.$$

Choose $\delta = \epsilon/(2N)$ and let G be any open set for which $\mathcal{L}(G) < \delta$. Since

$$E\cap G = \{x \in E\cap G : f(x) < N\} \cup \bigcup_{n=N+1}^{\infty}(G\cap E\cap A_n)$$

we have

$$\int_{E\cap G} f(x)\,dx \leq N\mathcal{L}(G) + \sum_{n=N+1}^{\infty}\int_{E\cap A_n} f(x)\,dx < \epsilon.$$

■

CHAPTER 9

Theory for the Integral

In this chapter we extend our understanding of the integral by introducing a number of integrability criteria. We are still lacking one of the deeper tools needed for our study, the Vitali covering theorem that will be addressed in a later chapter. This means we cannot yet complete the full theory of the integral.

9.1. Integrability criteria

The theoretical development of the integral depends on a number of integrability criteria.

9.1.1. First Cauchy criterion.

Theorem 9.1. A necessary and sufficient condition in order for a function $f : [a, b] \to \mathbb{R}$ to be integrable on a compact interval $[a, b]$ is that there is a number c so that for all $\epsilon > 0$ a Cousin cover β of $[a, b]$ can be found so that

$$\left| \sum_{(I,z)\in\pi} f(z)\mathcal{L}(I) - c \right| < \epsilon$$

for all partitions π of $[a, b]$ contained in β.

Proof. In Exercise 6.8 we checked that this condition is sufficient. The proof for necessity is similar to the analogous statement for sequence limits in Section 1.9. ■

9.1.2. Second Cauchy criterion.

Theorem 9.2. A necessary and sufficient condition in order for a function $f : [a, b] \to \mathbb{R}$ to be integrable on a compact interval $[a, b]$ is that, for all $\epsilon > 0$, a Cousin cover β of $[a, b]$ can be found so that

$$\left| \sum_{(I,z)\in\pi} \sum_{(I',z')\in\pi'} [f(z) - f(z')]\mathcal{L}(I \cap I') \right| < \epsilon$$

for all partitions π, π' of $[a, b]$ contained in β.

Proof. The proof is similar to the analogous statement for sequence limits in Section 1.9. Here are some details.

Start by checking that when π and π' are both partitions of the same interval $[a, b]$ then, for any $(I, x) \in \pi$,

$$\mathcal{L}(I) = \sum_{(I',x')\in\pi'} \mathcal{L}(I \cap I')$$

from which it is easy to see that

$$\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi}\sum_{(I',x')\in\pi'} f(x)\mathcal{L}(I\cap I').$$

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

$$\left|\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \sum_{(I',x')\in\pi'} f(x')\mathcal{L}(I')\right|$$

to assume the simple form

$$\left|\sum_{(I,x)\in\pi}\sum_{(I',x')\in\pi'} [f(x)-f(x')]\mathcal{L}(I\cap I')\right|. \tag{9.1}$$

■

9.1.3. Integrability on subintervals.

LEMMA 9.3 (Integrability on subintervals). If $f : [a,b] \to \mathbb{R}$ is integrable then it is also integrable on any compact subinterval of $[a,b]$.

Proof. Let $\epsilon > 0$. Suppose that f is integrable on $[a,b]$ and $[c,d]$ is a compact subinterval. Take any Cousin cover β of $[a,b]$ so that the second Cauchy criterion is satisfied for β. Note that β is a Cousin cover of $[c,d]$. Observe that for every pair of partitions π_1, and $\pi_2 \subset \beta$ of the subinterval $[c,d]$, there is a subpartition π from β so that $\pi_1 \cup \pi$ and $\pi_1 \cup \pi$ are partitions of the full interval $[a,b]$. In particular then

$$\left|\sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I)\right| =$$

$$\left|\sum_{(I,x)\in\pi\cup\pi_1} f(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi\cup\pi_2} f(x)\mathcal{L}(I)\right| < \epsilon$$

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

9.1.4. The indefinite integral.

THEOREM 9.4 (The indefinite integral). If $f : [a,b] \to \mathbb{R}$ is integrable then there is a function $F : [a,b] \to \mathbb{R}$, called an indefinite integral for f, so that

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

for every compact subinterval $[c,d]$ of $[a,b]$.

Proof. Theorem 9.3 supplies the existence of the integral on the subintervals and Theorem 6.16 shows that the integral is an additive interval function, hence can be written as ΔF for some such function F. ■

9.2. Absolutely integrable functions

Using normal inequality techniques we easily observe that the expression (9.1) that we use for the second Cauchy criterion must be smaller than a quite similar expression:

$$\left|\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'}[f(z)-f(z')]\mathcal{L}(I\cap I')\right| \leq \sum_{(I,x)\in\pi}\ \sum_{(I',x')\in\pi'}|f(x)-f(x')|\,\mathcal{L}(I\cap I').$$

It takes a sharp (and young) eye to spot the difference, but the larger side of this inequality may be strictly larger. This leads to a stronger integrability criterion than that in the second Cauchy criterion. This is the motivation for the notion of absolutely integrable functions. The criterion is named after E. J. McShane.

DEFINITION 9.5 (McShane's criterion). A function $f:[a,b]\to\mathbb{R}$ is said to be *absolutely integrable* on $[a,b]$ provided that for all $\epsilon>0$ a Cousin cover β of $[a,b]$ can be found so that

$$\sum_{(I,z\in\pi}\ \sum_{(I',z')\in\pi'}|f(z)-f(z')|\,\mathcal{L}(I\cap I')<\epsilon$$

for all partitions π, π' of $[a,b]$ contained in β.

EXERCISE 9.6. Show that if f is absolutely integrable on $[a,b]$ then it is also absolutely integrable on any subinterval $[c,d]$.

EXERCISE 9.7. Suppose that f and g are both absolutely integrable on $[a,b]$. Show that so too is any linear combination $rf+sg$.

EXERCISE 9.8. Suppose that each of the functions f_1, f_2, $\dots$, $f_n:[a,b]\to\mathbb{R}$ is absolutely integrable on a compact interval $[a,b]$ and that a function $L:\mathbb{R}^n\to\mathbb{R}$ is given satisfying

$$|L(x_1,x_2,\dots,x_n)-L(y_1,y_2,\dots,y_n)|\leq M\sum_{i=1}^n|x_i-y_i|$$

for some number M and all $(x_1,x_2,\dots,x_n)$ and $(y_1,y_2,\dots,y_n)$ in $\mathbb{R}^n$. Show that the function $g(x)=L(f_1(x),f_2(x),\dots,f_n(x))$ is absolutely integrable on $[a,b]$.

9.2.1. "Absolutely" integrable and absolutely integrable. It is an easy matter to check that if f is absolutely integrable then necessarily both f and $|f|$ must be integrable. This observation and the fact that the converse is true suggests the terminology; we must wait quite some time, however, before we can tackle the converse.

THEOREM 9.9. If f is absolutely integrable on $[a,b]$ then $|f|$ is absolutely integrable, both f and $|f|$ are integrable there and

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

Proof. It is immediate that if f satisfies McShane's criterion it also satisfies Cauchy's second criterion. Thus an absolutely integrable function is integrable. We then observe that, since

$$||f(x)|-|f(x')|\leq|\,|f(x)-f(x')|\,,$$

it is clear that whenever f satisfies McShane's criterion so too does $|f|$. Thus $|f|$ is absolutely integrable and hence integrable on $[a, b]$. The inequalities of the theorem simply follow from the inequalities $-|f(x)| \leq f(x) \leq |f(x)|$ which hold for all x.
■

EXERCISE 9.10. Show that there is an integrable function on the interval $[0, 1]$ that does not satisfy McShane's criterion.

Hint: Find a differentiable function F for which $|F'|$ cannot be integrable on $[0, 1]$.

9.3. Continuous functions are absolutely integrable

THEOREM 9.3.1. *If $f : [a, b] \to \mathbb{R}$ is continuous then f is absolutely integrable on $[a, b]$.*

Proof. Let $\epsilon > 0$ and define

$$\beta = \{([x, y], z) : \omega f([x, y]) < \epsilon/2(b - a).\}$$

Check, using the continuity of f, that β is a Cousin cover of $[a, b]$. Verify that if (I, z) and (I', z') both belong to β then, either I and I' have no points in common or else $|f(z) - f(z')| < \epsilon/(b - a)$.

Complete the proof by checking that

$$\sum_{(I,x)\in\pi} \sum_{(I',x')\in\pi'} |f(x) - f(x')|\, \mathcal{L}(I \cap I') < \epsilon$$

for any pair of partitions π and π' of $[a, b]$. ■

9.4. Bounded, almost-continuous functions are absolutely integrable

Later on we will characterize the absolutely integrable functions as those that are almost continuous and satisfy a certain boundedness property. The following is a weaker relative. We already know (Theorem 8.5) that such functions are integrable, so this theorem is presented just to show that these functions satisfy the stronger McShane criterion.

THEOREM 9.4.1. *Every bounded, almost continuous function is absolutely integrable on any compact interval.*

Proof. Let $\epsilon > 0$ and suppose that $|f(x)| < M$ for all x. Fix an interval $[a, b]$ and choose an open set G so that

$$\mathcal{L}(G) < \epsilon/(4M)$$

and so that $f(x) = g(x)$ for all $x \notin G$ where $g : \mathbb{R} \to \mathbb{R}$ is a continuous function. Using this continuity we can select a $\delta > 0$ so that

$$|g(z) - g(z')| < \epsilon/(2[b - a])$$

whenever $|z - z'| < \delta$.

Let

$$\beta_1 = \{(I, x) : \ x \in G \text{ and } I \subset G\}$$

and

$$\beta_2 = \{(I, x) : \ \mathcal{L}(I) < \delta/2 \text{ and } x \in I \setminus G\}.$$

Check that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a, b]$.

Take any two partitions π and π' from β and any two elements $(I, z) \in \pi$ and $(I', z') \in \pi'$. If either z or z' is in G then necessarily $I \cap I' \subset G$ and

$$|f(z) - f(z')| \leq 2M.$$

If neither z nor z' is in G then either I and I' have no points in common or else, if they do, then necessarily $|z - z'| < \delta$, $f(z) = g(z)$, $f(z') = g(z')$ and so

$$|f(z) - f(z')| < \epsilon/(2[b-a]).$$

This allows the computation, needed to use Definition 9.5:

$$\sum_{(I,z)\in\pi} \sum_{(I',z')\in\pi'} |f(z) - f(z')|\,\mathcal{L}(I \cap I') \leq (2M)\mathcal{L}(G) + (b-a)\epsilon/(2[b-a]) < \epsilon.$$

Consequently f is absolutely integrable on $[a, b]$. ■

9.5. Henstock's integrability criterion

THEOREM 9.5.1. *A necessary and sufficient condition for a function $f : [a, b] \to \mathbb{R}$ to be integrable on a compact interval $[a, b]$ and for F to be its indefinite integral is that for every $\epsilon > 0$ there exists a Cousin cover β of $[a, b]$ such that*

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon, \tag{9.2}$$

for every partition π of $[a, b]$ contained in β.

Proof. Suppose that this criterion holds. Then (9.2) immediately shows that

$$\left| F(b) - F(a) - \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \right| = \left| \sum_{(I,x)\in\pi} F(I) - \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \right|$$
$$\leq \sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon.$$

It follows that $F(b) - F(a) = \int_a^b f(x)\,dx$ by the first Cauchy criteria. The same argument will work on any subinterval to check that F is an indefinite integral for f.

Conversely let us suppose that F is an indefinite integral for f on $[a, b]$ and $\epsilon > 0$. By the Cauchy criterion there is a Cousin cover β of $[a, b]$ such that

$$\left| \Delta F([a, b]) - \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \right| < \epsilon/4 \tag{9.3}$$

for every partition π of $[a, b]$ contained in β and it will be our goal to establish (9.2) from this.

Fix π and let $\pi' \subset \pi$ be any nonempty subset. We see exactly how to supplement the subpartition π' so as to form a full partition of $[a, b]$: we write

$$\pi \setminus \pi' = \{(I_1, x_1), (I_2, x_2), \ldots (I_k, x_k)\}.$$

Our hypothesis requires F to be an indefinite integral for f on each I_i ($i = 1, 2, \ldots, k$) and so for each $i = 1, 2, \ldots, k$ we are able to select a partition $\pi_i \subset \beta$ of the interval I_i in such a way that

$$\left|\Delta F(I_i) - \sum_{(I,x)\in\pi_i} f(x)\mathcal{L}(I)\right| < \epsilon/(4k). \tag{9.4}$$

Thus if we augment π' to form

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

we obtain a partition of $[a,b]$ contained in β and thus also satisfying an inequality of the form (9.3). Computing with these ideas, we see

$$\sum_{(I,x)\in\pi'} \Delta F(I) = \Delta F([a,b]) - \sum_{i=1}^{k} \Delta F(I_i)$$

and

$$\sum_{(I,x)\in\pi'} f(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi''} f(x)\mathcal{L}(I) - \sum_{i=1}^{k}\left(\sum_{(I,x)\in\pi_i} f(x)\mathcal{L}(I)\right).$$

Putting these together with the estimates (9.3) and (9.4) we obtain

$$\left|\sum_{(I,x)\in\pi'} [\Delta F(I) - f(x)\mathcal{L}(I)]\right| \leq \left|\Delta F([a,b]) - \sum_{(I,x)\in\pi''} f(x)\mathcal{L}(I)\right|$$

$$+\sum_{i=1}^{k}\left|\Delta F(I_i) - \sum_{(I,x)\in\pi_i} f(x)\mathcal{L}(I)\right| < \epsilon/4 + k(\epsilon/(4k) = \epsilon/2.$$

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

To complete the proof let

$$\pi^+ = \{(I,x)\in\pi : F(I) - f(x)\mathcal{L}(I) \geq 0\}$$

and

$$\pi^- = \{(I,x)\in\pi : F(I) - f(x)\mathcal{L}(I) < 0\}.$$

Then

$$\sum_{(I,x)\in\pi^+} |\Delta F(I) - f(x)\mathcal{L}(I)| = \sum_{(I,x)\in\pi^+} [\Delta F(I) - f(x)\mathcal{L}(I)] < \epsilon/2$$

and

$$\sum_{(I,x)\in\pi^-} |\Delta F(I) - f(x)\mathcal{L}(I)| = \sum_{(I,x)\in\pi^-} -[\Delta F(I) - f(x)\mathcal{L}(I)] < \epsilon/2.$$

Adding the two inequalities proves (9.2). ■

9.6. Absolute continuity of the indefinite integral

In order for a function $F : [a, b] \to \mathbb{R}$ to be the indefinite integral of an integrable function it is both necessary and sufficient that F be absolutely continuous. At this stage, however, we can prove only the necessity part.

THEOREM 9.11. A necessary condition for a function $F : [a, b] \to \mathbb{R}$ to be the indefinite integral of an integrable function is that F is absolutely continuous.

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

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon/2, \tag{9.5}$$

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

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) < \epsilon/(2n), \tag{9.6}$$

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 (9.5) and (9.6) we deduce that

$$\begin{aligned}\sum_{(I,x)\in\pi} |\Delta F(I)| &\le \sum_{(I,x)\in\pi} |f(x)\mathcal{L}(I)| + \epsilon/2 \\ &\le \sum_{(I,x)\in\pi} n\mathcal{L}(I) + \epsilon/2 < \epsilon.\end{aligned}$$

By definition this shows that F cannot grow on the set N_n. ∎

9.7. Riemann's integrability criterion

The computation of an integral $\int_a^b f(x)\,dx$ formally involves estimating the Cauchy sums

$$\sum_{\pi} f(x)\mathcal{L}(I)$$

over partitions of the interval $[a, b]$. Since Cousin covers are not necessarily constructible we cannot rely on computing an integral this way. The problem is that, in general, this is highly sensitive to the placement of the points x inside the intervals for $(I, x) \in \pi$.

We could ask what class of functions would allow for a much freer choice of points. A moment's reflection shows that such functions would have to be bounded and there would have to be some control on the size of $\sum_{\pi} \omega f(I)\mathcal{L}(I)$.

DEFINITION 9.12 (Riemann's criterion). A bounded function $f : [a,b] \to \mathbb{R}$ is said to satisfy *Riemann's integrability criterion* on an interval $[a,b]$ provided that for every $\epsilon > 0$ there is a partition π of $[a,b]$ for which

$$\sum_{\pi} \omega f(I)\mathcal{L}(I) < \epsilon.$$

As an exercise (with mild historical interest) we show that such functions directly satisfy McShane's criterion.

THEOREM 9.7.1. *Every bounded function $f : [a,b] \to \mathbb{R}$ that satisfies Riemann's integrability criterion is absolutely integrable.*

Proof. We suppose that $|f(x)| < M$ for all $x \in [a,b]$. Given $\epsilon > 0$, choose points $a = a_0 < a_1 < a_2 < \cdots < a_n < a_{n+1} = b$ so that

$$\sum_{i=1}^{n+1} \omega f([a_{i-1}, a_i])\mathcal{L}([a_{i-1}, a_i]) < \epsilon/2. \tag{9.7}$$

Let

$$\beta = \{(I,x) : \ I \subset [a_{i-1}, a_i] \text{ for some } i\}$$

and

$$\beta_2 = \{(I, a_i) : \text{for some } i \text{ and } \mathcal{L}(I) < \epsilon/(2nM)\}$$

Check that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a,b]$.

We now verify that this β satisfies McShane's condition (Definition 9.5). Let π, π' be partitions of $[a,b]$ contained in β. We need to estimate the sum

$$\sum_{([x,y],z)\in\pi} \ \sum_{([x',y'],z')\in\pi'} |f(z_i) - f(z')| \, \mathcal{L}([x,y] \cap [x',y']). \tag{9.8}$$

If $z = a_i$ or $z' = a_i$ for some $i = 1, 2, \dots, n$ then

$$|f(z_i) - f(z')| \, \mathcal{L}([x,y] \cap [x',y']) \le 2M\mathcal{L}([x,y] \cap [x',y'])$$

and the contribution of such elements in the sum is no greater than $2Mn[\epsilon/(2nM)] = \epsilon/2$. The remaining terms in the sum (9.8) are dominated by the terms in the sum (9.7) which is smaller than $\epsilon/2$. In total then the sum (9.8) is smaller than ϵ as required to verify this condition. ■

EXERCISE 9.13. Let $\pi = (I_i, x_i)$ $(i = 1, 2, \dots, n)$ be a partition of $[a,b]$. Show that

$$\left| \sum_{i=1}^{n} f(x_i')\mathcal{L}(I_i) - \sum_{i=1}^{n} f(x_i'')\mathcal{L}(I_i) \right| \le \sum_{i=1}^{n} \omega f(I_i)\mathcal{L}(I_i)$$

for any choices of $x_i', x_i'' \in I_i$.

EXERCISE 9.14. Show that every continuous function satisfies Riemann's integrability criterion.

EXERCISE 9.15. Show that every bounded function that is continuous at almost every point in an interval $[a,b]$ satisfies Riemann's integrability criterion.

CHAPTER 10

Dini Derivatives

This chapter goes deeper into the study of the derivative. The depth is obtained, first by breaking the study of derivatives into left-handed and right-handed versions. Then a special kind of covering relation, similar to the Cousin covers, is introduced that is designed to handle one-sided derivatives.

10.1. The Dini derivatives

For many functions a closer analysis is needed than would be available using the upper and lower derivates: we require one-sided versions.

DEFINITION 10.1. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that $x \in \mathbb{R}$. Then the four values

$$\overline{D}^{+} f(x) = \inf_{\delta>0} \sup \left\{ \frac{\Delta f([x, x+h])}{\mathcal{L}([x, x+h])} : 0 < h < \delta \right\}$$
$$\underline{D}^{+} f(x) = \sup_{\delta>0} \inf \left\{ \frac{\Delta f([x, x+h])}{\mathcal{L}([x, x+h])} : 0 < h < \delta \right\}$$
$$\overline{D}^{-} f(x) = \inf_{\delta>0} \sup \left\{ \frac{\Delta f([x-h, x])}{\mathcal{L}([x-h, x])} : 0 < h < \delta \right\}$$
$$\underline{D}^{-} f(x) = \sup_{\delta>0} \inf \left\{ \frac{\Delta f([x-h, x])}{\mathcal{L}([x-h, x])} : 0 < h < \delta \right\}$$

are called the *Dini derivatives* of f at x.

EXERCISE 10.2. Show that

$$\underline{D} f(x) \leq \underline{D}^{+} f(x) \leq \overline{D}^{+} f(x) \leq \overline{D} f(x) \quad \text{and} \quad \overline{D} f(x) = \max\{\overline{D}^{-} f(x), \overline{D}^{+} f(x)\}.$$

10.2. Theorem of Grace Chisolm Young

THEOREM 10.3 (Grace Chisolm Young). Let $f : \mathbb{R} \to \mathbb{R}$. Then the sets of points

$$\{x : \overline{D}^{-} f(x) < \underline{D}^{+} f(x)\}$$

and

$$\{x : \overline{D}^{+} f(x) < \underline{D}^{-} f(x)\}$$

are both countable.

Proof. Let

$$E = \{x : \overline{D}^{-} f(x) < \underline{D}^{+} f(x)\}$$

and, for each rational number r, let

$$E_r = \{x : \overline{D}^{-} f(x) < r < \underline{D}^{+} f(x)\}.$$

Note that E is the union of the countable collection of sets E_r taken over all rationals r. For each x in E_r there is a $\delta(x) > 0$ so that, for all $0 < h < \delta(x)$,

$$\Delta f([x-h,x]) < r\mathcal{L}([x-h,x])$$

and

$$\Delta f([x,x+h) > r\mathcal{L}([x,x+h])\}$$

because of the values of the Dini derivatives.

Let

$$E_{rn} = \{x \in E : \delta(x) > 1/n\}$$

and check that

$$E_r = \bigcup_{n=1}^{\infty} E_{rn}.$$

We claim that, for each n, the set E_{rn} is countable. Indeed there cannot be two points x and y with $x < y$ in E_{rn} closer together than $1/n$. For if so, let $h = y - x$, note that $0 < h < \delta(x) < 1/n$ and $0 < h < \delta(y) < 1/n$. That would mean that

$$\Delta f([x,y]) < r\mathcal{L}([x,y]) < \Delta f([x,y)$$

which is impossible. Accordingly each E_{rn} is countable and so too also is E. The other set of the theorem can be handled by an identical proof. ∎

10.2.1. Beppo Levi Theorem.

Corollary 10.4 (Beppo Levi). *Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that f has one-sided derivatives $D^+f(x)$ and $D^-f(x)$ at each point of a set E. Then the set of points x in E at which*

$$D^+f(x) \neq D^-f(x)$$

is countable.

Exercise 10.5. It is easy to misinterpret the theorem of Beppo Levi. To avoid this construct a continuous function $f : \mathbb{R} \to \mathbb{R}$ so that for some uncountable set E the right-hand derivative $D^+f(x)$ exists at each point of E and the left-hand derivative $D^-f(x)$ fails to exist at each point of E.

10.3. Theorem of William Henry Young

The following theorem is closely related to the Grace Chisholm Young theorem. (Indeed its author was himself closely related to Grace Chisholm Young: he was her mathematics tutor and, later, her husband. They worked jointly on the general problem of left and right asymmetry of properties of real functions.)

Theorem 10.6 (William Henry Young). *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then the sets of points*

$$\{x : \overline{D}^- f(x) = \overline{D}^+ f(x)\}$$

and

$$\{x : \underline{D}^- f(x) = \underline{D}^+ f(x)\}$$

are both residual subsets of $\mathbb{R}$.

Proof. Consider first the set

$$A = \{x : \overline{D}^- f(x) < \overline{D}^+ f(x)\}$$

and, for each rational number r, let

$$A_r = \{x : \overline{D}^- f(x) < r < \overline{D}^+ f(x)\}.$$

Note that A is the union of the countable collection of sets A_r taken over all rational numbers r.

For each x in A_r we have $\overline{D}^- f(x) < r$. Thus there is a $\delta(x) > 0$ so that, for all $0 < h < \delta(x)$,

$$f(x) - f(x-h) < rh.$$

For each $n = 1, 2, 3, \dots$ and each $k = 0, \pm 1, \pm 2, \dots$ write

$$A_{rnk} = \left[\frac{k-1}{n}, \frac{k}{n}\right] \cap \left\{x \in A_r : \delta(x) > \frac{1}{n}\right\}.$$

Notice that

$$f(x) - f(y) < r(x-y)$$

for all $x < y$ with x, $y \in A_{rnk}$ and check that

$$A_r = \bigcup_{k=-\infty}^{\infty} \bigcup_{n=1}^{\infty} A_{rnk}.$$

Finally let E_{rnk} denote the closure of the set A_{rnk}. Each set E_{rnk} is compact and we claim that it contains no subinterval; in particular then it is a meager subset of $\mathbb{R}$.

Should such a set E_{rnk} contain an interval $[a, b]$ then, by the continuity of f we must conclude that the inequality stated above would require, for all $a < y < x < b$, that

$$f(x) - f(y) \leq r(x-y).$$

Consequently there would be no points y in (a, b) at which $r < \overline{D}^+ f(y)$. But this is impossible since the set A_{rnk} is dense in the set E_{rnk}.

Thus we have displayed

$$A_r \subset \bigcup_{k=-\infty}^{\infty} \bigcup_{n=1}^{\infty} E_{rnk}$$

as a subset of a union of a sequence of meager subsets of $\mathbb{R}$.

It follows that the set A defined above is also a meager subset of $\mathbb{R}$. In a similar way we can conclude that each of the sets

$$\{x : \overline{D}^- f(x) > \overline{D}^+ f(x)\}$$

$$\{x : \underline{D}^- f(x) > \underline{D}^+ f(x)\}$$

and

$$\{x : \underline{D}^- f(x) < \underline{D}^+ f(x)\}$$

is a meager subset of $\mathbb{R}$. From this the theorem follows. ∎

EXERCISE 10.7. Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Show that the set of points at which f has a right-hand derivative but no left-hand derivative is a meager subset of $[a, b]$.

10.4. Theorem of Anthony P. Morse

We continue our study of Dini derivatives by presenting a further result that reveals the structure. The preliminary lemma (which along with a meager/residual argument is the cornerstone of the proof) is due to Zygmund.

LEMMA 10.8. Let $f : [a, b] \to \mathbb{R}$ be a continuous function with $f([a, b]) = [c, d]$. Write

$$D = \{x \in [a, b] : \overline{D}^+ f(x) \leq 0\}.$$

Then, either f is nondecreasing on $[a, b]$ or else $f(D)$ contains a compact subinterval of $[c, d]$.

Proof. Suppose that f is not nondecreasing on $[a, b]$. Then we can choose points $a \leq a' < b' \leq b$ with $f(b') < f(a')$. Thus $[f(b'), f(a')]$ is a nonempty compact subinterval of $[c, d]$. Take any y between $f(b')$ and $f(a')$. Let

$$M(y) = \sup\{x \in (a', b') : f(x) = y\}.$$

Check that $f(x) = y$ and that $\overline{D}^+ f(x) \leq 0$ whenever $x = M(y)$. Thus, $y \in f(D)$. Consequently $f(D)$ contains $(f(b'), f(a'))$ and so, also, all compact subintervals contained in this open interval. ■

10.4.1. Morse's Theorem.

THEOREM 10.9 (Morse). Let $f : [a, b] \to \mathbb{R}$ be a continuous function with $f([a, b]) = [c, d]$. Write

$$A = \{x \in [a, b] : \overline{D}^+ f(x) \geq 0\},$$

$$B = \{x \in [a, b] : \overline{D}^+ f(x) < 0\},$$

and

$$C = \{x \in [a, b] : \overline{D}^+ f(x) = 0\}.$$

Suppose that A is dense in $[a, b]$. Then B is a meager subset of $[a, b]$ and $f(B)$ is a meager subset of $[c, d]$. Moreover either f is nondecreasing on $[a, b]$ or else $f(C)$ contains a residual subset of some compact subinterval of $[c, d]$.

Proof. We break the proof into a number of steps that follow Morse's original exposition.

Step 1. Suppose that f is strictly decreasing on a compact set $E \subset [a, b]$. If E contains no subinterval, then we claim that $f(E)$ is a compact subset of $[c, d]$ that also contains no interval.

Using the same ideas as used in Section 4.12 we can suppose that there is a strictly decreasing, continuous function $g : \mathbb{R} \to \mathbb{R}$ so that $f(x) = g(x)$ for all x in E. We know that $f(E) = g(E)$ would be compact. Suppose, contrary to what we want, that $g(E)$ contains a subinterval J of $[c, d]$. We consider the inverse function g^{-1} which maps that subinterval J back into E. Such a function would be continuous and therefore maps J to some interval. That would require E to contain an interval.

Step 2. Define, for each integer $n = 1, 2, 3, \ldots,$

$$E_n = \{x \in [a, b] : f(x + h) - f(x) \leq -h/n \text{ whenever } 0 \leq h \leq 1/n\}$$

Then we will prove that E_n is a compact subset of $[a, b]$ that contains no interval and that $f(E_n)$ is a compact subset of $[c, d]$ that contains no interval.

It is easy to check, using the continuity of f, that E_n is closed. Thus both E_n and $f(E_n)$ must be compact. We subdivide $[a,b]$ into a finite collection $\{J_k\}$ of compact, nonoverlapping subintervals of $[a,b]$, covering all of that interval and each of length less than $1/n$. It is easy to see that f is strictly decreasing on each set $J_k \cap E_n$. By our hypotheses the set A is dense in $[a,b]$ so that no one of these sets $J_k \cap E_n$ can contain an interval. In particular E_n itself can contain no interval. Moreover, by step 1, we conclude that $f(J_k \cap E_n)$ is a compact set that contains no subintervals of $[c,d]$. It follows that $f(E_n)$ is contained in the finite union of such sets and so must itself contain no subintervals of $[c,d]$.

Step 3. The set B is a meager subset of $[a,b]$ and the set $f(B)$ is a meager subset of $[c,d]$. This follows from step 2 since B is the union of the sequence of sets $\{E_n\}$ each of which is a meager subset of $[a,b]$, while $f(B)$ is the union of the sequence of sets $\{f(E_n)\}$, each of which is a meager subset of $[c,d]$.

Step 4. Suppose now that f is not nondecreasing on $[a,b]$. Then we can choose points $a \leq a' < b' \leq b$ with $f(b') < f(a')$. Thus $[f(b'), f(a')]$ is a nonempty compact subinterval of $[c,d]$. We know from the proof of the preliminary lemma that f maps the set

$$D = \{x \in [a,b] : \overline{D}^+ f(x) \leq 0\}$$

onto a set containing the open interval $(f(b'), f(a'))$. But we already have established that the set $f(B)$ is a meager subset of $[c,d]$. Using the fact that $B \cup C = D$, we conclude that $f(B) \cup f(C) = f(D) \supset (f(b'), f(a'))$. Thus $f(C)$ must contain a residual subset of the interval $[f(b'), f(a')]$. ■

10.4.2. Darboux property of Dini derivatives.

EXERCISE 10.10 (A Darboux property of Dini derivatives). Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and suppose that the Dini derivative $\overline{D}^+ f(x)$ is unbounded both above and below on each interval. Show, for every real number r and compact interval $[a,b]$, that f maps the set

$$E_r = \{x \in [a,b] : \overline{D}^+ f(x) = r\}$$

onto a residual subset of some compact interval. (In particular $\overline{D}^+ f(x)$ assumes every real number at many points in any subinterval.)

10.5. Measure properties of Dini derivatives

The Dini derivatives have a property similar to Theorem 7.14, but here we must assume that the function is continuous.

LEMMA 10.11. For any continuous function $f : \mathbb{R} \to \mathbb{R}$ and any real number r the sets

$$\{x : \overline{D}^+ f(x) \leq r\} \text{ and } \{x : \underline{D}^+ f(x) \leq r\}$$

are almost closed.

Proof. For example, consider the set

$$E = \{x : \overline{D}^+ f(x) < r\}$$

and write, for positive integers m and n,

$$E_{mn} = \{x : \ f(x+t) - f(x) - rt + t/m \leq 0 \text{ for all } 0 \leq t \leq 1/n\}.$$

Since f is continuous, we can check that each set E_{mn} is closed. But

$$E = \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} E_{mn}$$

reveals that E must be almost closed. ■

10.6. Quasi-Cousin covers

We require a variant on the notion of a Cousin cover that is more appropriate for handling Dini derivatives.

10.6.1. Motivation. The Cousin covers are particularly suited to describing properties of the ordinary derivative. For example if $\underline{D}F(x) > r$ then the covering relation

$$\beta = \{(I, x) : \Delta F(I) > r\mathcal{L}(I)$$

has the property that for some $\delta > 0$, if $x \in I$ and $\mathcal{L}(I) < \delta$ then necessarily $(I, x) \in \beta$. Indeed $\underline{D}F(x) > r$ if and only if β has this property.

In this chapter we investigate Dini derivatives and so will require a one-sided analogue. The simplest version could come from the observation that $\underline{D}^+F(x) > r$ if and only if the covering relation

$$\beta = \{(I, x) : \Delta F(I) > r\mathcal{L}(I)$$

has the property that for some $\delta > 0$, if $0 < h < \delta$ then necessarily $([x, x+h], x) \in \beta$.

But in fact our covering relation needs to be designed to handle the upper Dini derivative, not the lower. For that the description is more delicate: $\overline{D}^+F(x) > r$ if and only if the covering relation

$$\beta = \{(I, x) : \Delta F(I) > r\mathcal{L}(I)\}$$

has the property that for any $\epsilon > 0$, there is at least one value of h with $0 < h < \epsilon$ for which $([x, x + h], x) \in \beta$. We strengthen this by insisting that F is continuous. In that case, if we found h so that

$$\frac{F(x + h) - F(x)}{(x + h) - x} > r,$$

notice that there must be a $\delta > 0$ so that

$$\frac{F(x' + h) - F(x)}{(x' + h) - x} > r$$

for every value of x' in the interval $[x - \delta, x]$.

10.6.2. Quasi-Cousin cover.

Definition 10.12. Let K be a compact set with endpoints $a = \inf K$ and $b = \sup K$. A covering relation β is said to be a *quasi-Cousin cover* of K provided that

(a) There is at least one pair $([a, d], a) \in \beta$ with $a < d \leq b$.
(b) For every $a < x < b$, $x \in K$ there is a $\delta > 0$ so that there is at least one $x < d \leq b$ for which all pairs $([c, d], x) \in \mathcal{S}$ whenever $x - \delta < c \leq x$.
(c) There is a $\delta > 0$ so that all pairs $([c, b], b) \in \beta$ whenever $b - \delta < c < b$.

EXERCISE 10.13. Suppose that $F : \mathbb{R} \to \mathbb{R}$ is a continuous function and that $\overline{D}^+ F(x) > r$ at every point of an interval $[a, b]$. Verify that the covering relation

$$\beta = \{(I, x) : \Delta F(I) > r\mathcal{L}(I)$$

satisfies the first two conditions (but not necessarily the third) in Definition 10.12.

EXERCISE 10.14. Continuing the previous exercise, let $\epsilon > 0$ and let

$$\beta' = \{([x-t, x)], x) : |F(x-t) - F(x)| < \epsilon\}.$$

Show that $\beta \cup \beta'$ is a quasi-Cousin cover of $[a, b]$.

EXERCISE 10.15. Show that every Cousin cover of an interval $[a, b]$ is also a quasi-Cousin cover for any compact subset of $[a, b]$.

10.7. Quasi-Cousin Covering Lemma

Even though the notion of a quasi-Cousin cover is much weaker than that of a Cousin cover the covering lemma generalizes.

LEMMA 10.16 (Quasi-Cousin covering lemma). Let β be a quasi-Cousin cover of a compact set K with endpoints $a = \inf K$ and $b = \sup K$. Then β contains a subpartition π so that

$$K \subset \bigcup_{(I,x)\in\pi} I \subset [a, b].$$

Proof. Let us assume first that $K = [a, b]$. Let E be the set of all points z, with $b \geq z > a$ and with the property that β contains a partition π of $[a, z]$.

Argue that (i) $E \neq \emptyset$, (ii) if $\sup E = t$ then t cannot be less than b, (iii) if $\sup E = b$ then $b \in E$.

We know that (i) is true since there is at least one pair $([a, d], a) \in \beta$ with $a < d \leq b$ and so $d \in E$. Thus we may set $t = \sup E$ and be assured that $a < d \leq t \leq b$. To see (ii) note that it is not possible for $t < b$ for if so then there is a $\delta > 0$ and $d' > t$ for which all pairs $(t, [c, d']) \in \beta$ with $t - \delta < c \leq t$. But that supplies a point $t' \in (c - \delta, c] \cap E$ and the partition of $[a, t']$ can be enlarged by including $(t, [t', d'])$ to form a partition of $[a, d']$; thus $d' \in E$. But this violates $t = \sup E$.

Finally for (iii) if $t = b$ and yet $b \notin E$ then, repeating much the same argument, there is a $\delta > 0$ for which all pairs $(b, [c, b]) \in \beta$ with $b - \delta < c < b$. But that supplies a point $t' \in (b - \delta, b) \cap E$ and the partition for $[a, t']$ can be enlarged by including $(b, [t', b])$ to form a partition π for $[a, b]$. This shows that $b \in E$ after all.

Now let us handle the general case for an arbitrary compact set $K \subset [a, b]$. Let $G = (a, b) \setminus K$ and

$$\beta_1 = \{(I, x) : \ x \in I \text{ and } I \subset G\}.$$

Since β is a quasi-Cousin cover of K we can check that $\beta \cup \beta_1$ is a quasi-Cousin cover of $[a, b]$. By the first part of the proof there is a partition $\pi \subset \beta \cup \beta_1$ of $[a, b]$. Remove those elements of π that do not belong to β to form a subpartition with exactly the required properties. ■

The proof contains explicitly the statement of the corollary:

COROLLARY 10.17. Let β be a quasi-Cousin cover of a compact interval $[a, b]$. Then β contains a partition of $[a, b]$ (although not necessarily of other subintervals of $[a, b]$).

10.7.1. Another variant. For some arguments it is awkward to arrange for the cover to have exactly the correct properties at the right end-point of the compact set. Thus occasionally we shall use the following very minor variant of the quasi-Cousin covering lemma. Note that the subpartition here may extend past the interval $[a, b]$.

LEMMA 10.18. Let K be a compact set and β a covering relation. Suppose that, for each $x \in K$, there are s, $t > 0$ so that all pairs

$$([x', x+s], x) \in \beta$$

whenever $x - t \leq x' \leq x$. Then β contains a subpartition π for which

$$K \subset \bigcup_{(I,x)\in\pi} I.$$

10.8. Growth properties from Dini derivatives

In Section 5.5 we developed a number of growth implications arising from upper and lower derivatives. Exactly the same kind of conditions can be proved for the Dini derivatives under appropriate assumptions.

THEOREM 10.19. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous at each point of an open interval (a, b) and suppose that $\overline{D}^+ f(x) > m$ for each $x \in (a, b)$. Then $f(d) - f(c) > m(d-c)$ for each $[c, d] \subset (a, b)$.

Proof. The student should review the proof of Theorem 5.20. The proof is nearly identical but using quasi-Cousin covers and exploiting the continuity assumption. ∎

We leave it to the reader to formulate and prove versions of this theorem that would allow countable or null exceptional sets.

EXERCISE 10.20. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that the function $\overline{D}^+ f(x)$ is finite-valued and continuous at a point x_0. Show that f is differentiable at x_0.

10.9. Estimates of integrals from Dini derivatives

For Dini derivatives there is a weaker version of Theorem 6.12 available using similar arguments (but employing quasi-Cousin covers as well as Cousin covers). Note that this weaker version uses lower and upper rather than upper and lower integrals; in particular no corollary can be derived asserting the integrability of the Dini derivative (indeed it may not be integrable).

THEOREM 10.21. Suppose that $F : [a, b] \to \mathbb{R}$ is continuous and that g is a finite-valued function. If $\overline{D}^+ F(x) \geq g(x)$ at every point $a < x < b$, then,

$$F(b) - F(a) \geq \underline{\int_a^b} g(x)\,dx. \tag{10.1}$$

If $\underline{D}^+ F(x) \leq g(x)$ at every point $a < x < b$, then

$$F(b) - F(a) \leq \overline{\int_a^b} g(x)\,dx. \tag{10.2}$$

Proof. Let $\epsilon > 0$. Take the covering relation β_1 of all pairs $([x,y],z)$ with

$$\Delta F([x,y]) \geq (f(z) - \epsilon)\mathcal{L}([x,y])$$

and β_2 of all pairs $([a,y],a)$ and $([x,b],b)$ for which

$$\Delta F([a,y]) - f(a)\mathcal{L}([a,y]) > -\epsilon$$

and

$$\Delta F([x,b]) \geq f(b)\mathcal{L}([x,b]) - \epsilon.$$

It is easy to verify that $\beta = \beta_1 \cup \beta_2$ is a quasi-Cousin cover of $[a,b]$. At the endpoints a or b the continuity of F needs to be used in the verification, while at the points in (a,b) the inequality $\overline{D}^+ F(x) \geq g(x)$ is used.

This may not seem too much of a help since the integral is defined by Cousin covers, not by quasi-Cousin covers. But let β_3 be *any* Cousin cover of $[a,b]$. Check that, as defined, $\beta_3 \cap \beta$ must be a quasi-Cousin cover of $[a,b]$. Thus there is at least one partition π from β_3 that is also in β. For that partition a familiar argument gives us

$$\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) \leq \sum_{(I,x)\in\pi} [\Delta F(I) + \epsilon\mathcal{L}([x,y])] + 2\epsilon = F(b) - F(a) + \epsilon(2 + b - a).$$

Note that this means any Cousin cover of $[a,b]$ contains at least one partition π with this property. Thus, while we can say nothing about the *upper* integral, we certainly can assert that the *lower* integral must always be lesser than $F(b)-F(a)+\epsilon(2+b-a)$ and from this the theorem follows. ■

As a consequence of this theorem we observe that if an everywhere finite function g is *assumed* to be integrable on $[a,b]$ and lies between the two derivates then an integral identity holds. The assumption that g is integrable cannot be dropped here.

COROLLARY 10.22. Let $F : [a,b] \to \mathbb{R}$ be continuous and $g : [a,b] \to \mathbb{R}$ be integrable on $[a,b]$ and suppose that

$$\underline{D}^+ F(x) \leq g(x) \leq \overline{D}^+ F(x)$$

at every point x on $[a,b]$. Then

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

10.10. Growth lemmas on compact sets

In all of the growth lemmas we have studied we have obtained comparisons between the measure of the interval $[a,b]$ and the increment $\Delta f([a,b])$. We now generalize this to arbitrary compact sets. Let f be continuous and K a compact set. What can we deduce about the relation between $\mathcal{L}(K)$ and the measure of the image, $\mathcal{L}(f(K))$, if we are given some information about the derivates? [Remember $f(K)$ is compact if f is continuous and K is compact.]

Our two lemmas use the Dini derivatives to obtain the comparison. Many other variants are possible.

LEMMA 10.23. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, nondecreasing function. Let K be a compact subset with the property that $\underline{D}^+ f(x) < p$ for each $x \in K$. Then

$$\mathcal{L}(f(K)) \leq p\mathcal{L}(K) \tag{10.3}$$

Proof. To prove this we exploit the geometry of the derivative situation, expressed in the following covering relation:

$$\beta_p = \left\{(I,z) : z \in K \cap I, \quad \frac{\Delta f(I)}{\mathcal{L}(I)} < p\right\}.$$

The only detailed work that needs checking here is to verify that the covering relation has the property expressed in Lemma 10.18. This is true directly from the fact that $\underline{D}^+ f(x) < p$ at each point $x \in K$ and the continuity of f at x.

Let G be any open set containing K. Prune out unnecessary parts of β_p by defining

$$\beta_p' = \{(I,x) \in \beta_p : I \subset G\}.$$

This does not change the covering properties and so we can use Lemma 10.18. Thus there is a subpartition $\pi \subset \beta_p'$ covering K in the sense that allows us to compute

$$\mathcal{L}(f(K)) \leq \sum_{(I,x)\in\pi} \Delta f(I) \leq \sum_{(I,x)\in\pi} p\mathcal{L}(I) \leq p\mathcal{L}(G).$$

Since this is true for all such open sets G we deduce (10.3). ■

Lemma 10.24. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, nondecreasing function. Let K be a compact subset with the property that $\overline{D}^+ f(x) > q$ for each $x \in K$. Then

$$\mathcal{L}(f(K)) \geq q\mathcal{L}(K) \tag{10.4}$$

Proof. To prove this second lemma we consider the covering relation:

$$\beta_q = \left\{([x,y],z) : z \in K \cap [x,y], \frac{\Delta f(I)}{\mathcal{L}(I)} > q\right\}.$$

Check that Lemma 10.18 can applied, as before:, using the fact that $\overline{D}^+ f(x) \geq q$ at each point $x \in K$ and the continuity of f at x.

Let G be any open set containing $f(K)$ and prune out unnecessary parts of β_q by defining

$$\beta_q' = \{(I,x) \in \beta_q : I \subset f^{-1}(G)\}.$$

This does not change the covering properties and so by Lemma 10.18 again there is a subpartition $\pi \subset \beta_q'$ covering K in the sense that we will be able to compute

$$q\mathcal{L}(K) \leq \sum_{\pi} q\mathcal{L}(I) \leq \sum_{\pi} \Delta f(I) \leq \mathcal{L}(G).$$

Since this is true for all such open sets G we deduce (10.4). ■

From these lemmas we deduce quite easily an important fact about the differentiation structure of continuous monotonic functions. We use this in Section 10.11 to show that such functions have a derivative almost everywhere.

Corollary 10.25. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, nondecreasing function and let K be a compact set. Suppose that at every point $x \in K$

$$\underline{D}^+ f(x) < p < q < \overline{D}^+ f(x).$$

Then both K and $f(K)$ are null sets.

Proof. Applying Lemmas 10.23 and 10.24, we obtain

$$\mathcal{L}(f(K)) \leq p\mathcal{L}(K) < q\mathcal{L}(K) \leq \mathcal{L}(f(K)).$$

This would be possible only if $\mathcal{L}(K) = \mathcal{L}(f(K)) = 0$ and so these two sets are null sets as we require. ■

COROLLARY 10.26. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, nondecreasing function and let E be an arbitrary almost closed set. Suppose that at every point $x \in E$

$$\underline{D}^+ f(x) < p < q < \overline{D}^+ f(x).$$

Then E is a null set.

Proof. Since E is an almost closed set, we may use the principle from Section 7.8 to test that the set is null: E is null if all its compact subsets are null. But the previous corollary (Corollary 10.25) shows that, indeed, all compact subsets of E are null. ■

10.11. The Lebesgue differentiation theorem

The growth properties allow us to prove a differentiation theorem that is central to the calculus program. It asserts that monotonic functions are differentiable at most points, indeed at almost every point.

THEOREM 10.27. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous monotonic function. Then f has a derivative almost everywhere.

Proof. We prove first that the set of points where there is no right-hand derivative is a null set. That means we are analysing the set

$$E = \{x : \underline{D}^+ f(x) < \overline{D}^+ F(x)\}.$$

At each point $x \in E$ there must exist a pair of rational numbers p and q for which

$$\underline{D}^+ f(x) < p < q < \overline{D}^+ F(x).$$

Consequently the set E can be expressed as the countable union of the collection of the sets

$$E_{pq} = \{x : \underline{D}^+ f(x) < p < q < \overline{D}^+ F(x)\}$$

taken over all rational numbers $p < q$. If each of these is a null set then so too is E.

Using Lemma 10.11, we can prove that each such set E_{pq} is almost closed. Thus we may apply Corollary 10.26 to deduce that E_{pq} is null. It follows now that E is null. In other words, almost everywhere,

$$\underline{D}^+ f(x) = \overline{D}^+ F(x).$$

We must still consider the possibility that

$$\underline{D}^+ f(x) = \overline{D}^+ F(x) = \infty$$

and we now show that the set of these points is also null. Define

$$\{x : A = \underline{D}^+ f(x) = \infty\}. \tag{10.5}$$

Exactly as before this set can be shown to be almost closed. We use the principle from Section 7.8 to test that the set is null: A is null if all its compact subsets are null. Every compact subset K of A must satisfy

$$n\mathcal{L}(K) \leq \mathcal{L}(f(K)) \leq f(b) - f(a)$$

for all integers n, because of Lemma 10.24. But this can hold only if $\mathcal{L}(K) = 0$. Consequently, A is a null set.

In summary what we know so far is that the function f has a finite right-hand derivative almost everywhere. Applying that observation to the function $-f(-x)$ shows that, in addition, the function f has a finite left-hand derivative almost everywhere in $[a, b]$.

The theorem of Beppo Levi finishes off the proof because left and right derivatives can differ only on a countable set. ■

10.12. Differentiation of the integral

So far we have studied the relation between the derivative and the integral in one direction only. If $F' = f$ then we can integrate f to recover the increment of F. The indefinite integral of f is exactly F.

Is the converse true? If F is the indefinite integral of f can we deduce that $F'(x) = f(x)$? At points of continuity of f this is pretty easy to manage. We need to do better.

The differentiation of the integral in general requires some of the heavy machinery of the more advanced chapters. Provided we are allowed to restrict the function f to be an almost continuous function, our elementary covering lemmas allow us to answer the question. Note that this extra hypothesis is a burden to us (since we would prefer to prove that integrable functions are indeed almost continuous anyway); this is not a burden to the Lebesgue program since, in that program, all functions to be integrated are assumed in advance to be almost continuous.

THEOREM 10.28. *Let $f : \mathbb{R} \to \mathbb{R}$ be an almost continuous function that is integrable on an interval $[a, b]$ and let F be its indefinite integral. Then $F'(x) = f(x)$ for almost every x in $[a, b]$.*

Proof. We show that $\overline{D}^+F(x) = \underline{D}^+F(x) = f(x)$ outside of a null set. The set of points where $\underline{D}^+F(x) < f(x)$,

$$E = \{x \in [a, b] : \underline{D}^+F(x) < f(x)\}$$

can be analysed by considering the sets

$$E_{pq} = \{x \in [a, b] : \underline{D}^+F(x) < p < q < f(x)\}$$

for rational numbers $p < q$. To prove that E is a null set, it is enough to show that each set E_{pq} is a null set, since then E is the union of this countable collection of null sets.

We need to know that each of these sets E_{pq} is almost closed. Since f is almost continuous we know that

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

is almost closed. We also know that for Dini derivatives of continuous functions the set

$$\{x \in [a, b] : \underline{D}^+F(x) < p\}$$

is almost closed. But E_{pq} is the intersection of these two almost closed sets and so is almost closed as well. Accordingly then, to show E_{pq} is a null set, it suffices (because of Lemma 7.8.1) to show that every compact subset of E_{pq} is null.

To this end, let $\epsilon > 0$ and select β_0 to be a Cousin cover of $[a, b]$ chosen so that the Henstock criterion in Section 9.1 is satisfied, i.e., so that

$$\sum_{(I,x)\in\pi} |f(x)\mathcal{L}(I) - \Delta F(I)| < \epsilon.$$

for any choice of subpartition π from β_0.

Let K be any compact subset of E_{pq} and let

$$\beta_1 = \{(I, x) : x \in K,\ \ \Delta F(I) < p\mathcal{L}(I) < q\mathcal{L}(I) < f(x)\mathcal{L}(I)\}.$$

The hypotheses of Lemma 10.18 can be verified by using the inequalities $\underline{D}^+F(x) < p < q < f(x)$ and the continuity of F at each point. The intersection $\beta_0 \cap \beta_1$ also satisfies these same hypotheses. Select then, by Lemma 10.18, a subpartition π from $\beta_0 \cap \beta_1$ covering K in a manner allowing this computation:

$$(q-p)\mathcal{L}(K) \le (q-p)\sum_{(I,x)\in\pi} \mathcal{L}(I) \le \sum_{(I,x)\in\pi} [q\mathcal{L}(I) - p\mathcal{L}(I) \le$$

$$\sum_{(I,x)\in\pi} |f(x)\mathcal{L}(I) - \Delta F(I)| < \epsilon.$$

As $\epsilon > 0$ is arbitrary and $q - p > 0$ it follows that $\mathcal{L}(K) = 0$. Since this is true for any compact set $K \subset E_{pq}$ our principle requires E_{pq} to be a null set. As announced above, this shows that the set of points where $\underline{D}^+F(x) < f(x)$ 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

$$\underline{D}^+F(x) = \overline{D}^+F(x) = f(x).$$

This result can be applied to the functions $f(-x)$ and $F(-x)$ to show that, again outside of a null set

$$\underline{D}^-F(x) = \overline{D}^-F(x) = f(x).$$

In particular, combining these facts, we have proved the theorem. ■

EXERCISE 10.29. Let f be integrable on an interval $[a, b]$ and let F be its indefinite integral. Then $F'(x) = f(x)$ for every x in $[a, b]$ that is a point of continuity of f.

CHAPTER 11

The Vitali Covering Theorem

We turn now to a deeper study of the classical measures of Lebesgue. Our presentation uses covering arguments and will require an extension of our repertoire. The *fine covers* that we now introduce comprise a notion dual to that of a full cover that can be motivated in much the same way as the quasi-Cousin covers.

11.1. Full and fine covers

Full covers contain *all sufficiently small* intervals near a point, while fine covers contain always *arbitrarily small* intervals near the point. We recognize the full covers as closely related to the Cousin covers. The fine covers are defined as a dual concept, originating from work of Vitali in the early 20th century. We repeat the definition of full cover and follow it with the very similar but contrasting fine cover.

DEFINITION 11.1. A covering relation β is said to be a *full cover* for a set E if for every $z \in E$ there is a $\delta > 0$ so that every pair $([x, y], z)$ with $0 < y - x < \delta$ and $z \in [x, y]$ must belong to β.

DEFINITION 11.2. A covering relation β is said to be a *fine cover* for a set E if for every $z \in E$ and every $\delta > 0$ there must exist at least one pair $([x, y], z)$ with $0 < y - x < \delta$ and $z \in [x, y]$ that belongs to β.

11.1.1. Exercises.

EXERCISE 11.3. Show that a full or a fine cover β of a set E ignores no point of E in the sense that, for all $x \in E$, there is at least one pair (I, x) belonging to β.

EXERCISE 11.4. Show that if β is a full cover of a set E then β is a fine cover E.

EXERCISE 11.5. Show that if β is a full cover of a compact set K then β is a Cousin cover for K. Is the converse true?

11.1.2. Motivation for fine covers. The Cousin covers and full covers are particularly suited to describing properties of the ordinary derivative. For example if $\underline{D}F(x_0) > r$ then the covering relation

$$\beta = \{(I, x) : \Delta F(I) > r\mathcal{L}(I)\}$$

has the property that for some $\delta > 0$, if $x_0 \in I$ and $\mathcal{L}(I) < \delta$ then necessarily $(I, x_0) \in \beta$. Thus β is a full cover of the set $\{x : \underline{D}F(x) > r\}$.

On the other hand, for the upper derivate, if $\overline{D}F(x_0) > r$ then this same covering relation has the property that for any $\epsilon > 0$, there is at least one interval I containing x_0 for which $\mathcal{L}(I) < \epsilon$. Thus this same β is a fine cover of the larger set $\{x : \overline{D}F(x) > r\}$.

11.1.3. Properties of full/fine covers.

LEMMA 11.6 (Full filtering property). Let β_1 and β_2 be full covers of a set E. If a covering relation

$$\beta \supset \beta_1 \cap \beta_2,$$

then β is also a full cover of E.

LEMMA 11.7 (Fine filtering property). Let β_1 be a full cover of a set E and let β_2 be a fine cover of E. If a covering relation

$$\beta \supset \beta_1 \cap \beta_2,$$

then β is a fine cover of E.

11.1.4. Prunings. By a *pruning* of a covering relation β we mean a device to select a subset. If E is a real set we write

$$\beta(E) = \{(I, x) \in \beta : I \subset E\}$$

and

$$\beta[E] = \{(I, x) \in \beta : x \in E\}.$$

LEMMA 11.8 (Pruning by open sets). Let $E \subset G$ where G is open. If β is a full [fine] cover of E, then so too is $\beta(G)$.

LEMMA 11.9 (Assembling the pieces). Let $\{E_i\}$ be a sequence of sets with $E \subset \bigcup_{i=1}^{\infty} E_i$ and let β_i be a sequence of covering relations. Suppose that each β_i is a full [fine] cover of E_i. Then $\bigcup_{i=1}^{\infty} \beta_i[E_i]$ is a full [fine] cover of E.

EXERCISE 11.10. If G is open there is full cover β of $\mathbb{R}$ with the property that $\beta[G] = \beta(G)$, i.e., that for all $(I, x) \in \beta$, if $x \in G$ then necessarily $I \subset G$.

11.1.5. Decomposition of full covers. There is a decomposition of full covers that is often of use in constructing a proof. Here is a good place to put it for easy reference, although it is entirely unmotivated for the moment.

LEMMA 11.11 (Decomposition Lemma). Let β be a full cover of a set E. Then there is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E_n so that if x is any point in E_n and I is any subinterval of I_{kn} that contains x then (I, x) belongs to $\beta([E_n \cap I_{kn}])$.

Proof. Let β be a full cover of a set E. By the nature of the cover there must exist a positive function δ on E with the property that (I, x) belongs to β whenever if $x \in E$, $x \in I$ and $\mathcal{L}(I) < \delta(x)$. Define

$$E_n = \{x \in E : \delta(x) > 1/n\}.$$

This is an expanding sequence of subsets of E whose union is E itself. If I is any compact interval that contains a point x in E_n and has length $\mathcal{L}(I) \leq 1/n$, then (I, x) must belong to β.

A way of exploiting this property is to introduce the intervals

$$I_{mn} = [m/n, (m + 1)/n)]$$

for integers $m = 0, \pm 1, \pm 2, \ldots$. Then $\beta([E_n \cap I_{mn}])$ has this property: if x is any point in $E_n \cap I_{mn}$ and I is any subinterval of I_{mn} that contains x then (I, x) is a member of $\beta([E_n \cap I_{mn}])$.

Thus the condition of being a full cover, which is a local condition defined in a special way at each point, has been made uniform throughout each piece of the decomposition. If we relabel these sets in a convenient way then we now have our decomposition property. ■

11.2. Variation over covering relations

Let h be a real-valued function defined on covering relations. Sums of the form

$$\sum_{(I,x)\in\pi} h(I,x)$$

where π is a partition or a subpartition have been and continue to be central to our concerns.

DEFINITION 11.12 (Variation over a subpartition). If π is a subpartition then

$$V(h,\pi) = \sum_{(I,x)\in\pi} |h(I,x)|$$

is called the variation of h over the subpartition.

DEFINITION 11.13 (Variation over a covering relation)). If β is a covering relation then

$$V(h,\beta) = \sup_{\pi\subset\beta} V(h,\pi),$$

where the supremum is taken over all subpartitions π contained in β, is called the variation for h over β.

In this chapter we shall make much use of special cases of these expressions:

$$V(\mathcal{L},\pi) = \sum_{(I,x)\in\pi} \mathcal{L}(I)$$

and, for a function $f:\mathbb{R}\to\mathbb{R}$,

$$V(f\mathcal{L},\pi) = \sum_{(I,x)\in\pi} |f(x)|\mathcal{L}(I).$$

EXERCISE 11.14. Let β be a covering relation and $h:\beta\to\mathbb{R}$. If $\{I_k\}$ is a sequence of nonoverlapping subintervals of an interval I (open or closed) then show that

$$\sum_{k=1}^{\infty} V(h,\beta(I_k)) \le V(h,\beta(I)).$$

11.3. Variational Measures

The variational measures are central to the calculus study.

DEFINITION 11.15 (Full and Fine Variations). Let h be a real-valued function defined on covering relations and let E be any set of real numbers. Then we define the full and fine variational measures associated with h by the expressions:

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

and

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

LEMMA 11.16. Let h be a real-valued function defined on covering relations. Then, for any set E,

$$V_*(h, E) \leq V^*(h, E).$$

Proof. This follows from the fact that every full cover is fine. ■

11.4. Subadditive property of the variation

LEMMA 11.17 (Subadditivity property). Let h_1 and h_2 be real-valued functions defined on covering relations. Then, for any set E,

$$V_*(h_1 + h_2, E) \leq V_*(h_1, E) + V^*(h_2, E)$$

and

$$V^*(h_1 + h_2, E) \leq V^*(h_1, E) + V^*(h_2, E).$$

Proof. Note that there is no typo in the first inequality: the full variation is needed on the right-hand side. The second inequality, the easier to check, follows from the fact that the intersection of two full covers is again full. The first inequality follows from the fact that the intersection of two covering relations, one of which is full and the other fine, is again fine. ■

11.5. Measure properties of the variation

LEMMA 11.18 (Countable subadditive property). Let h be a real-valued function defined on covering relations. Let E, E_1, E_2, ... be a sequence of subsets of $\mathbb{R}$ for which $E \subset \bigcup_{i=1}^{\infty} E_i$. Then

$$V^*(h, E) \leq \sum_{i=1}^{\infty} V^*(h, E_i)$$

and

$$V_*(h, E) \leq \sum_{i=1}^{\infty} V_*(h, E_i).$$

Proof. The proof is identical to that for Theorem 4.10. ■

LEMMA 11.19 (Additive over separated sets). Let h be a real-valued function defined on covering relations. Let E_1 and E_2 be sets that are separated by open sets (e.g., E_1 and E_2 are disjoint closed sets). Then

$$V^*(h, E_1 \cup E_2) = V^*(h, E_1) + V^*(h, E_2)$$

and

$$V_*(h, E_1 \cup E_2) = V_*(h, E_1) + V_*(h, E_2).$$

Proof. From the preceding lemma we know that

$$V^*(h, E_1 \cup E_2) \leq V^*(h, E_1) + V^*(h, E_2).$$

Let us prove the opposite direction. Let β be any full cover of $E_1 \cup E_2$. Select G_1 and G_2, disjoint open sets containing E_1 and E_2 (respectively). Then $\beta(G_1 \cup G_2)$ is necessarily a full cover of $E_1 \cup E_2$. Note that $\beta(G_1)$ is a full cover of E_1 and that $\beta(G_2)$ is a full cover of E_2. Now check that

$$V^*(h, E_1) + V^*(h, E_2) \leq V(h, \beta(G_1)) + V(h, \beta(G_2)) = V(h, \beta(G_1 \cup G_2)) \leq V(h, \beta).$$

This is true for all such full covers β of $E_1 \cup E_2$ from which the inequality

$$V^*(h, E_1) + V^*(h, E_2) \leq V^*(h, E_1 \cup E_2)$$

follows, proving the identity. A similar argument handles the fine variation. ■

Corollary 11.20 (Countable additive property). Let h be a real-valued function defined on covering relations. Let E_1, E_2, E_3, … be a sequence of disjoint closed subsets of $\mathbb{R}$. Then

$$V^*\left(h, \bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} V^*(h, E_i)$$

and

$$V_*\left(h, \bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} V_*(h, E_i).$$

11.6. The Vitali covering theorem

The general version of this theorem is the assertion, that under certain conditions on a function h, the two measures

$$V_*(h, E) = V^*(h, E)$$

agree on every set. We say that the function h has the *Vitali property* when this holds. The classical case, proved later on is the following assertion:

> Let $F : \mathbb{R} \to \mathbb{R}$. A necessary and sufficient condition for ΔF to have the Vitali property is that F is continuous and has σ-finite variation.

The particular case $F(x) = x$ is the original version; it asserts that Lebesgue's measure can be characterized by either full or fine covers. This is the content of the next section. Our strategy of proof is to obtain this result first and then to lift it to the most general class of functions that we can.

11.7. Full and fine versions of Lebesgue measure

We write for any real set E

$$\mathcal{L}^*(E) = V^*(\mathcal{L}, E)$$

and

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

We consider these measures as versions of Lebesgue measure and we will show eventually that $\mathcal{L}$, $\mathcal{L}^*$, and $\mathcal{L}_*$ are identical. We know already from an earlier chapter that $\mathcal{L} = \mathcal{L}^*$.

11.8. Radó Covering Lemma

Towards the end of relating the measure $\mathcal{L}$ to the two variational versions of Lebesgue measure using full and fine covers we first present a general covering theorem due to T. Radò from 1928.

LEMMA 11.21 (Radó Covering Lemma). Let β be a covering relation that ignores no point in a set E of finite Lebesgue measure. Then there is a subpartition $\pi \subset \beta$ so that

$$\mathcal{L}(E) \leq 3\left\{\sum_{(I,x)\in\pi} \mathcal{L}(I)\right\}.$$

Proof. Let

$$G_0 = \bigcup \{(c,d) : ([c,d],x) \in \beta\}$$

We 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 $1/n$. Thus no two points of C_n can be closer together than $1/n$. It follows that each C_n is countable. Consequently C is countable.

Since G_0 is open, by Exercise 2.13, 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 we assume that $\mathcal{L}(E)$ is finite and $\mathcal{L}(E) \leq \mathcal{L}(G_0)$, there must exist an integer N large enough so that

$$\mathcal{L}(E) \leq \frac{3}{2}\mathcal{L}\left(\bigcup_{i=1}^{N} (a_i,b_i)\right).$$

From the finite sequence, $\{([a_i,b_i],x_i)\}$ $(i = 1,2,3,\dots,N)$, discard all redundant elements, i.e., elements 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

$$\pi_1 = \{([a_i,b_i],x_i) :\ i \text{ is odd}\}$$

and

$$\pi_2 = \{([a_i,b_i],x_i) :\ i \text{ is even}\}.$$

By the way that this has been constructed both of these are subpartitions and

$$\frac{2}{3}\mathcal{L}(E) \le \frac{2}{3}\mathcal{L}(G_0) \le \mathcal{L}\left(\bigcup_{i=1}^{N}(a_i, b_i)\right) \le \sum_{(I,x)\in\pi_1} \mathcal{L}(I) + \sum_{(I,x)\in\pi_2} \mathcal{L}(I).$$

The larger of these two sums on the right-hand side of the inequality finishes off the proof. ■

11.9. Vitali Covering Theorem

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

THEOREM 11.9.1 (Vitali Covering Theorem).

$$\mathcal{L} = \mathcal{L}_* = \mathcal{L}^*.$$

We already know that

$$\mathcal{L} = \mathcal{L}^* \ge \mathcal{L}_*.$$

Thus the proof of the Vitali covering theorem can be obtained by verifying the inequality, for any set E,

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

11.9.1. Vitali covering theorem. Perhaps the deepest theorem of the calculus is this connection between these three versions of Lebesgue's measure. These two measures are expressed by a duality; the fact that they are equal is essential to much of the theory. We shall first show that the theorem is equivalent to 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.

LEMMA 11.22 (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 $\pi \subset \beta$ so that*

$$\mathcal{L}\left(E \setminus \bigcup_{(I,x)\in\pi} I\right) < \epsilon. \tag{11.1}$$

Proof. The proof appears in Section 11.9.2. The remaining exercises in the section show that geometrical form is equivalent to the measure form $\mathcal{L} = \mathcal{L}_* = \mathcal{L}^*$. ■

EXERCISE 11.23. Show that the identity

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

implies Lemma 11.22.

Hint: Let $\epsilon > 0$. Choose G open so that $E \subset G$ and $\mathcal{L}(G) < \mathcal{L}(E) + \epsilon/2$. Let β be an arbitrary fine cover of E; then $\beta(G)$ is also a fine cover of E. There must be a subpartition $\pi \subset \beta(G)$, so that

$$\mathcal{L}_*(E) \le V(\mathcal{L}, \beta(G)) < V(\mathcal{L}, \pi) + \epsilon/2.$$

Check that

$$\mathcal{L}\left(G \setminus \sigma(\pi)\right) = \mathcal{L}(G) - \sum_{(I,x)\in\pi} \mathcal{L}(I).$$

Use the hypothesis of the exercise to check that

$$\mathcal{L}\left(E \setminus \bigcup_{(I,x)\in\pi} I\right) \le \mathcal{L}\left(G \setminus \bigcup_{(I,x)\in\pi} I\right) \le \mathcal{L}_*(E) + \epsilon/2 - \sum_{(I,x)\in\pi} \mathcal{L}(I) < \epsilon.$$

Exercise 11.24. Deduce the identity

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

for arbitrary sets, granted that it holds for sets of finite Lebesgue measure.

Exercise 11.25. Deduce Theorem 11.9.1 from Lemma 11.22.

Hint: For any set of real numbers E for which $\mathcal{L}(E) < \infty$ and any fine cover β of E select $\pi \subset \beta$ so that (11.1) 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

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

11.9.2. Proof of the Vitali covering theorem. Our proof uses the Radó covering lemma to obtain a proof for Lemma 11.22 (i.e., the geometrical form of the Vitali theorem).

Proof. We use the notation, for any subpartition π,

$$V(\mathcal{L}, \pi) = \sum_{(I,x)\in\pi} \mathcal{L}(I)$$

and

$$\sigma(\pi) = \bigcup_{(I,x)\in\pi} I$$

to simplify the writing. Note that $\sigma(\pi)$ is a finite union of compact intervals, and so also a compact set.

Start with an open set $G_0 \supset E$ with $\mathcal{L}(G_0) < \infty$ and prune β by considering $\beta_0 = \beta(G_0)$. This β_0 is a fine cover of E and so, by Radó's lemma, we can choose a subpartition $\pi_1 \subset \beta_0$ for which

$$\mathcal{L}(E) \leq 3V(\mathcal{L}, \pi_1).$$

Let

$$G_1 = G_0 \setminus \sigma(\pi_1)$$

which is evidently an open set that contains $E_1 = E \setminus \sigma(\pi_1)$. 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 $E_0 = E$, $\beta_0 = \beta(G_0)$. For any $n = 1, 2, \dots$, unless E_{n-1} is empty or has zero measure we select, by Radó's lemma, a subpartition $\pi_n \subset \beta_{n-1}$ for which

$$\mathcal{L}(E_{n-1}) \leq 3V(\mathcal{L}, \pi_n).$$

Let

$$G_n = G_{n-1} \setminus \sigma(\pi_n)$$

which is evidently an open set that contains $E_n = E_{n-1} \setminus \sigma(\pi_1)$. Set $\beta_n = \beta(G_n)$.

None of the subpartitions π_n overlap and so, in particular,

$$\pi'_N = \bigcup_{n=1}^{N} \pi_n$$

is itself a subpartition contained in β. If the process stops then it is easy to verify that

$$\mathcal{L}\left(E \setminus \sigma(\pi'_N)\right) = 0.$$

If the process does not stop then for some large N we must have

$$\mathcal{L}\left(E \setminus \sigma(\pi'_N)\right) < \epsilon. \tag{11.2}$$

To see this note that all the subpartitions have been pruned to lie in the open set G_0. From this it follows that

$$\sum_{i=1}^{\infty} V(\mathcal{L}, \pi_n) \leq \mathcal{L}(G_0) < \infty.$$

Choose N so large that $V(\mathcal{L}, \pi_N) < \epsilon/3$ and it will follow that

$$\mathcal{L}(E_{N-1}) \leq 3V(\mathcal{L}, \pi_N) < \epsilon.$$

But this is exactly (11.2). ■

Exercise 11.26. Let E be a set of finite Lebesgue measure and β any fine cover of E. Then there exists a finite or infinite sequence

$$(I_i, x_i) \quad i = 1, 2, 3, \dots$$

from β so that I_i and I_j are disjoint for $i \neq j$ and

$$\mathcal{L}\left(E \setminus \bigcup_i I_i\right) = 0. \tag{11.3}$$

Hint: Extract this information from the proof.

11.10. Fundamental Limit Theorems

The two measures $\mathcal{L}^*$ and $\mathcal{L}_*$ play a key role in estimates of sets arising from limits and derivatives. In contrast, it is difficult and arduous to connect limits and derivatives to the measure $\mathcal{L}$ using only the original definitions. The Vitali covering theorem allows us to use either the full or fine variant in any computation.

11.10.1. Limits. Let h be any real-valued function defined on covering relations. We have already defined the two measures $V^*(h, E)$ and $V_*(h, E)$ for any set E of real numbers. Let us define as well some limiting operations. Since the full covers are filtering there is a natural limiting operation associated with them. Here we give a simple inf/sup version; later we will see how closely connected these limits are to full and fine covers.

Definition 11.27. Define the limits

$$\limsup_{(I,x)\Rightarrow x} h(I, x) = \inf_{\delta>0} \left(\sup\{h(I, x) : \ x \in I \text{ and } \mathcal{L}(I) < \delta \ \}\right)$$

and

$$\liminf_{(I,x)\Rightarrow x} h(I, x) = \sup_{\delta>0} \left(\inf\{h(I, x) : \ x \in I \text{ and } \mathcal{L}(I) < \delta \ \}\right).$$

As usual if the lim sup and lim inf are same then the common value (including ∞ and $-\infty$) would be written as

$$\lim_{(I,x)\Rightarrow x} h(I, x).$$

11.10.2. Limsup comparison. There are close connections between limit properties and measure properties. We shall be considering limits of expressions of the form
$$\left|\frac{h(I,x)}{\mathcal{L}(I)}\right|.$$

LEMMA 11.28. (Limsup Comparison) Suppose that, for every x in a set E,
$$s < \limsup_{(I,x)\Rightarrow x}\left|\frac{h(I,x)}{\mathcal{L}(I)}\right| < r$$
Then $s\mathcal{L}(E) \leq V^*(h,E) \leq r\mathcal{L}(E)$ and $V_*(h,E) \leq r\mathcal{L}(E)$.

Proof. Let
$$\beta_1 = \{(I,x) : s\mathcal{L}(I) < |h(I,x))\}$$
and
$$\beta_2 = \{(I,x) : |h(I,x))| < r\mathcal{L}(I)\}.$$
Note that β_1 is a fine cover of E and that β_2 is a full cover of E. Let β be any full cover of E and note that $\beta_1 \cap \beta$ is a fine cover of E and that $\beta_2 \cap \beta$ is a full cover of E. Thus
$$V^*(h,E) \leq V(h,\beta\cap\beta_2) \leq rV(\mathcal{L},\beta\cap\beta_2) \leq rV(\mathcal{L},\beta).$$
From this it follows that
$$V^*(h,E) \leq r\mathcal{L}^*(E).$$
Similarly
$$sV_*(\mathcal{L},E) \leq V(s\mathcal{L},\beta\cap\beta_1) \leq V(h,\beta\cap\beta_1) \leq V(h,\beta).$$
From this it follows that
$$s\mathcal{L}_*(E) \leq V^*(h,E).$$
■

11.10.3. Liminf comparison. A similar theorem with a similar proof uses the limit inferior.

LEMMA 11.29. (Liminf Comparison) Suppose that, for every x in a set E,
$$s < \liminf_{(I,x)\Rightarrow x}\left|\frac{h(I,x)}{k(I,x)}\right| < r$$
Then $s\mathcal{L}(E) \leq V_*(h,E) \leq r\mathcal{L}(E)$ and $s\mathcal{L}(E) \leq V^*(h,E)$.

Proof. Use the methods already seen for Lemma 11.28. ■

11.11. Fundamental theorem of the calculus

From the limsup and liminf comparison lemmas we can easily deduce one of the fundamental tools of the calculus, our main method for determining a relationship between integrals and measures on one hand, and derivatives on the other.

THEOREM 11.30. Let h be a real-valued function defined on covering relations and E a set of real numbers for which $V^*(h,E) = 0$. Then
$$\lim_{(I,x)\Rightarrow x}\frac{h(I,x)}{\mathcal{L}(I)} = 0$$
for almost every x in E.

THEOREM 11.31. Let h be a real-valued function defined on covering relations and E a set of real numbers for which

$$\lim_{(I,x)\Rightarrow x} \frac{h(I,x)}{\mathcal{L}(I)} = 0$$

for every x in E. Then $V^*(h,E) = 0$.

11.11.1. The fundamental theorem of the calculus. As standard exercises we now leave it the student to apply these theorems to establish the exact relation between the derivative and the integral. We have already proved partial versions of these before (e.g., under additional hypotheses in Section 10.12).

THEOREM 11.32 (Derivative of the integral). Let $f : [a,b] \to \mathbb{R}$ be an integrable function on a compact interval $[a,b]$ with F as its indefinite integral. Then

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

for almost every point x in $[a,b]$.

COROLLARY 11.33. Let $f : [a,b] \to \mathbb{R}$ be an integrable function on a compact interval $[a,b]$. Then f must be almost continuous.

11.12. Variational characterization of Lebesgue's measure

We know that if $V^*(h - \mathcal{L}, \mathbb{R}) = 0$ then, for every set $E \subset \mathbb{R}$,

$$\mathcal{L}(E) = V^*(h,E) = V_*(h,E).$$

The following theorem presents a converse to this.

THEOREM 11.34. Let h be a real-valued function of interval-point pairs for which, for every set $E \subset \mathbb{R}$,

$$\mathcal{L}(E) = V^*(h,E) = V_*(h,E).$$

Then

$$V^*(|h| - \mathcal{L}, \mathbb{R}) = 0, \tag{11.4}$$

$$\lim_{(I,x)\Rightarrow x} \frac{|h(I,x)|}{\mathcal{L}(I)} = 1 \tag{11.5}$$

for almost every real number x, and

$$\int_a^b |h| = b - a \tag{11.6}$$

for every compact interval $[a,b]$.

Proof. Let us prove assertion (11.5). Assume that it fails. Then there must exist a set E and rational numbers $0 < t < 1$ and $s > 0$ so that $\mathcal{L}(E) > s$ and

$$\limsup_{(I,x)\Rightarrow x} \frac{|h(I,x)|}{\mathcal{L}(I)} > 1 + t \tag{11.7}$$

for all x in E, or else

$$\liminf_{(I,x)\Rightarrow x} \frac{|h(I,x)|}{\mathcal{L}(I)} < 1 - t \tag{11.8}$$

for all x in E.

In the former case the limsup comparison lemma provides the inequality

$$V^*(h, E) \geq (1+t)\mathcal{L}(E)$$

which contradicts the identity that $V^*(h, E) = \mathcal{L}(E)$. On the other hand, if instead (11.8) holds, the liminf comparison lemma yields

$$\mathcal{L}(E) \geq (1-t)^{-1} V_*(h, E)$$

which again contradicts the identity that $V_*(h, E) = \mathcal{L}(E)$.

For the remainder of the proof we note merely that (11.4) follows easily from (11.5) and that (11.6) follows from (11.4). ■

11.13. The Density theorem

DEFINITION 11.35. Let S be a set of real numbers and let x_0 be any point in $\mathbb{R}$. Then x_0 is a *point of density* for S provided that

$$\lim_{(I,x_0)\Rightarrow x_0} \frac{\mathcal{L}(S \cap I)}{\mathcal{L}(I)} = 1.$$

If S is open then certainly every point in S is a point of density; some points outside of S may also be points of density. A null set has no points of density.

THEOREM 11.36 (Density theorem). Let S be an almost closed set. Then almost every point of S is a point of density.

Proof. We suppose that S is bounded, since the proof can be readily obtained once the bounded case is handled. Let n be a fixed integer and suppose that T_n is the set of points x in S at which

$$\liminf_{(I,x)\Rightarrow x} \frac{\mathcal{L}(S \cap I)}{\mathcal{L}(I)} < 1 - \frac{1}{n}.$$

If we write S^c for the set complementary to S then, since S is almost closed,

$$\mathcal{L}(I) = \mathcal{L}(S \cap I) + \mathcal{L}(S^c \cap I).$$

Thus we can also describe T_n as the set of points where

$$\limsup_{(I,x)\Rightarrow x} \frac{\mathcal{L}(S^c \cap I)}{\mathcal{L}(I)} > \frac{1}{n}.$$

We claim that T_n is a null set.

If not then $\mathcal{L}(T_n) = k > 0$ and we can choose an open set $G \supset S$ so that

$$\mathcal{L}(G) < \mathcal{L}(S) + \frac{k}{n}.$$

Write h for the interval function $h(I) = \mathcal{L}(S^c \cap I)$ and apply the limsup comparison lemma (Lemma 11.28) to the fact that, for all $x \in T_n$,

$$\limsup_{(I,x)\Rightarrow x} \frac{h(I)}{\mathcal{L}(I)} > \frac{1}{n}$$

to obtain

$$\mathcal{L}(T_n)/n \leq V^*(h, T_n).$$

Note that

$$V^*(h, T_n) \leq \mathcal{L}(S^c \cap G).$$

This follows from the fact that for any full cover β of T_n, $\beta(G)$ is also a full cover of T_n and so

$$V^*(h, T_n) \leq V(h, \beta(G)) \leq \mathcal{L}(S^c \cap G).$$

In particular

$$\mathcal{L}(S^c \cap G) \geq \frac{k}{n}.$$

Now, on the other hand,

$$G = (G \cap S) \cup (G \cap S^c)$$

and

$$G \cap S = S$$

so that

$$\mathcal{L}(G) = \mathcal{L}(S) + \mathcal{L}(G \cap S^c)$$

requiring that

$$\mathcal{L}(G \cap S^c) < k/n.$$

This contradiction establishes that T_n is a null set. Since this is true for all n the set of points at which

$$\liminf_{(I,x) \Rightarrow x} \frac{\mathcal{L}(S \cap I)}{\mathcal{L}(I)} < 1$$

must also be a null set and the theorem is proved. ■

11.14. Approximate continuity

A function f is continuous at a point x_0 if and only if the following situation holds: for every pair of rational numbers r and s for which

$$r < f(x_0) < s$$

the set

$$E_{rs} = \{x : r < f(x) < s\}$$

contains an open set that includes x_0. In particular x_0 would have to be a point of density of E_{rs}. Thus the following definition is much weaker than continuity.

Definition 11.37. Let $f : \mathbb{R} \to \mathbb{R}$ be an almost continuous function. Then f is *approximately continuous* at a point x_0 if, for every pair of rational numbers r and s for which

$$r < f(x_0) < s,$$

the point x_0 is a point of density of the set

$$E_{rs} = \{x : r < f(x) < s\}.$$

As an exercise the student should have little trouble using the density theorem to establish the following theorem.

Theorem 11.38. Let $f : \mathbb{R} \to \mathbb{R}$ be an almost continuous function. Then f is approximately continuous at almost every point.

11.15. s-dimensional measures

DEFINITION 11.39. Let $0 < s < 1$. Then the set functions $\mathcal{L}_s$ and $\mathcal{L}^s$, defined for all subsets E of $\mathbb{R}$ by

$$\mathcal{L}_s(E) = V_*(\mathcal{L}^s, E)$$

and

$$\mathcal{L}^s(E) = V^*(\mathcal{L}^s, E),$$

are called the *fine and full s-dimensional measures* on the real line.

These measures arise in many situations, in particular in the study of the s-dimensional Lipschitz numbers:

DEFINITION 11.40. Let $F : \mathbb{R} \to \mathbb{R}$, let $0 < s < 1$ and let x_0 be a point on the real line. Then we write

$$\overline{D_s}(F, x_0) = \limsup_{(I,x_0) \Rightarrow x_0} \frac{|\Delta F(I)|}{\mathcal{L}(I)^s}$$

and

$$\underline{D_s}(F, x_0) = \liminf_{(I,x_0) \Rightarrow x_0} \frac{|\Delta F(I)|}{\mathcal{L}(I)^s}$$

and refer to these as the *s-dimensional Lipschitz numbers* of F at x_0.

THEOREM 11.41 (The s-dimensional density theorem). Let $F : \mathbb{R} \to \mathbb{R}$ be a continuous function, let $0 < s < 1$ and let E be a set of real numbers. Then

$$\frac{V_*(\Delta F, E)}{\sup_{x \in E} \overline{D_s}(F, x)} \leq \mathcal{L}_s(E) \leq \frac{V^*(\Delta F, E)}{\inf_{x \in E} \overline{D_s}(F, x)}$$

and

$$\frac{V_*(\Delta F, E)}{\sup_{x \in E} \underline{D_s}(F, x)} \leq \mathcal{L}^s(E) \leq \frac{V^*(\Delta F, E)}{\inf_{x \in E} \underline{D_s}(F, x)}$$

The proof is left to the student. The methods we have developed in this chapter can be used. Note that, unlike the situation for Lebesgue measure where $\mathcal{L}* = \mathcal{L}^*$, the s-dimensional measures generally satisfy $\mathcal{L}_s < \mathcal{L}^s$. Also the larger measure is unusual in a certain sense: it assumes only the values 0 and ∞. The main use of such measures is in the notion of a rarefaction index (which the reader is encouraged to pursue elsewhere).

CHAPTER 12

Jordan variation of a real function

The concept of variation of a function was introduced by Camille Jordan in his 1882-1887 studies of the following two problems.

> A. What is the smallest linear space of functions on a compact interval $[a,b]$ if that space must include all monotonic functions?

and

> B. Let $f : [a,b] \to \mathbb{R}$ be a continuous function. What should we take for the length of the graph of f, i.e., what is the length of the set of points $\{(x, f(x)) : x \in [a,b]\}$ in $\mathbb{R}^2$?

Both problems lead to a detailed study of the sums

$$\sum_{(I,x)\in\pi} |\Delta f(I)|$$

taken over partitions of the interval $[a,b]$. From our point of view such a study belongs to the integration theory, the integration theory of interval functions. Thus we begin this chapter with the study of integrals of interval functions.

Later measure-theoretic methods were used to discuss the variation of f on arbitrary sets. In this chapter we consider the older methods, now largely subsumed by the later variational measures of Chapter 13.

12.1. Integration of interval functions

We require a method for integrating the interval function $|\Delta f|$. In fact the methods we have already studied suffice to define an integral for arbitrary real-valued functions of intervals (i.e., functions of the form $h(I)$ for I an arbitrary compact interval). In fact our methods will work for functions of the form $h(I,x)$ defined on any covering relation; we can pass to these when needed.

Definition 12.1. Suppose that $h(I)$ is defined as a real number for all compact subintervals of $[a,b]$. Then we define

$$\overline{\int_a^b} h = \inf_\beta \sup \left\{ \sum_{(I,x)\in\pi} h(I) : \ \pi \subset \beta \text{ and } \pi \text{ a partition of } [a,b] \right\}$$

where the infimum is taken over all Cousin covers β of $[a,b]$.

Similarly we define

$$\underline{\int_a^b} h.$$

As usual we say h is *integrable* if the upper and lower integrals agree and are finite. The common value is denoted as

$$\int_a^b h.$$

We are mainly interested in this chapter in applying this notation for

$$h(I) = |\Delta f(I)|$$

and similar expressions arising from a function $f : [a,b] \to \mathbb{R}$. The special case $h(I,x) = f(x)\mathcal{L}(I)$ corresponds to the usual calculus integral. The further special case

$$h(I,x) = \Delta F(I) - f(x)\mathcal{L}(I),$$

as we shall shortly see, offers a compact way of writing the Henstock criterion for integration.

EXERCISE 12.2. Show that $\int_a^b \Delta F = \Delta F([a,b]$.

Hint: Check that $\sum_{(I,x)\in\pi} \Delta F(I) = \Delta F([a,b]$ for any partition π of $[a,b]$.

EXERCISE 12.3. Let h be a nonnegative interval function. If $a < b < c$ show that (assuming these integrals are finite)

$$\overline{\int_a^b} h + \overline{\int_b^c} h \leq \overline{\int_a^c} h.$$

Give an example illustrating strict inequality, even in the case where h is integrable on both subintervals $[a,b]$ and $[b,c]$.

12.2. Henstock criterion

The Henstock criterion of Section 9.5 for ordinary integrals transfers with few changes to this setting and is fundamental to its study. Note that it has a particularly attractive expression now, not needing epsilons and Cousin covers to describe.

THEOREM 12.4 (Henstock criterion). Suppose that h is an interval function, integrable on a compact interval $[a,b]$. Then h is integrable on all compact subintervals $[c,d]$ of $[a,b]$, and there is an indefinite integral $H : [a,b] \to \mathbb{R}$ for which

$$\Delta H([c,d]) = \int_c^d h \qquad (a \leq c < d \leq b)$$

and

$$\int_a^b |\Delta H - h| = 0. \tag{12.1}$$

Conversely if (12.1) is true then h is integrable on $[a,b]$ with H as an indefinite integral.

Proof. The proof from Section 9.5 can be repeated with only minor changes. ∎

12.3. Differentiation of the integral

The Henstock criterion combines immediately with the fundamental theorem of the calculus (Theorems 11.31 and 11.32 from Section 11.11) to provide a version for integrals of interval functions.

THEOREM 12.5 (Derivative of the integral). Suppose that h is an interval function, integrable on a compact interval $[a, b]$ with an indefinite integral $H : [a, b] \to \mathbb{R}$. Then

$$\lim_{I \Rightarrow x} \frac{\Delta H(I) - h(I)}{\mathcal{L}(I)} = 0$$

at almost every point $x \in [a, b]$ and

$$H'(x) = \lim_{I \Rightarrow x} \frac{h(I)}{\mathcal{L}(I)}$$

at almost every point x at which H is differentiable.

THEOREM 12.6 (Integral of derivatives). Suppose that $H : [a, b] \to \mathbb{R}$, that h is an interval function, and that

$$\lim_{I \Rightarrow x} \frac{\Delta H(I) - h(I)}{\mathcal{L}(I)} = 0$$

at every point x in the compact interval $[a, b]$ excepting a set Z for which

$$V^*(h, Z) = V^*(\Delta H, Z) = 0.$$

Then h is integrable on $[a, b]$ and H is an indefinite integral.

12.4. Integrability of subadditive, continuous functions

An interval function h is *subadditive* on $[a, b]$ (we recall) if

$$h(J) \leq \sum_{(I,x) \in \pi} h(I)$$

for every compact subinterval J of $[a, b]$ and every partition π of J. As usual we say that h is *continuous* if $V^*(h, C) = 0$ for every countable set. By familiar methods we can show that a subadditive interval function h that is continuous is also continuous in another familiar sense: for every $\epsilon > 0$ there is a $\delta > 0$ so that $|h(I)| < \epsilon$ whenever I is a compact subinterval of $[a, b]$ for which $\mathcal{L}(I) < \delta$.

THEOREM 12.7. Let h be a subadditive, continuous, nonnegative interval function, defined on subintervals of a compact interval $[a, b]$. Then

$$\underline{\int_a^b} h = \overline{\int_a^b} h \tag{12.2}$$

and h is integrable on $[a, b]$ if and only if these expressions are finite.

Proof. Certainly if the upper or lower integral is infinite then h is not integrable. So we need only prove that the identity (12.2) holds. Take any number t for which

$$t < \overline{\int_a^b} h$$

and let $\epsilon > 0$. There must be at least one partition π_0 for which

$$t < \sum_{(I,x)\in\pi_0} h(I).$$

Write the partition as $\pi_0 = \{([a_{i-1}, a_i], x_i)\}$ where $a = a_0 < a_1 < \cdots < a_n = b$.

Choose $\delta > 0$ so that $|h(I)| < \epsilon/(3n)$ whenever I is a compact subinterval of $[a, b]$ for which $\mathcal{L}(I) < \delta$.Write

$$\beta_1 = \{(I, x) : \; x \in I \text{ and } I \subset [a_{i-1}, a_i] \text{ for some } i = 1, 2, \dots, n\}$$

and

$$\beta_2 = \{(I, x) : x = a_i \text{ for some } i \text{ and } \mathcal{L}(I) < \delta\}.$$

Now construct $\beta = \beta_1 \cup \beta_2$. Check that β is a Cousin cover of $[a, b]$.

The feature we require of β is this: if π is any partition of $[a, b]$ from β then, if there happens to be an interval-point pair $([c, d], a_i) \in \pi$ with $c < a_i < d$, we can substitute it with the two elements $([c, a_i], a_i)$ and $([a_i, d], a_i)$ from β without altering the sum $\sum_\pi h$ by more than ϵ/n. We do this if necessary for each i and obtain a new partition π' constructed from π for which

$$|\sum_{(I,x)\in\pi} h(I) - \sum_{(I,x)\in\pi'} h(I)| < \epsilon.$$

By the nature of the construction, the new partition π' must itself contain a partition of every interval appearing in π_0 (i.e., if $(I, x) \in \pi$ then $\pi'(I)$ is a partition of I). Thus the subadditivity of h gives us the computation

$$t < \sum_{(I,x)\in\pi_0} h(I) \leq \sum_{(I,x)\in\pi'} h(I)$$

Apply this to any original partition π from β and we have that

$$t - 2\epsilon < \sum_{(I,x)\in\pi} h(I)$$

must hold for all partitions π of the interval $[a, b]$ from β. Thus

$$t - 2\epsilon < \underline{\int_a^b} h$$

and the identity (12.2) follows. ∎

12.5. Jordan variation

We use the integral just defined as our definition of the Jordan variation. We restrict attention to continuous functions so that we may take advantage of the fact that continuous, subadditive interval functions are integrable.

Definition 12.8. Let $f : [a, b] \to \mathbb{R}$ be a continuous function. The *Jordan variation* of f over $[a, b]$ is defined to be the expression

$$\int_a^b |\Delta f|$$

and the function

$$T_f(x) = \int_a^x |\Delta f| \quad (a \leq x \leq b)$$

is called *the total variation function* for f. We say that f has *bounded [Jordan] variation* on $[a, b]$ if this is finite.

If $f : \mathbb{R} \to \mathbb{R}$ has bounded variation throughout every compact interval it is said to be *locally of bounded Jordan variation.*

EXERCISE 12.9. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that

$$\int_a^b |\Delta f| = \int_a^b \omega f.$$

EXERCISE 12.10. If $f : [a, b] \to \mathbb{R}$ is not continuous we use the expression

$$\overline{\int_a^b} |\Delta f|$$

for the Jordan variation. Determine the elementary properties of this variation.

12.6. Jordan decomposition

It follows from simple properties of the integral that the class of all continuous functions having bounded variation throughout a compact interval $[a, b]$ includes all continuous monotonic functions and is closed under linear combinations. An important converse is the observation that every continuous function of bounded Jordan variation on $[a, b]$ can be written as a linear combination of monotonic functions, so such functions have a simple basic structure.

THEOREM 12.11 (Jordan decomposition theorem). Let $f : [a, b] \to \mathbb{R}$ be a continuous function that has bounded Jordan variation throughout $[a, b]$. Then f and T_f can be expressed as the difference and sum of two continuous nondecreasing functions P and N,

$$f(x) - f(a) = P(x) - N(x) \quad \text{and} \quad T_f(x) = P(x) + N(x) \tag{12.3}$$

where

$$P(x) = \int_a^x [\Delta f]^+ \quad \text{and} \quad N(x) = \int_a^x [\Delta f]^-.$$

Moreover $T_f'(x) = |f'(x)|$, $P'(x) = \max\{f'(x), 0\}$, $N'(x) = \max\{-f'(x), 0\}$, and $P'(x)N'(x) = 0$ at almost every point x in $[a, b]$.

Proof. Theorem 12.7 assures us that each of the functions $|\Delta f|$, $[\Delta f]^+$ and $[\Delta f]^-$ is integrable. Thus the identities in (12.3) follow immediately from the identities

$$\Delta f = [\Delta f]^+ - [\Delta f]^- \quad \text{and} \quad |\Delta f| = [\Delta f]^+ + [\Delta f]^-$$

merely by integrating.

For the differentiation statements, first observe that the derivatives $P'(x)$ and $N'(x)$ must exist for almost all x in $[a, b]$ by the Lebesgue differentiation theorem. The Henstock criterion provides

$$\int_a^b |\Delta P - [\Delta f]^+| = 0.$$

From Lemma 12.5 we obtain the differentiation statement

$$P'(x) = \lim_{(I,x)\Rightarrow x} \frac{[\Delta f(I)]^+}{\mathcal{L}(I)}$$

almost everywhere. This must be the same as $f'(x)$ if $f'(x) > 0$ and must be 0 if $f'(x) < 0$. The other differentiation statements are obtained in a similar manner. ■

EXERCISE 12.12. Obtain a version of the Jordan decomposition theorem for functions of bounded variation that are not assumed continuous.

Hint: Define

$$P(x) = \frac{1}{2}\left\{\overline{\int_a^x}|\Delta f| + [f(x) - f(a)]\right\}$$

and

$$N(x) = \frac{1}{2}\left\{\overline{\int_a^x}|\Delta f| - [f(x) - f(a)]\right\}$$

In order to check that P is a monotonic nondecreasing function, first use Exercise 12.3 to obtain for any interval $[c,d] \subset [a,b]$ the inequality

$$TV_f(d) - TV_f(c) = \overline{\int_a^d}|\Delta f| - \overline{\int_a^c}|\Delta f| \geq \overline{\int_c^d}|\Delta f| \geq |\Delta f([c,d])| = |f(d) - f(c)|.$$

Now $P(d) - P(c)$, the increment of P on this interval, can be computed as

$$\frac{1}{2}\left[TV_f(d) - TV_f(c) + (f(d) - f(c))\right]$$

which is nonnegative. Almost identical computations show that N too is a monotonic nondecreasing function.

12.7. Absolute continuity in sense of Vitali

Recall from Section 4.9 that a function f is absolutely continuous in the sense of Vitali on a compact interval $[a,b]$ provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever π is any subpartition from $[a,b]$ for which

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) < \delta \quad \text{implies that} \quad \sum_{(I,x)\in\pi} |\Delta f(I)| < \epsilon.$$

THEOREM 12.13. A function $f : [a,b] \to \mathbb{R}$ is absolutely continuous in the sense of Vitali on $[a,b]$ if and only if f is absolutely continuous and has bounded variation.

Proof. Suppose that f is absolutely continuous in the sense of Vitali on $[a,b]$. Then we already know that f is absolutely continuous. To see that f has bounded variation let us suppose that

$$\sum_{(I,x)\in\pi} |\Delta f(I)| < 1$$

whenever a partition π has total length less than δ. Consider a Cousin cover β chosen in such a way that any pair (I,x) belonging to β must have length smaller than $\delta/2$. Take any partition π of $[a,b]$ contained in β and any integer $m > (b-a)/\delta$. We may subdivide π into at most m subcollections $\{\pi_i\}$ each of total length smaller than δ. Consequently

$$\sum_{(I,x)\in\pi} |\Delta f(I)| = \sum_i \sum_{(I,x)\in\pi_i} |\Delta f(I)| < m.$$

From this it follows that

$$\int_a^b |\Delta f| < m < \infty$$

and so f has bounded variation on $[a,b]$.

In the converse direction suppose that f is absolutely continuous and has bounded variation. Then there is a null set N so that $f'(x)$ exists at each x in the set $E = [a,b] \setminus N$ and $V^*(\Delta F, N) = 0$. Note that this means that for any subinterval $[c,d] \subset [a,b]$,

$$|\Delta F([c,d]| \leq V^*(\Delta F, [c,d])$$

$$= V^*(\Delta F, E \cap [c,d]) = V^*(f'\mathcal{L}, E \cap [c,d]) = \int_{E\cap[c,d]} |f'(x)|\, dx.$$

Since f' is almost continuous and

$$\int_E |f'(x)|\, dx = V^*(\Delta, [a,b]) < \infty$$

we can apply the absolute continuity theorem (Theorem 8.14) for the Lebesgue "integral." This translates neatly into the statement that F is absolutely continuous in the sense of Vitali. ■

Corollary 12.14. A function $f : [a,b] \to \mathbb{R}$ of bounded variation is absolutely continuous in the sense of Vitali on $[a,b]$ if and only if f is the indefinite integral of its derivative f'.

Every function that is absolutely continuous in the sense of Vitali has the three properties asserted in the following theorem. The converse will be established in Chapter 13. A version for absolute continuity in the more general sense will also be given.

Theorem 12.15. If a function $f : [a,b] \to \mathbb{R}$ is absolutely continuous in the sense of Vitali on $[a,b]$ then each of the following are true:

(a) f is continuous.
(b) f has bounded variation on $[a,b]$.
(c) $\mathcal{L}(f(N)) = 0$ for every subset N of $[a,b]$ that has $\mathcal{L}(N) = 0$.

Proof. Suppose that f is absolutely continuous in the sense of Vitali on $[a,b]$. Then we know from the previous theorem that f is absolutely continuous (hence also continuous) and that f has bounded variation.

Now we check the third condition (known as Lusin's condition). Let $\epsilon > 0$. Using the definition we can choose a $\delta > 0$ so that whenever π is any subpartition from $[a,b]$ for which

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) < \delta \quad \text{implies that} \quad \sum_{(I,x)\in\pi} |\omega f(I)| < \epsilon.$$

This is only a minor modification of the definition itself.

Let Z be a null set and choose an open set G containing Z but with $\mathcal{L}(G) < \delta$. Then if $\{(a_k, b_k)\}$ are the component intervals of G we must have

$$\mathcal{L}(f(Z)) \leq \mathcal{L}(f(G)) \leq \sum_k \mathcal{L}(f([a_k,b_k]) \leq \sum_k \omega f([a_k,b_k]) \leq \epsilon.$$

From this we deduce that the third condition must hold. ■

12.8. Mutually singular functions

DEFINITION 12.16. Let $f,\ g : [a,b] \to \mathbb{R}$ be continuous functions of bounded variation. Then f and g are said to be *mutually singular* provided that

$$\int_a^b \sqrt{|\Delta f \cdot \Delta g|} = 0.$$

LEMMA 12.17. Let $f,\ g : [a,b] \to \mathbb{R}$ be continuous functions of bounded variation. If f and g are mutually singular, then $f'(x)g'(x) = 0$ almost everywhere in $[a,b]$.

Proof. Since

$$\int_a^b \sqrt{|\Delta f \cdot \Delta g|} = 0$$

we can use Lemma 12.5 to deduce that

$$\lim_{(I,x)\Rightarrow x} \sqrt{\frac{\Delta f(I)}{\mathcal{L}(I)}}\sqrt{\frac{\Delta f(I)}{\mathcal{L}(I)}} = 0$$

at almost every x in $[a,b]$. Since $f'(x)$ and $g'(x)$ exist almost everywhere we conclude that $f'(x)g'(x) = 0$ almost everywhere in $[a,b]$. ■

THEOREM 12.18. Let $f,\ g : [a,b] \to \mathbb{R}$ be continuous functions of bounded variation. Then f and g are mutually singular on $[a,b]$ if and only for every $\epsilon > 0$ there is a Cousin cover β of $[a,b]$ with the property that every partition π of $[a,b]$ contained in β can be split into two disjoint subpartitions $\pi = \pi' \cup \pi''$ so that

$$\sum_{(I,x)\in\pi'} |\Delta f(I)| < \epsilon$$

and

$$\sum_{(I,x)\in\pi''} |\Delta g(I)| < \epsilon.$$

Proof. Suppose that

$$\int_a^b \sqrt{|\Delta f \cdot \Delta g|} = 0.$$

Let $\epsilon > 0$ and select a Cousin cover β so that

$$\sum_{(I,x)\in\pi} \sqrt{|\Delta f(I) \cdot \Delta g(I)|} < \epsilon$$

for all partitions π of $[a,b]$ contained in β. Split such a π as follows:

$$\pi' = \{(I,x) : |\Delta f(I)| \leq |\Delta g(I)|\}$$

and

$$\pi'' = \{(I,x) : |\Delta f(I)| > |\Delta g(I)|\}.$$

Verify that $\pi = \pi' \cup \pi''$ and that

$$\sum_{(I,x)\in\pi'} |\Delta f(I)| \leq \sum_{(I,x)\in\pi'} \sqrt{|\Delta f(I) \cdot \Delta g(I)|} < \epsilon$$

and that

$$\sum_{(I,x)\in\pi''} |\Delta g(I)| \leq \sum_{(I,x)\in\pi''} \sqrt{|\Delta f(I) \cdot \Delta g(I)|} < \epsilon.$$

This proves one direction in the theorem.

For the converse select a number $M > 0$ and a Cousin cover β_1 so that

$$\sum_{(I,x)\in\pi} [|\Delta f(I)| + |\Delta g(I)|] < M$$

for all partitions π of $[a, b]$ from β_1. This is possible merely because the functions f and g have bounded variation. Select a Cousin cover β_2 with the property presented in the statement of the theorem (for ϵ). Let $\beta = \beta_1 \cap \beta_2$. This is a Cousin cover of $[a, b]$. Consider any partition π of $[a, b]$ contained in β. There must be, by hypothesis, a split $\pi = \pi' \cup \pi''$ so that

$$\sum_{(I,x)\in\pi'} |\Delta f(I)| < \epsilon$$

and

$$\sum_{(I,x)\in\pi''} |\Delta g(I)| < \epsilon.$$

We now compute

$$\sum_{(I,x)\in\pi} \sqrt{|\Delta f(I) \cdot \Delta g(I)|} = \sum_{(I,x)\in\pi'} \sqrt{|\Delta f(I) \cdot \Delta g(I)|} + \sum_{(I,x)\in\pi''} \sqrt{|\Delta f(I) \cdot \Delta g(I)|}$$

$$\leq \sqrt{\sum_{(I,x)\in\pi'} |\Delta f(I)|}\sqrt{\sum_{(I,x)\in\pi'} |\Delta g(I)|} + \sqrt{\sum_{(I,x)\in\pi''} |\Delta f(I)|}\sqrt{\sum_{(I,x)\in\pi''} |\Delta g(I)|}$$

$$\leq 2\sqrt{M\epsilon}.$$

Here we have used the Cauchy-Schwartz inequality. Since ϵ is an arbitrary positive number it follows that

$$\int_a^b \sqrt{|\Delta f \cdot \Delta g|} = 0.$$

Consequently f and g must be mutually singular. ∎

12.9. Properties of the Jordan decomposition

As an application of the methods for mutually singular functions let us show that the two nondecreasing functions supplied by the Jordan decomposition must be mutually singular.

THEOREM 12.19. Let $f : [a, b] \to \mathbb{R}$ be a continuous function that has bounded Jordan variation throughout $[a, b]$. Then the functions

$$P(x) = \int_a^x [\Delta f]^+ \quad \text{and} \quad N(x) = \int_a^x [\Delta f]^-$$

are mutually singular on $[a, b]$

Proof. Using the Henstock criterion we may select, for any $\epsilon > 0$ a Cousin cover β of $[a, b]$ in such a way that

$$\sum_{(I,x)\in\pi} |\Delta P(I) - [\Delta f(I)]^+| < \epsilon$$

and

$$\sum_{(I,x)\in\pi} |\Delta N(I) - [\Delta f(I)]^-| < \epsilon$$

for any subpartition π chosen from β.

Now to verify that P and N are mutually singular we show that the condition in Theorem 12.18 must hold for this β and ϵ. Let π be any partition of $[a,b]$ chosen from β and write

$$\pi_1 = \{(I,x) \in \pi : \Delta f(I) \geq 0\}$$

and

$$\pi_2 = \{(I,x) \in \pi : \Delta f(I) < 0\}.$$

Certainly $\pi = \pi_1 \cup \pi_2$ and

$$\sum_{(I,x)\in\pi_2} |\Delta P(I)| = \sum_{(I,x)\in\pi} |\Delta P(I) - [\Delta f(I)]^+| < \epsilon$$

and

$$\sum_{(I,x)\in\pi_1} |\Delta N(I)| = \sum_{(I,x)\in\pi} |\Delta N(I) - [\Delta f(I)]^-| < \epsilon.$$

It follows that P and N are mutually singular. ■

EXERCISE 12.20. Let $f : [a,b] \to \mathbb{R}$ be absolutely continuous in the sense of Vitali on $[a,b]$. Show that the parts of the Jordan decomposition of f are necessarily also absolutely continuous in the sense of Vitali.

12.10. Singular functions

A function f that is mutually singular with the identity function $g(x) = x$ is said to be *singular*. It is not immediately clear that there can exist continuous singular functions other than the constant functions. The student should seek out a popular account of the devil's staircase in order to find an appropriate example of such a function.

This definition can be easily reconciled with the definition of the term "singular" given in an earlier chapter. The reader is invited to check this.

DEFINITION 12.21. Let $f : [a,b] \to \mathbb{R}$ be a continuous function of bounded variation. Then f is *singular* provided that

$$\int_a^b \sqrt{|\Delta f| \cdot \mathcal{L}} = 0.$$

THEOREM 12.22. Let $f : [a,b] \to \mathbb{R}$ be a continuous function of bounded variation. Then f is singular if and only if $f'(x) = 0$ almost everywhere in $[a,b]$.

Proof. Apply Lemma 12.17 with g chosen so that $g(x) = x$ for all x in $[a,b]$. This proves that whenever f is a singular function its derivative $f'(x)$ must be zero almost everywhere.

Conversely suppose that $f'(x) = 0$ almost everywhere. Let $\epsilon > 0$ and choose an open set G with $\mathcal{L}(G) < \epsilon$ so that $f'(x) = 0$ for all $x \in [a,b] \setminus G$. Define

$$\beta_1 = \{(I,x) : x \in I \subset [a,b] \text{ and } |\Delta f(I)| < \epsilon\mathcal{L}(I)/(b-a)\}$$

and

$$\beta_2 = \{(I,x) : x \in I \subset G \cap [a,b]\}.$$

Then $\beta = \beta_1 \cap \beta_2$ is a Cousin cover of $[a,b]$. Note that if π is a subpartition contained in β_1 then

$$\sum_{(I,x)\in\pi} |\Delta f(I)| \leq \sum_{(I,x)\in\pi} \epsilon\mathcal{L}(I)/(b-a) < \epsilon.$$

Note that if π is a subpartition contained in β_2 then

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) \le \mathcal{L}(G) < \epsilon.$$

Thus any partition of $[a,b]$ chosen from β can be split into two subpartitions with these inequalities. This verifies the conditions asserted in Theorem 12.18 for f and the function $g(x) = x$. Consequently f and g are mutually singular, or equivalently, f is singular.

■

THEOREM 12.23 (Lebesgue decomposition). Let f be a continuous function of bounded variation on a compact interval. Then $f = g + h$ where g is a continuous singular function and h is absolutely continuous in the sense of Vitali. Moreover g and h are mutually singular.

Proof. Define

$$h(t) = \int_a^t f'(x)\,dx$$

for all t in $[a,b]$. Then h is absolutely continuous in the sense of Vitali because of Theorem 12.14. The function $g = f - h$ is easily checked to be singular because $g'(x) = f'(x) - h'(x) = f'(x) - f'(x) = 0$ almost everywhere. Finally g and h are mutually singular merely because $g'(x)h'(x) = 0$ almost everywhere. ■

COROLLARY 12.24. Let $f : [a,b] \to \mathbb{R}$ be a continuous function of bounded variation that is both singular and absolutely continuous in the sense of Vitali on $[a,b]$. Then f is constant.

Proof. If f is singular then we know that f' is a null function. If we know that f is absolutely continuous then we know that f is the indefinite integral of its derivative f'. Consequently f must be constant. ■

12.11. Length of curves

A curve is a pair of continuous functions $f,\ g : [a,b] \to \mathbb{R}$. We consider that the curve is the pair, rather than that the curve is the geometric set of points

$$\{(f(x), g(x)) : x \in [a,b]\}$$

that is the object we might likely think about when contemplating a curve.

DEFINITION 12.25. Suppose that $f,\ g : [a,b] \to \mathbb{R}$ is a pair of continuous functions. By the *length* of the curve given by f and g we shall mean

$$\int_a^b \sqrt{[\Delta f]^2 + [\Delta g]^2}.$$

Note that as f and g are continuous, then so too is the interval function

$$h(I) = \sqrt{[\Delta f(I)]^2 + [\Delta g(I)]^2}.$$

A simple application of the Pythagorean theorem will verify that the function h here is then a continuous subadditive interval function. Thus the integral in the definition must necessarily exist, although it might have an infinite value.

A curve given by a pair of continuous functions $f,\ g : [a,b] \to \mathbb{R}$ has finite length if and only if both functions f and g have bounded variation. This is easy to verify because of the elementary inequalities in the lemma.

LEMMA 12.26. For any pair of continuous functions $f,\ g : [a,b] \to \mathbb{R}$ the following inequalities hold:

$$\max\left\{\int_a^b |\Delta f|, \int_a^b |\Delta g|\right\} \le \int_a^b \sqrt{[\Delta f]^2 + [\Delta g]^2}$$

and

$$\int_a^b \sqrt{[\Delta f]^2 + [\Delta g]^2} \le \int_a^b |\Delta f| + \int_a^b |\Delta g|.$$

12.11.1. Formula for the length of curves. In the elementary (computational) calculus one usually assumes that a curve is given by a pair of continuously differentiable functions (i.e., a pair f, g of continuous functions for which f' and g' are also continuous). In that case the familiar formula for length used in elementary applications is

$$\int_a^b \sqrt{[f'(x)]^2 + [g'(x)]^2}\, dx.$$

We study this now.

LEMMA 12.27. For any pair of continuous functions $f, g : [a,b] \to \mathbb{R}$ of bounded variation on $[a,b]$ define the following function

$$L(t) = \int_a^t \sqrt{[\Delta f]^2 + [\Delta g]^2} \qquad (a < t \le b).$$

Then

$$L'(t) = \sqrt{[f'(t)]^2 + [g'(t)]^2}$$

almost everywhere in $[a,b]$.

Proof. From methods, by now quite familiar (Lemma 12.5), we know that

$$\lim_{(I,x)\Rightarrow x} \frac{\Delta L(I) - \sqrt{[\Delta f(I)]^2 + [\Delta g(I)]^2}}{\mathcal{L}(I)} = 0$$

for almost every x in $[a,b]$. The statement of the theorem follows from this since we know that $L'(x)$, $f'(x)$, and $g'(x)$ exist for almost every x. ∎

LEMMA 12.28. The function L in the lemma is absolutely continuous in the sense of Vitali if and only if both f and g are absolutely continuous in that sense.

Proof. This follows easily from the inequalities of Lemma 12.26. ∎

The length of the curve is now available as a familiar formula precisely in the case where the two functions defining the curve are absolutely continuous.

LEMMA 12.29. For any pair of continuous functions $f, g : [a,b] \to \mathbb{R}$ of bounded variation on $[a,b]$,

$$\int_a^b \sqrt{[\Delta f]^2 + [\Delta g]^2} \ge \int_a^b \sqrt{[f'(x)]^2 + [g'(x)]^2}\, dx.$$

The two expressions are equal if and only if both f and g are absolutely continuous on $[a,b]$.

Proof. Using the function L introduced above we see that this assertion is easily deduced from the fact that

$$L(t) \geq \int_a^t L'(x)\, dx$$

with equality precisely when L is absolutely continuous. ■

EXERCISE 12.30. For any continuous function $f : [a, b] \to \mathbb{R}$ define the length of the graph of f to mean

$$\int_a^b \sqrt{\mathcal{L}^2 + [\Delta f]^2}.$$

Show that the graph has finite length if and only if f has bounded variation. Discuss the availability of the the familiar formula for length used in elementary applications:

$$\int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

12.12. The Indicatrix

Let $f : [a, b] \to \mathbb{R}$ and let $y \in \mathbb{R}$. We can count the number of times that $f(x) = y$ by assigning a number $N_f(y)$ described as the number of elements in the set

$$\{x \in [a, b] : f(x) = y\}.$$

The function N_f is called *the indicatix* for f. We must include the possibility that $N_f(y) = \infty$ if this value y is assumed infinitly often. None of our integration and measure methods permit functions to assume infinite values, but we can work around this. The simplest manoeever is just to assume that the indicatrix is finite almost everywhere. Then the infinite values can be ignored by taking any finite-valued function which is almost everywhere the same as the infinite-valued function.

THEOREM 12.31 (Indicatrix theorem). Let $f : [a, b] \to \mathbb{R}$ be a continuous function and suppose that the indicatrix N_f is finite almost everywhere. Then

$$\int_{\mathbb{R}} N_f(y)\, dy = \int_a^b |\Delta f|.$$

Proof. If f is constant then the identity must hold trivially. If f is not constant then the image of the interval $[a, b]$ is a compact interval $[A, B]$. Take any compact subinterval I of $[a, b]$ and let $M_I(y) = 1$ if y belongs to $f(I)$ and $M_I(y) = 0$ if y does not belong to $f(I)$. Observe that

$$\int_A^B M_I(y)\, dy = \omega f(I).$$

This identity is trivial but is the key to the theorem.

Select a sequence of partitions $\{\pi_j\}$ of the interval $[a, b]$ so that the following conditions are met:

(a)
$$\lim_{j\to\infty} \sum_{(I,x)\in\pi_j} \omega f(I) = \int_a^b \omega f.$$

(b) For each $j = 2, 3, \ldots$ and each $(I, x) \in \pi_{j-1}$, the collection π_j contains a partition of I.

(c) For each $j = 1, 2, 3, \dots$ and each $(I, x) \in \pi_j$, the length of the interval I is smaller than $1/j$.

Define

$$g_j(y) = \sum_{(I,x)\in\pi_j} M_I(y)$$

and observe that

$$\int_A^B g_j(y)\,dy = \sum_{(I,x)\in\pi_j} \omega f(I) \nearrow \int_a^b \omega f.$$

We see that $\{g_j\}$ is a nondecreasing sequence of functions, each of which is integrable on the interval $[A, B]$.

Let $y \in [A, B]$ and suppose that $N_f(y) = k$ for some integer $k = 1, 2, 3, \dots,$. That means that there are exactly k points in $[a, b]$ at which f assumes the value k. Choose j large enough so that no pair of these points is closer together than $1/j$. Then for each member $(I, x) \in \pi_j$ the interval I can contain at most one of these points. Indeed there are exactly k members $(I, x) \in \pi_j$ with the property that $M_I(y) = 1$ and $M_I(y) = 0$ for the remaining members. Consequently

$$N_f(y) = g_j(y)$$

for large enough j at every point y at which N_f is finite.

From the monotone convergence theorem, since $g_j \to N_f$ almost everywhere in $[A, B]$,

$$\int_A^B N_f(y)\,dy = \lim_{j\to\infty} \int_A^B g_j(y)\,dy = \lim_{j\to\infty} \sum_{(I,x)\in\pi_j} \omega f(I) = \int_a^b \omega f.$$

To complete the theorem now it is enough to check that the integral can be expressed as a measure, i.e., that

$$\int_A^B N_f(y)\,dy = \int_{\mathbb{R}} N_f(y)\,dy$$

and to recall that

$$\int_a^b \omega f = \int_a^b |\Delta f|.$$

■

EXERCISE 12.32. Let $f : [a, b] \to \mathbb{R}$ be a continuous function and suppose that the indicatrix N_f is infinite on a set of positive measure. Show that

$$\int_a^b |\Delta f| = \infty.$$

CHAPTER 13

Variational Measures

The Jordan variation is restricted to the study of functions of bounded variation on a compact interval $[a, b]$. For a large part of the calculus program this is a sufficiently useful tool. But there are differentiable functions which do not have bounded variation and there are integrable functions whose indefinite integrals are not of bounded variation.

Jordan's theory was extended in the early 20th century to handle functions of finite variation on arbitrary compact sets by A. Denjoy, N. Lusin, and S. Saks. This theory was clarified later by the introduction, by R. Henstock, of measures carrying the variational information of a function. This theory includes the Jordan version and the Denjoy-Lusin-Saks versions and is the appropriate technical tool for the full range of problems arising in the calculus program.

13.1. Variational measures

Let $f : \mathbb{R} \to \mathbb{R}$. Recall that we use the following notation for the variation of f over a partition π or an arbitrary covering relation β:

$$V(\Delta f, \pi) = \sum_{(I,x)\in\pi} |\Delta f(I)|$$

and

$$V(\Delta f, \beta) = \sup_{\pi\subset\beta} V(\Delta f, \pi),$$

where in the latter the supremum is taken over all subpartitions π contained in β.

DEFINITION 13.1 (Full and Fine Variations). Let $f : \mathbb{R} \to \mathbb{R}$ and let E be any set of real numbers. Then we define the full and fine variational measures associated with f by the expressions:

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

and

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

The two measures $\mathcal{V}_f{}^*$ and $\mathcal{V}_{f*}$ together express the variation of the function f. These set functions share the same properties as the measure $\mathcal{L}$. Specifically they are countably subadditive for sequences of sets and they are countably additive for disjoint sequences of closed sets. See Section 11.5 for details.

EXERCISE 13.2. Let $f : \mathbb{R} \to \mathbb{R}$. Establish the following relation between the Jordan variation and the variational measures:

$$\mathcal{V}_f{}^*((a, b)) \leq \overline{\int_a^b} |\Delta f| \leq \mathcal{V}_f{}^*([a, b]) = \mathcal{V}_f{}^*((a, b)) + \mathcal{V}_f{}^*(\{a\}) + \mathcal{V}_f{}^*(\{b\}).$$

In particular show that $\mathcal{V}_f{}^*((a,b)) = \mathcal{V}_f{}^*([a,b])$ is exactly the Jordan variation if f is continuous at a and b.

EXERCISE 13.3. Let $E \subset (a,b)$ be a compact set and let $\{(a_i,b_i)\}$ be the component intervals of $(a,b)\setminus E$. Suppose that f is a continuous function satisfying $f(x) = 0$ for all $x \in E$ and that

$$\sum_i \omega f([a_i,b_i]) < \infty.$$

Show that $\mathcal{V}_f{}^*(E) = 0$.

13.2. Variational estimates

In special cases it is easy to estimate the full and fine variations. Note that as a result of this computation we see that continuous, increasing functions possess the Vitali property.

13.2.1. Variation of continuous, increasing functions.

THEOREM 13.4. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and strictly increasing. Then, for any set E,

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

Proof. If β is a full [fine] cover of E then check that

$$\beta' = \{(f(I), f(x)) : (I,x) \in \beta\}$$

is a full [fine] cover of $f(E)$. Note too that $\Delta f(I) = \mathcal{L}(f(I))$ for such a function. From this we deduce that

$$\mathcal{L}^*(f(E)) = \mathcal{V}_f{}^*(E)$$

and

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

■

13.2.2. Variation and image measure. In general the full variation is larger than the image measure.

THEOREM 13.5. For an arbitrary function $f : \mathbb{R} \to \mathbb{R}$ and any real set E,

$$\mathcal{L}(f(E)) \leq \mathcal{V}_f{}^*(E).$$

Proof. Let $\mathcal{V}_f{}^*(E) < t$ and select a full cover β of E so that $V(\Delta f, \beta) < t$. We apply the decomposition lemma, Lemma 11.11, for β. There is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty}$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E so that if x is any point in E_n and I is any subinterval of I_k that contains x then (I,x) belongs to $\beta([E_n \cap I_{kn}])$.

Thus let us estimate the $\mathcal{L}$-measure of the set $f(E_n \cap I_{kn})$. Our estimate need only be crude: if $f(x_1)$, $f(x_2)$ with $x_1 < x_2$ are any two points in this set then certainly $([x_1,x_2],x_1) \in \beta(I_k)$. Thus

$$|f(x_1) - f(x_2)| = |\Delta f([x_1,x_2])| \leq V(\Delta f, \beta(I_{kn}))$$

so it follows that

$$\mathcal{L}(f(E_n \cap I_{kn}) \leq V(\Delta f, \beta(I_{kn}).$$

Hence, using Exercise 11.14 and properties of Lebesgue measure (Theorem 7.19), we have that

$$\mathcal{L}(f(E_n)) \leq \sum_k \mathcal{L}(f(E_n \cap I_{kn}) \leq \sum_k V(\Delta f, \beta(I_{kn}) \leq V(\Delta f, \beta) < t.$$

Note that the sequence $\{E_n\}$ is expanding and that its union is the whole set E; it follows that $\{f(E_n)\}$ is expanding and that its union is the whole set $f(E)$. Accordingly then, by the usual properties of Lebesgue measure (Theorem 7.19),

$$\lim_{n\to\infty} \mathcal{L}(f(E_n)) = \mathcal{L}(f(E)).$$

It follows that

$$\mathcal{L}(f(E)) \leq t.$$

Since t was merely chosen so that $\mathcal{V}_f{}^*(E) < t$ it follows that $\mathcal{L}(f(E)) \leq \mathcal{V}_f{}^*(E)$ as required. ■

13.3. Lipschitz numbers

We have defined the notion of a Lipschitz number as a local estimate of the growth of a function. We refine this a bit by introducing a lower estimate. In Section 13.4 we show how these numbers relate to the variations.

13.3.1. Upper and lower Lipschitz numbers.

DEFINITION 13.6. Let $f : \mathbb{R} \to \mathbb{R}$. Then

$$\overline{lip}_f(x) = \limsup_{(I,x)\Rightarrow x} \left|\frac{\Delta f(I)}{\mathcal{L}(I)}\right|$$

$$\underline{lip}_f(x) = \liminf_{(I,x)\Rightarrow x} \left|\frac{\Delta f(I)}{\mathcal{L}(I)}\right|$$

are called the *upper and lower Lipschitz number* of f at a point x.

LEMMA 13.7. Let $f : \mathbb{R} \to \mathbb{R}$. For any real number r the sets

$$\{x : \overline{lip}_f(x) < r\} \text{ and } \{x : \underline{lip}_f(x) < r\}$$

are almost closed.

Proof. This is nearly identical to Lemma 7.14. ■

13.3.2. Lipschitz numbers and derivatives. We are interested often in deducing the existence of the derivative from the Lipschitz numbers. The exercises reveal a close connection between the concepts.

EXERCISE 13.8. Let $f : \mathbb{R} \to \mathbb{R}$. Verify that

$$\overline{lip}_f(x) = \max\{|\overline{D}f(x)|, |\underline{D}f(x)|\}$$

and also

$$\overline{lip}_f(x) = \max\{|\overline{D}^+ f(x)|,\ |\underline{D}^+ f(x)|,\ |\overline{D}^- f(x)|,\ |\underline{D}^- f(x)|\}.$$

EXERCISE 13.9. Let $f : \mathbb{R} \to \mathbb{R}$. Suppose that f has a derivative at x (finite or infinite). Show that $\overline{lip}_f(x) = \underline{lip}_f(x) = |f'(x)|$.

EXERCISE 13.10. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, and suppose that $\overline{lip}_f(x) = \underline{lip}_f(x) < \infty$. Show that f has a finite derivative at x and that

$$\overline{lip}_f(x) = \underline{lip}_f(x) = |f'(x)|.$$

Hint: Use the Darboux property of continuous functions. As a more challenging exercise the student may wish to prove this without the assumption of continuity.

EXERCISE 13.11. If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $\underline{lip}_f(x) = \infty$ then either $f'(x) = \infty$ or $f'(x) = -\infty$. Give an example to show that continuity cannot be dropped.

13.4. Six growth lemmas

The growth lemmas we present all follow easily from the general limit theorems of Section 11.10 and proofs are left to the student. (Make liberal use of the Vitali covering theorem, i.e., replace $\mathcal{L}_*$ or $\mathcal{L}^*$ by $\mathcal{L}$ as needed.)

LEMMA 13.12. Let $f : \mathbb{R} \to \mathbb{R}$. If $\underline{lip}_f(z) < r$ for every $z \in E$ then

$$\mathcal{V}_{f*}(E) \leq r\mathcal{L}(E).$$

LEMMA 13.13. Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{lip}_f(z) > r > 0$ for every $z \in E$ then

$$r\mathcal{L}(E) \leq \mathcal{V}_f{}^*(E).$$

LEMMA 13.14. Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{lip}_f(z) < r$ for every $z \in E$ then

$$\mathcal{V}_f{}^*(E) \leq r\mathcal{L}(E).$$

LEMMA 13.15. Let $f : \mathbb{R} \to \mathbb{R}$. If $\underline{lip}_f(z) > r > 0$ for every $z \in E$ then

$$r\mathcal{L}(E) \leq \mathcal{V}_{f*}(E).$$

LEMMA 13.16. Let $f : \mathbb{R} \to \mathbb{R}$. If $\mathcal{V}_f{}^*(E) < \infty$, then $\overline{lip}_f(x) < \infty$ for almost every point x in E.

LEMMA 13.17. Let $f : \mathbb{R} \to \mathbb{R}$. If $\mathcal{V}_{f*}(E) < \infty$ then $\underline{lip}_f(z) < \infty$ for almost every point x in E.

13.5. Variational classifications of real functions

We shall use the following terminology to describe properties of real functions, properties each of which can be described directly in terms of the variational measures. In particular note that continuity is now a variational concept, distinguishing it from the notion of uniform continuity on an interval.

Let $f : \mathbb{R} \to \mathbb{R}$ and let E be any set of reals. The following terminology is used to describe different properties that a function might have that can be characterized by the variation. Most of them are concerned with the situation in which a measure is zero or is finite. [Some repeat or are closely related to earlier definitions that did not use a variational measure.]

(Zero variation): f has *zero variation* on E if

$$\mathcal{V}_f{}^*(E) = 0.$$

(Finite variation): f has *finite variation* on E if

$$\mathcal{V}_f{}^*(E) < \infty.$$

(σ-finite variation): f has *σ-finite variation* on E if

$$E \subset \bigcup_{k=1}^{\infty} E_k$$

so that, for each k,

$$\mathcal{V}_f{}^*(E_k) < \infty.$$

(Kolmogorov equivalent): f and g are *Kolmogorov equivalent* on E if

$$V^*(\Delta f - \Delta g, E) = 0.$$

(Vitali property on a set): f has the *Vitali property* on E provided that, for all subsets A of E,

$$\mathcal{V}_f{}^*(A) = \mathcal{V}_{f*}(A).$$

(Continuous at a point): f is *continuous* at a point x_0 provided that

$$\mathcal{V}_f{}^*(\{x_0\}) = 0.$$

(Weakly continuous at a point): f is *weakly continuous* at a point x_0 provided that

$$\mathcal{V}_{f*}(\{x_0\}) = 0.$$

($\mathcal{L}$-absolutely continuous on a set): f is *$\mathcal{L}$-absolutely continuous* on E provided that, for every set $N \subset E$ that has Lebesgue measure zero,

$$\mathcal{V}_f{}^*(N) = 0.$$

($\mathcal{L}$-singular on E): f is *$\mathcal{L}$-singular* on E provided

$$\mathcal{V}_f{}^*(E \setminus N) = 0$$

for some set $N \subset E$ that has Lebesgue measure zero.

(Mutually singular): Two functions f and g are said to be *mutually singular* on a set E if $E = E_1 \cup E_2$ and

$$\mathcal{V}_f{}^*(E_2) = \mathcal{V}_g{}^*(E_1) = 0.$$

(Saltus function): f is a *saltus function* on an interval $[a, b]$ if there is a countable set C so that

$$\mathcal{V}_f{}^*([a, b] \setminus C) = 0 \quad \text{and} \quad \mathcal{V}_f{}^*([a, b] \cap C) < \infty.$$

13.5.1. Some connections. Because of the close connection in this theory among the concepts of variation, derivative and integral these notions described above will be seen to have important implications as regards the differentiation and integration properties associated with the functions. The exercises are intended to show how easy it is to pass back and forth among the different concepts.

EXERCISE 13.18 (Kolmogorov equivalence). Let $f,\ g : \mathbb{R} \to \mathbb{R}$. Show that, if f and g are Kolmogorov equivalent on a set E, then $\mathcal{V}_f{}^*(E) = \mathcal{V}_g{}^*(E)$ and $\mathcal{V}_{f*}(E) = \mathcal{V}_{g*}(E)$.

EXERCISE 13.19 (Kolmogorov equivalence). Let $f,\ g : \mathbb{R} \to \mathbb{R}$. Show that, if $f'(x) = g'(x)$ at every point of a set E then f and g are Kolmogorov equivalent on E.

EXERCISE 13.20 (Singular functions). Let $f : \mathbb{R} \to \mathbb{R}$. Show that f is $\mathcal{L}$-singular on a set E if $f'(x) = 0$ at every point x of E.

Hint: If $f'(x) = 0$ for all $x \in E \setminus N$ use the fundamental theorem of the calculus (Theorem 11.32 from Section 11.11) to show that $\mathcal{V}_f{}^*(E \setminus N) = 0$.

EXERCISE 13.21 (Singular functions). Let $f : \mathbb{R} \to \mathbb{R}$. Show that if f is $\mathcal{L}$-singular on a set E then $f'(x) = 0$ at almost every point x of E.

Hint: Use the fundamental theorem of the calculus (Theorem 11.31 from Section 11.11).

EXERCISE 13.22 (Continuity). Show that if $f : \mathbb{R} \to \mathbb{R}$ has finite variation or σ-finite variation on a set E then f is continuous at each point of E with countably many exceptions.

Hint: Let $C = \{x \in E : \mathcal{V}_f{}^*(\{x\}) > 0\}$ and let $C_n = \{x \in E : \mathcal{V}_f{}^*(\{x\}) > 1/n\}$.

EXERCISE 13.23 (Weak continuity). Let $f : \mathbb{R} \to \mathbb{R}$. Show that f must be weakly continuous at every point with at most countably many exceptions.

Hint: Let

$$E = \{x : \liminf_{(I,x) \Rightarrow x} |\Delta f(I)| > 0\}$$

and

$$E = \{x : \liminf_{(I,x) \Rightarrow x} |\Delta f(I)| > 1/n\}.$$

The set of points where f is not weakly continuous is exactly the set $E = \bigcup_n E_n$. Note that $\beta = \{(I,x) : |\Delta f(I)| > 1/n\}$ is a full cover of E_n and apply the decomposition lemma from Section 11.1.5.

13.6. Local behaviour of functions

13.6.1. Local recurrence.

DEFINITION 13.24. A function $f : \mathbb{R} \to \mathbb{R}$ is *locally recurrent* at a point x if there is a sequence of points x_n with $x_n \neq x$ and $\lim_{n\to\infty} x_n = x$ so that $f(x) = f(x_n)$ for all n.

THEOREM 13.25. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that f is locally recurrent at every point of a set E. Then $\mathcal{V}_{f*}(E) = 0$.

Proof. If f is locally recurrent at every point of a set E then

$$\beta = \{(I,x) : \Delta f(I) = 0\}$$

is a fine cover of E. Thus

$$\mathcal{V}_{f*}(E) \le V(\Delta f, \beta) = 0.$$

■

13.6.2. Local monotonicity.

DEFINITION 13.26. A function $f : \mathbb{R} \to \mathbb{R}$ is *locally nondecreasing* at a point x if there is a $\delta > 0$ so that $\Delta f(I) \geq 0$ for every compact interval I containing x for which $\mathcal{L}(I) < \delta$.

THEOREM 13.27. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that f is locally nondecreasing at every point of a set E and that $\mathcal{V}_f{}^*(\{x\}) < \infty$ for each x in E. Then f has σ-finite variation on E.

Proof. Define
$$\beta = \{(I, x) : \Delta f(I) \geq 0\}$$
and notice that this is a full cover of E. Apply the decomposition from Section 11.11 for β. There is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^\infty E_n$ and a sequence of compact intervals $\{I_{kn}\}$ covering E so that if x is any point in E_n and I is any subinterval of I_{kn} that contains x then (I, x) belongs to β.

We check that f is nondecreasing on each set $D_{nk} = E_n \cap I_{kn}$ in a certain strong way. For if either x or y belongs to the set D_{nk} and $[x, y] \subset I_{kn}$ then one of the pairs $([x, y], x)$ or $([x, y], y)$ belongs to β which requires that $f(x) \leq f(y)$.

Let $c = \inf D_{nk}$ and $d = \sup D_{nk}$. Suppose that $c = d$. Then D_{nk} contains a single point c and $\mathcal{V}_f{}^*(\{c\}) < \infty$, i.e., $\mathcal{V}_f{}^*(D_{nk}) < \infty$. Suppose instead that $c < d$. Let $D'_{nk} = D_{nk} \cap (c, d)$ so that D_{nk} contains, at most, two points c and d more than the set D'_{nk}. Let $\beta' = \beta[D_{nk}] \cap \beta((c, d))$. Then β' is a full cover of D'_{nk}. Let $\pi = \{\{[c_i, d_i], x_i)\}$ be any subpartition contained in β'. We see from the manner in which f increases relative to the set D_{nk} that
$$\sum_i |f(d_i) - f(c_i)| \leq 2[f(d) - f(c)].$$
It follows that
$$\mathcal{V}_f{}^*(D'_{nk}) \leq V(\Delta f, \beta') \leq 2[f(d) - f(c)] < \infty.$$
Consequently,
$$\mathcal{V}_f{}^*(D_{nk}) \leq \mathcal{V}_f{}^*(D'_{nk}) + \mathcal{V}_f{}^*(\{c\}) + \mathcal{V}_f{}^*(\{d\}) < \infty$$
too, so that in either case $\mathcal{V}_f{}^*(D_{nk})$ is finite. It follows that $\mathcal{V}_f{}^*$ is σ-finite on the set E since that set has been expressed as a union of a sequence of sets on each of which $\mathcal{V}_f{}^*$ is σ-finite. ■

13.7. Derivates and variation

The finiteness of the derivates of a function $f : \mathbb{R} \to \mathbb{R}$ on a set E has implications for the variation $\mathcal{V}_f{}^*$ on E. The first version we can state is an immediate consequence of Lemma 13.14.

13.7.1. Lipschitz numbers and variation.

THEOREM 13.28. Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{lip}_f(z) < \infty$ for every $z \in E$ then f has σ-finite variation in E and is $\mathcal{L}$-absolutely continuous there.

Proof. Write
$$E_n = \{x \in E : \overline{lip}_f(z) < n\}.$$
By Lemma 13.14
$$\mathcal{V}_f{}^*(E_n \cap [-n, n]) \leq n\mathcal{L}(E_n \cap [-n, n]) < \infty.$$

It follows that f has σ-finite variation in E. Note then, that if N is a null subset of E,

$$\mathcal{V}_f{}^*(N) \leq \sum_{n=1}^{\infty} \mathcal{V}_f{}^*(E_n \cap N) \leq n\mathcal{L}(E_n \cap N) = 0.$$

This proves the final assertion. ■

13.7.2. Ordinary derivates and variation.

THEOREM 13.29. Let $f : \mathbb{R} \to \mathbb{R}$ and suppose at every point x of a set E that $\mathcal{V}_f{}^*(\{x\}) < \infty$ and that either $\overline{D}f(x) < \infty$ or $\underline{D}f(x) > -\infty$. Then f has σ-finite variation in E.

Proof. For example let us consider that the set E consists of all points at which $\underline{D}f(x) > -\infty$. Write

$$E_n = \{x : \underline{D}f(x) > -n\}.$$

Note that E is the union of the sequence of sets $\{E_n\}$.

Observe that the function $f_n(x) = f(x) + nx$ is locally nondecreasing at each $x \in E_n$. It follows (from Theorem 13.27) that f_n has σ-finite variation on E_n. But

$$\mathcal{V}_f{}^* \leq \mathcal{V}^*_{f_n} + n\mathcal{L}^*.$$

Thus f too has σ-finite variation on E_n. In consequence, f has σ-finite variation on E. ■

13.7.3. Dini derivatives and variation.

THEOREM 13.30. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and suppose that at every point x of a set E either

$$-\infty < \underline{D}^+ f(x) \leq \overline{D}^+ f(x) < \infty$$

or

$$-\infty < \underline{D}^- f(x) \leq \overline{D}^- f(x) < \infty.$$

Then f has σ-finite variation in E and is $\mathcal{L}$-absolutely continuous there.

Proof. We first show that, for any positive integer c, f has σ-finite variation and is $\mathcal{L}$-absolutely continuous on the set of points

$$A = \{x : -c < \underline{D}^+ f(x) \leq \overline{D}^+ f(x) < c\}.$$

The geometry of this situation is expressed by the covering relation

$$\beta = \{([x, x+h], x) : |\Delta f([x, x+h])| < c\mathcal{L}([x, x+h])\}.$$

This relation has none of the properties we have so far encountered, but a modification of our methods will handle.

First apply the ideas of the decomposition from Section 11.11 for β. There is an increasing sequence of sets $\{A_n\}$ with $A = \bigcup_{n=1}^{\infty} A_n$ and a sequence of compact intervals $\{I_{kn}\}$ covering A so that if x is any point in A_n and $[x, x+h]$ is any subinterval of I_{kn} then $([x, x+h], x)$ belongs to β.

In particular if $\{[c_i, d_i]\}$ is a sequence of subintervals of I_{kn} with endpoints in the set A_n, then a brief computation shows that

$$\sum_{i=1}^{\infty} \omega f([c_i, d_i]) \leq \sum_{i=1}^{\infty} 2c\mathcal{L}([c_i, d_i]) \leq 2c\mathcal{L}(I_{kn}).$$

Let C_{nk} denote the closure of the set $A_n \cap I_{kn}$. Since f is continuous this same inequality extends to points in that closure. Thus if $\{[c_i, d_i]\}$ is a sequence of intervals with endpoints in the compact set C_{nk}, then

$$\sum_{i=1}^{\infty} \omega f([c_i, d_i]) \leq \sum_{i=1}^{\infty} 2c\mathcal{L}([c_i, d_i]) \leq 2c\mathcal{L}(I_{kn}) < \infty.$$

Define a function g_n so that $g_n(x) = f(x)$ for all $x \in C_{nk}$ and extend to all of the real line using the device used in Section 4.12. Such a function g_n is evidently continuous and has bounded variation. The same inequality shows that g_n is absolutely continuous in the sense of Vitali and so also $\mathcal{L}$-absolutely continuous.

The computations of Exercise 13.3 can be used here to check that

$$V^*(\Delta f - \Delta g_n, C_{kn}) = 0.$$

This shows that f is Kolmogorov equivalent on each set C_{nk} to a continuous function of bounded variation. In particular $\mathcal{V}_f{}^*$ is finite on each set C_{nk}. It follows that $\mathcal{V}_f{}^*$ is σ-finite on A. The function f also inherits from g_n the property of being $\mathcal{L}$-absolutely continuous on C_{nk}.

Finally the set E of the theorem can be expressed as a union of a sequence of sets of the same type as A, so that $\mathcal{V}_f{}^*$ is σ-finite and vanishes on null subsets of each member of the sequence. The theorem follows. ■

13.8. Continuous functions have σ-finite fine variation

We turn now to the property for a function of σ-finite variation. "Most" continuous functions do not have σ-finite full variation. We begin by observing that all continuous functions have σ-finite *fine* variation.

THEOREM 13.31. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then $\mathcal{V}_{f*}$ must be σ-finite.

Proof. Define three sets E_1, E_2, and E_3. E_1 is the set of points at which f is locally nondecreasing. E_2 is the set of points at which $-f$ is locally nondecreasing. E_3 is the set of points at which f is locally recurrent. Since f is continuous it has the Darboux property. From that we see that $E_1 \cup E_2 \cup E_3 = \mathbb{R}$ since there are no other possibilities.

But $\mathcal{V}_{f*}(E_3) = 0$ and $\mathcal{V}_f{}^*$ is σ-finite on E_1 and E_2 (Theorem 13.27). It follows that the smaller measure $\mathcal{V}_{f*}$ must be σ-finite. ■

COROLLARY 13.32. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then $\underline{lip}_f(z) < \infty$ for almost every point x in E.

Proof. This follows immediately from Lemma 13.17 now that we know that, for every continuous function f, the measure $\mathcal{V}_{f*}$ must be σ-finite. ■

13.9. Functions having σ-finite full variation

THEOREM 13.33. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and E a real set. Then the following are equivalent:

(a) f has σ-finite variation on E,
(b) there is a sequence $\{E_n\}$ of compact sets covering E so that f has finite variation on each E,

(c) there is a sequence $\{E_n\}$ of compact sets covering E so that on each E_n, f is Kolmogorov equivalent to some continuous function of bounded variation.

Proof. The implication (b)⇒(a) is trivial. The implication (c)⇒(b) is easy: if (c) holds then, for some function of bounded variation $g_n : \mathbb{R} \to \mathbb{R}$, the equivalence relation

$$V^*(\Delta f - \Delta g_n, E_n) = 0$$

implies that $\mathcal{V}_f{}^*(E_n) = \mathcal{V}^*_{g_n}(E_n) < \infty$.

Thus the proof is completed by showing that (a)⇒(c). It is enough to consider the situation for which E is a bounded set for which $\mathcal{V}_f{}^*(E) < \infty$. Choose a full cover β of E and a real number t so that

$$V(\Delta f, \beta) < t < \infty.$$

Apply the decomposition in Lemma 11.11 to β. Accordingly there is an increasing sequence of sets $\{B_n\}$ with $E = \bigcup_{n=1}^{\infty} B_n$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E so that if x is any point in B_n and I is any subinterval of I_{kn} that contains x then (I, x) belongs to β.

Let $A_{kn} = B_n \cap I_{kn}$. We check some facts about the variation of f on A_{kn}. Suppose that $\{[a_i, b_i]\}$ is any disjointed sequence of compact subintervals of I_{kn} each of which contains at least one point, say x_i, of B_n. Then $\{([a_i, b_i], x_i)\}$ must form a subpartition contained in β. Consequently

$$\sum_i |f(b_i) - f(a_i)| \leq V(\Delta f, \beta) < t.$$

Now let C_{kn} denote the closure of A_{kn}, i.e., C_{kn} is the smallest compact set that contains A_{kn}. We extend these considerations to estimating the variation of f on the larger set C_{kn}. Suppose now that $\{[a_i, b_i]\}$ is any disjointed sequence of compact subintervals of I_{kn} each of which contains at least one point of C_{nk}. We enlarge each interval slightly as needed to ensure that the intervals remain disjointed but contain also a point, now, of the dense subset A_{kn}. As f is continuous we can do this without much of an increase in the sums, and so we can certainly guarantee that for the given sequence $\{[a_i, b_i]\}$ that

$$\sum_i |f(b_i) - f(a_i)| < 2t < \infty.$$

Let us define a function g_{nk} so as to be equal to $f(x)$ on the compact set C_{kn} and extended to the real line by the device used in Section 4.12. Such a function g_{nk} is continuous and has bounded variation.

The computations of Exercise 13.3 can be used here to check that

$$V^*(\Delta f - \Delta g_{nk}, C_{kn}) = 0.$$

As every compact set from the sequence $\{C_{kn}\}$ can be treated the same way, we have verified the implication (a)⇒(c) provided we merely relabel the full collection $\{C_{kn}\}$ as a single sequence $\{E_n\}$. ■

13.10. Variation on compact sets

We can refine our analysis of σ-finite variation with a few further steps.

THEOREM 13.34. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and E a compact set. Then the following are equivalent:

(a) f has σ-finite variation on E.
(b) Every nonempty compact subset S of E has a portion $S \cap (a, b)$ on which f has finite variation.
(c) f has σ-finite variation on every null set $Z \subset E$ that is a $\mathcal{G}_\delta$ set.

Proof. By a $\mathcal{G}_\delta$ set we mean a set Z of the form $Z = \bigcap_{n=1}^{\infty} G_n$ for some sequence $\{G_n\}$ of open sets. Every closed set can be written in this form.

We begin with (a)⇒(b). As we have seen in Theorem 13.33, if f has σ-finite variation on E, then there is a sequence of compact sets $\{E_n\}$ covering the compact set S so that $\mathcal{V}_f{}^*(E_n) < \infty$ for each n. By the Osgood-Baire theorem (Theorem 3.35) there must be a portion $S \cap (a, b)$ of E contained in one at least from the sequence $\{E_n\}$. In particular, for some n, $\mathcal{V}_f{}^*(S \cap (a, b) \leq \mathcal{V}_f{}^*(E_n) < \infty$ as required to prove (b).

Let us now prove that (b)⇒(a) Suppose that every nonempty closed subset S of E has a portion $S \cap (a, b)$ on which f has finite variation. Let G denote the real set consisting of all real x with the property that there is a $\delta(x) > 0$ so that f has σ-finite variation on the set $E \cap (x - \delta(x), x + \delta(x))$. Note that

$$G = \bigcup_{x \in G} (x - \delta(x), x + \delta(x))$$

so G is open.

Consider the set $G \cap E$. Any point in this set would be contained in an open interval (c, d) with rational endpoints so that f has σ-finite variation on $G \cap (c, d)$. It follows that f has σ-finite variation on $G \cap E$. If $G \supset E$ then, we deduce that f has σ-finite variation on E as we wished to prove to verify (a).

Suppose, in order to obtain a contradiction that G does not contain E. Let $E' = E \setminus G$. This would be a nonempty closed subset of E and so, by hypothesis, there would have to be a portion $E' \cap (a, b)$ on which f has finite variation. But if f has finite variation on $E' \cap (a, b)$ and also, evidently, has σ-finite variation on $E \setminus E'$ then f must have σ-finite variation on $E \cap (a, b)$. Every point of this set should belong to G which is impossible in view of the assumption that $E' \cap (a, b)$ is a portion. This contradiction completes our proof that (b)⇒(a).

The implication (a)⇒(c) is trivial. To complete the proof, then, it will suffice to verify that (c)⇒(b). Suppose that f has σ-finite variation on every set $Z \subset E$ that is a $\mathcal{G}_\delta$ set of $\mathcal{L}$-measure zero.

Let S be a nonempty closed subset of E. To verify (b) we need to find a portion of S on which f has finite variation. If S is a null set then we are almost there. A closed set is also of type $\mathcal{G}_\delta$. Thus, $\mathcal{V}_f$ is σ-finite on S by hypothesis. As we have already argued above, in this situation we are assured that S has a portion $S \cap (a, b)$ on which f has finite variation.

Suppose instead that S is a closed set having positive measure. Exercise 13.35, which follows the proof, shows exactly how to choose a null subset Z of S that is a $\mathcal{G}_\delta$-set that is dense in S. By our assumption (c), there must be a portion $Z \cap (a, b)$

on which f has σ-finite variation. We apply Theorem 13.33 to obtain a sequence of compact sets $\{K_n\}$ whose union includes $Z \cap (a,b)$ so that each $\mathcal{V}_f{}^*(K_n) < \infty$.

Apply the Osgood-Baire theorem, now to the sequence of compact sets $\{K_n\}$ that covers the $\mathcal{G}_\delta$-set $Z \cap (a,b)$. Recall that the Osgood-Baire theorem, stated in Section 3.8 for closed sets, applies equally well to $\mathcal{G}_\delta$-sets. Thus we can conclude that there is a portion $Z \cap (c,d)$ and an integer k so that $Z \cap (c,d) \subset K_k$. Since Z is dense in the compact set S we also have $S \cap (c,d) \subset K_k$. In particular

$$\mathcal{V}_f{}^*(S \cap (c,d)) \leq \mathcal{V}_f{}^*(K_n) < \infty.$$

We have obtained again (but this time without the additional assumption that S has measure zero) exactly property (b). ■

EXERCISE 13.35. Let S be a compact set. Show that there is a subset Z of S that is of type $\mathcal{G}_\delta$, is a null set, and is dense in S.

Hint: If S is a null set then $Z = S$ solves the exercise. Otherwise construct such a set by first taking a countable dense subset Z_1 of S. [The endpoints of the complementary intervals will suffice, unless S contains an interval. If S does contain an interval then include all rational numbers in that interval.] Now Z_1 is a countable subset of S and so has measure zero. For each integer n choose an open set G_n containing Z_1 with $\mathcal{L}(G_n) < 1/n$. Finally check that $Z = S \cap \bigcap_{n=1}^{\infty} G_n$ is a $\mathcal{G}_\delta$-set and that $\mathcal{L}(Z) = \mathcal{L}(Z_1) = 0$.

13.11. $\mathcal{L}$-absolutely continuous functions

As a corollary to Theorem 13.34 immediately we have a special observation, since an $\mathcal{L}$-absolutely continuous function must have finite variation (indeed zero variation) on every set of $\mathcal{L}$-measure zero.

COROLLARY 13.36. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function that is $\mathcal{L}$-absolutely continuous on a compact set E. Then f has σ-finite variation there.

13.12. Vitali property and differentiability

In this section we show that differentiability on a set implies the Vitali property on that set. The converse appears in the next section: the Vitali property on a set implies almost everywhere differentiability.

THEOREM 13.37. Let $f : \mathbb{R} \to \mathbb{R}$ have a finite derivative at every point of a set E. Then f has the Vitali property on E and, moreover,

$$\mathcal{V}_f{}^*(E) = \mathcal{V}_{f*}(E) = \int_E |f'(x)|\,dx.$$

Proof. Since f is differentiable at each point x of E we have for such points

$$\lim_{I \Rightarrow x} \frac{\Delta f(I) - f'(x)\mathcal{L}(I)}{\mathcal{L}(I)} = 0.$$

The fundamental theorem of the calculus (Theorem 11.32 from Section 11.11) supplies the equivalence relation

$$V^*(\Delta f - f'\mathcal{L}, E) = 0.$$

From this we know immediately that

$$\mathcal{V}_f{}^*(E) = V^*(\Delta f, E) = V^*(f'\mathcal{L}, E) = \int_E |f'(x)|\,dx.$$

and that

$$\mathcal{V}_{f*}(E) = V_*(\Delta f, E) = V_*(f'\mathcal{L}, E).$$

Let A be the set of *all* points where f has a derivative. From Exercise 7.15 we recall that this is almost closed. The function f' is defined on A and it is almost continuous. Thus from the Vitali covering theorem itself (i.e., the identity of $\mathcal{L}_*$ and $\mathcal{L}$) we can deduce that this variation is expressible as a usual Lebesgue integral

$$V_*(f'\mathcal{L}, E) = \int_E |f'(x)|\, dx.$$

This verifies the identity. ∎

13.13. Differentiability properties from the Vitali property

THEOREM 13.38. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function that has the Vitali property on a set E. Then f has a finite derivative at almost every point of E and, except at the points of a set N for which $\mathcal{V}_f{}^*(N) = 0$, f has a finite or infinite derivative $f'(z)$.

Proof. We need work only with the Lipschitz numbers here. Recall that if $\underline{lip}_f(z) = \infty$ then necessarily f has an infinite derivative, $f'(z) = \infty$ or $f(z) = -\infty$ (see Exercise 13.11). Also if

$$\underline{lip}_f(z) = \overline{lip}_f(z) < \infty$$

then f has a finite derivative at z (see Exercise 13.10).

It is enough to prove the theorem under the assumption that E is a bounded set. We examine

$$A = \{x \in E : \underline{lip}_f(x) < \overline{lip}_f(x)\}.$$

As is usual in arguments of this type, introduce rational numbers $0 < r < s$ and the subsets

$$A_{rs} = \{x \in A : \underline{lip}_f(x) < r < s < \overline{lip}_f(x)\}.$$

Note that A is the countable union of this collection of sets taken over all rationals r and s with $r < s$.

By the growth lemmas of Section 13.4 we obtain

$$\mathcal{V}_{f*}(A_{rs}) \le r\mathcal{L}(A_{rs}) \le s\mathcal{L}(A_{rs}) \le \mathcal{V}_f{}^*(A_{rs}).$$

Our assumption that f has the Vitali property on E gives the identity $\mathcal{V}_f = \mathcal{V}_{f*}$ on each of these subsets of E. None of these numbers are infinite, $r < s$, and so the inequality makes sense only in the case that $\mathcal{V}_f{}^*(A_{rs}) = \mathcal{L}(A_{rs}) = 0$. Consequently $\mathcal{V}_f{}^*(A) = \mathcal{L}(A) = 0$.

At every point x in $E \setminus A$ we know that either

$$\underline{lip}_f(x) = \overline{lip}_f(x) < \infty$$

or else

$$\underline{lip}_f(x) = \overline{lip}_f(x) = +\infty.$$

In the former case, as we have already noted, f has a finite derivative and in the latter case f has an infinite derivative. This latter case can occur only on a set of Lebesgue measure zero (as a consequence of Lemma 13.17). ∎

13.14. The Vitali property and variation

The Vitali property is closely related to the finiteness of the variation. Indeed, since the fine variation $\mathcal{V}_{f*}$ of a continuous function f is always σ-finite, we know that the identity $\mathcal{V}_{f*}(E) = \mathcal{V}_f{}^*(E)$ can only hold if f has σ-finite variation on E.

13.14.1. Monotonic functions.

THEOREM 13.39. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, strictly increasing function. Then f has the Vitali property.

Proof. Theorem 13.4 supplies the identity

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

■

THEOREM 13.40. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous, monotonic nondecreasing function. Then f has the Vitali property.

Proof. Let $\epsilon > 0$ and define a new function $g(x) = f(x) + \epsilon x$. The function g is continuous and strictly increasing so, by the previous theorem, $\mathcal{V}_{g*} = \mathcal{V}_g{}^*$. From Lemma 11.17 we deduce the inequalities

$$\mathcal{V}_f{}^* \leq \mathcal{V}_g{}^* \leq \mathcal{V}_f{}^* + \epsilon\mathcal{L}^*$$

and

$$\mathcal{V}_{f*} \leq \mathcal{V}_{g*} \leq \mathcal{V}_{f*} + \epsilon\mathcal{L}^*.$$

From these two inequalities and the identity $\mathcal{V}_{g*} = \mathcal{V}_g^*$ we can deduce $\mathcal{V}_f{}^* = \mathcal{V}_{f*}$. ■

EXERCISE 13.41. Let $f : \mathbb{R} \to \mathbb{R}$ be a monotonic, nondecreasing function. Show that if $\mathcal{V}_{f*}(\{x\}) = \mathcal{V}_f{}^*(\{x\})$ for a point x then f must be continuous at x.

13.14.2. Functions of bounded variation.

THEOREM 13.42. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function that is locally of bounded variation. Then f has the Vitali property on the real line.

Proof. Fix a compact interval $[a, b]$ and let g be the total variation function of f on $[a, b]$. We know that this relation between a function and its total variation function requires the identity

$$V^*(\Delta g - |\Delta f|, (a, b)) = 0.$$

In particular $\mathcal{V}_f{}^*(E) = \mathcal{V}_g{}^*(E)$ and $\mathcal{V}_{f*}(E) = \mathcal{V}_{g*}(E)$ for all subsets E of (a, b). By the previous theorem $\mathcal{V}_g{}^*(E) = \mathcal{V}_{g*}(E)$ and so $\mathcal{V}_f{}^*(E) = \mathcal{V}_{f*}(E)$ follows. This argument produces the identity we require on all bounded sets, and the extension to arbitrary sets follows from measure properties. ■

13.14.3. Functions of σ-finite variation.

THEOREM 13.43. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then f has σ-finite variation on a set E if and only if f has the Vitali property on E.

Proof. We already know that the Vitali property for a continuous function will imply σ-finite variation. Let us prove the converse.

Suppose that f is continuous function that has σ-finite variation on E. By Theorem 13.33 there is a sequence of compact sets $\{E_n\}$ covering E and a sequence of functions g_n each continuous and locally of bounded variation so that

$$V^*(\Delta f - \Delta g_n, E_n) = 0 \tag{13.1}$$

We know then, from the previous theorem, that $\mathcal{V}_{g_n *} = \mathcal{V}^*_{g_n}$. We also know that the equivalence (13.1) requires that $\mathcal{V}_{g_n}{}^* = \mathcal{V}_f{}^*$ and $\mathcal{V}_{g_n *} = \mathcal{V}_{f*}$ on all subsets of E_n.

Introduce the notation

$$A_n = E_n \setminus \bigcup_{k<n} E_k$$

so that $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty E_n$ and the sets $\{A_n\}$ are pairwise disjoint, almost closed sets. The student should justify that the following computations are permitted:

$$\mathcal{V}_f{}^*(E) = \sum_{n=1}^\infty \mathcal{V}_f{}^*(E \cap A_n) = \sum_{n=1}^\infty \mathcal{V}_{g_n}{}^*(E \cap A_n) =$$

$$\sum_{n=1}^\infty \mathcal{V}_{g_n *}(E \cap A_n) = \sum_{n=1}^\infty \mathcal{V}_{f*}(E \cap A_n) = \mathcal{V}_{f*}(E).$$

As this applies as well to any subset of E we see that f must have the Vitali property on E as required. ■

Corollary 13.44. If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function that is $\mathcal{L}$-absolutely continuous on a compact set E, then f has the Vitali property on E.

Proof. Use Corollary 13.36. ■

13.15. Characterization of the Vitali property

The class of functions satisfying the Vitali property on a set is fundamental to an understanding of the calculus program demanding the relation among the concepts of derivative, integral and variation. We have already found a number of characterizations in Theorem 13.33 and Theorem 13.34. Here are some more. Some are easy consequences of what we have proved [e.g., (a) and Theorem 13.31 immediately imply (b)]. Others are left as entertainments for the student.

Theorem 13.45. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous real function and let E be a compact set. The following are equivalent:

(a) f has the Vitali property on E.
(b) f has σ-finite variation on E.
(c) there is a sequence of compact sets $\{E_n\}$ with $E = \bigcup_{n=1}^\infty E_n$ so that for each n there is a continuous function g_n that is locally of bounded variation so that f and g_n are Kolmogorov equivalent on E_n.
(d) f has a derivative (finite or infinite) at $\mathcal{V}_f{}^*$-almost every point of E.
(e) There is a continuous, increasing function g so that

$$\limsup_{(I,x)\Rightarrow x} \left| \frac{\Delta f(I)}{\Delta g(I)} \right| < \infty$$

at every point $x \in E$.

(f) There is a continuous, increasing function g and a real function f_1 so that
$$V^*(\Delta f - f_1 \Delta g, E) = 0.$$
(g) There is a continuous, increasing function g so that the composed function $f \circ g$ has a finite derivative everywhere in the compact set $g^{-1}(E)$.

13.16. Characterization of $\mathcal{L}$-absolute continuity

The Vitali property expresses the most important property arising in studies of the derivative in the calculus. The special subclass of $\mathcal{L}$-absolutely continuous functions plays its most significant role in the integration theory. Here are some similar characterizations for this class, most easily proved from previously proved statements or techniques.

Theorem 13.46. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and let E be a compact set. The following are equivalent:

(a) f is $\mathcal{L}$-absolutely continuous on E.
(b) f has σ-finite variation on E and is $\mathcal{L}$-absolutely continuous there.
(c) there is a sequence of compact sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ so that for each n there is a continuous function g_n that is of locally of bounded variation and absolutely continuous in the sense of Vitali so that f and g_n are Kolmogorov equivalent on E_n.
(d) f has a finite derivative at $\mathcal{V}_f{}^*$-almost every point of E.
(e) There is an increasing, $\mathcal{L}$-absolutely continuous function g so that
$$\limsup_{(I,x)\to x} \left| \frac{\Delta f(I)}{\Delta g(I)} \right| < \infty$$
at every point $x \in E$.
(f) There is an increasing, $\mathcal{L}$-absolutely continuous function g and a real function f_1 so that
$$V^*(\Delta f - f_1 \Delta g, E) = 0.$$

13.17. Mapping properties

For any set E and any function $f : \mathbb{R} \to \mathbb{R}$ the image of E under the mapping f is written as
$$f(E) = \{f(x) : x \in E\}.$$
We already know some properties of the image set for continuous functions. From Section 4.10 we recall that the image of any compact interval $[a, b]$ under f is again a compact interval. It is easy to check that that the image of any compact set E under f is again a compact set $f(E)$. A natural question is whether the image of an almost closed set must also be almost closed.

Theorem 13.47. Let $f : \mathbb{R} \to \mathbb{R}$ be an almost continuous function and P an almost closed set. The following are equivalent:

(M): $f(E)$ is almost closed for every almost closed subset E of P,
(N): $\mathcal{L}(f(N)) = 0$ for every subset N of P for which $\mathcal{L}(N) = 0$.

Proof. Suppose that E is almost closed and that the second statement of the theorem holds. We need consider only the case where E is bounded. Since f is almost continuous, then by definition, we can find open sets G_n so that $\mathcal{L}(G_n) <$

$1/n$, $E \setminus G_n$ is compact and f is equal to a continuous function $g_n : \mathbb{R} \to \mathbb{R}$ on the compact set $E \setminus G_n$.

In particular

$$E = Z \cup \bigcup_{n=1}^{\infty} (E \setminus G_n)$$

where

$$Z = E \cap \bigcap_{n=1}^{\infty} G_n$$

has $\mathcal{L}$-measure zero. By hypothesis $f(Z)$ must be a set of $\mathcal{L}$-measure zero and hence is almost closed. Also each

$$f(E \setminus G_n) = g_n(E \setminus G_n)$$

is a compact set (since the continuous function g_n maps compact sets to compact sets). In particular each set here is also almost closed. Thus

$$f(E) = f(Z) \cup \bigcup_{n=1}^{\infty} f(E \setminus G_n)$$

displays $f(E)$ as the union of a sequence of almost closed sets. Thus $f(E)$ is also almost closed.

Conversely suppose that the first statement of the theorem does not hold, yet the second does. Then there is a set $Z \subset P$ for which $\mathcal{L}(Z) = 0$ and yet $f(Z)$ does not have $\mathcal{L}$-measure zero. For (b) to be true, however, $f(Z)$ should be an almost closed set of positive measure. Such a set must have a subset A that is not almost closed.

We shall not pause to prove this assertion but leave it as a project for the student to find elsewhere (or prove). A proof will require use of a logical principle that is beyond our elementary calculus course.

Then there is a set $Z_1 \subset Z$ with $f(Z_1) = A$. The set Z_1 must be almost closed merely because $\mathcal{L}(Z_1) \leq \mathcal{L}(Z) = 0$. But then f maps an almost closed set Z_1 to a set $f(Z_1) = A$ that is not almost closed. We have contradicted the second statement thus completing the proof. ∎

13.18. Lusin's conditions

Definition 13.48. A function $f : \mathbb{R} \to \mathbb{R}$ is said to satisfy *Lusin's conditions* on a set P when these equivalent conditions hold:

- **(M):** $f(E)$ is almost closed for every almost closed subset E of P,
- **(N):** $\mathcal{L}(f(N)) = 0$ for every subset N of P for which $\mathcal{L}(N) = 0$.

Theorem 13.49. If $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{L}$-absolutely continuous on an almost closed set P then f satisfies Lusin's conditions on P.

Proof. This follows immediately from Theorem 13.5 that asserts that $\mathcal{L}(f(N))$ is smaller than the full variation of f on N. Thus for every null set $N \subset P$,

$$\mathcal{L}(f(N)) \leq {\mathcal{V}_f}^*(N) = 0.$$

∎

13.19. Banach-Zarecki Theorem

In the converse direction we should expect that Lusin's conditions play a role in characterizing the important property of absolutely continuity.

THEOREM 13.50 (Banach-Zarecki). Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and E a compact set. Then the following are necessary and sufficient conditions in order that f is $\mathcal{L}$-absolutely continuous on E:

(a) f has σ-finite variation on E, and
(b) f satisfies Lusin's condition on E.

Proof. Certainly if f is $\mathcal{L}$-absolutely continuous then we already know that (a) holds because of Theorem 13.34 and that (b) holds because of Theorem 13.49.

Conversely let us suppose that (a) and (b) now hold. We know from Theorem 13.33 that when f has σ-finite variation on a compact set E, there is a sequence $\{E_n\}$ of compact sets covering E and a sequence of continuous functions of bounded variation g_n so that f and g_n are Kolmogorov equivalent on E_n. Recall in the proof that the construction there required $f = g_n$ on the set E_n. We can insist on that here. Moreover the functions g_n in the proof that extended f were also chosen to be merely linear or constant in the intervals complementary to E_n. We can insist also on that here.

We note that the condition (b) of the theorem asserting that f satisfies Lusin's condition on E means that g_n satisfies this same condition on E_n. Moreover by the nature of the construction the function g_n satisfies Lusin's condition on all sets. The proof is completed now by addressing the special case of proving that g_n is $\mathcal{L}$-absolutely continuous.

Note that each g_n constructed in our proof above satisfies the hypotheses of Exercise 13.52 below. Indeed, since g_n has bounded variation on every interval it is differentiable outside of a set N of $\mathcal{L}$-measure zero. The assumption of Lusin's condition on g_n then provides $\mathcal{L}(g_n(N)) = 0$. The finiteness of

$$\mathcal{V}_{g_n}(\mathbb{R} \setminus N) = \int_{\mathbb{R}\setminus N} |g_n'(x)|\, dx$$

follows from the fact that g_n, as constructed have finite variation.

Now let Z be any set for which $\mathcal{L}(Z) = 0$. Let $\epsilon > 0$ and choose $\delta > 0$ by applying the Exercise 13.52 to this function g_n. Choose an open set $G \subset Z$ with $\mathcal{L}(G) < \delta$. Choose any full cover β of Z; then $\beta(G)$ is also a full cover of Z and the exercise provides

$$V^*(\Delta g_n, Z) \leq V(\Delta g_n, \beta(G)) < \epsilon.$$

From this we deduce that $\mathcal{V}_{g_n}(Z) = 0$. In consequence g_n is $\mathcal{L}$-absolutely continuous.

From this we can prove that f is $\mathcal{L}$-absolutely continuous on the set E in question. For if Z is a set of $\mathcal{L}$-measure zero then ${\mathcal{V}_{g_n}}^*(Z) = 0$ will imply that

$${\mathcal{V}_f}^*(E_n \cap Z) = {\mathcal{V}_{g_n}}^*(E_n \cap Z) = 0$$

and hence that

$${\mathcal{V}_f}^*(E \cap Z) \leq \sum_{n=1}^{\infty} {\mathcal{V}_f}^*(E_n \cap Z) = 0.$$

This will then show that f is $\mathcal{L}$-absolutely continuous on E. ∎

COROLLARY 13.51. Let $f : [a,b] \to \mathbb{R}$. The following are necessary and sufficient conditions in order that f is absolutely continuous in the sense of Vitali on $[a,b]$:

(a) f is continuous,
(b) f has bounded Jordan variation on $[a,b]$, and
(c) f satisfies Lusin's conditions on $[a,b]$.

A crucial step in the proof of the theorem uses the following classical problem:

EXERCISE 13.52. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Suppose that g is differentiable at each point with the exception of points in a set N for which $\mathcal{L}(g(N)) = 0$ and suppose that $\int_{\mathbb{R}\setminus N} |g'(x)|\,dx < \infty$. Show that, for every $\epsilon > 0$, there is a $\delta > 0$ so that any sequence of nonoverlapping intervals $\{[c_n, d_n]\}$ for which $\sum_n \mathcal{L}([c_n, d_n]) < \delta$ it follows that

$$\sum_n |\Delta g([c_n, d_n])| < \epsilon.$$

Hint: Take g' to denote the derivative of g where that exists and 0 otherwise; such a function is almost continuous and we will be able to apply Theorem 8.14.

Observe first that if $[c,d]$ is any compact interval then

$$|\Delta g([c,d])| \le \int_{[c,d]} |g'(x)|\,dx.$$

This follows from the fact that g is continuous so that

$$|\Delta g([c,d])| \le \mathcal{L}(g([c,d]) \le \mathcal{L}(g([c,d] \cap N) + \mathcal{L}(g([c,d] \setminus N) =$$

$$\mathcal{L}(g([c,d] \cap N) \le \mathcal{V}_g([c,d] \cap N) = \int_{[c,d]} |g'(x)|\,dx.$$

Use Theorem 13.5 and Theorem 13.37.

Now apply Theorem 8.14 to obtain, for every $\epsilon > 0$, a $\delta > 0$ so that if G is an open set with $\mathcal{L}(G) < \delta$ then

$$\int_G |g'(x)|\,dx < \epsilon.$$

In particular, if we are given any sequence of nonoverlapping intervals $\{[c_n, d_n]\}$ for which $\sum_n \mathcal{L}([c_n, d_n]) < \delta$ then there is an open set G covering these intervals for which $\mathcal{L}(G) < \delta$; it follows that

$$\sum_n |\Delta g([c_n, d_n])| \le \sum_n \int_{[c_n,d_n]} |g'(x)|\,dx \le \int_G |g'(x)|\,dx < \epsilon.$$

CHAPTER 14

The Integral

In this chapter we return to complete our account of the integral

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

Now we are furnished with the measure and variational tools of the preceding chapters. We will restate some results from earlier chapters, but most results in this chapter are less easily proved without the variational techniques we have just learned.

We can summarize, a bit, what we know so far.

The relation of a function f to its indefinite integral $F : [a,b] \to \mathbb{R}$

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

signifies that $F'(x) = f(x)$ at all points of the interval $[a,b]$ excepting some null set and that the function F does not grow on that null set. Such a function F is absolutely continuous (not necessarily in the narrower Vitali sense) and has σ-finite variation. These ideas can be expressed using the measure $\mathcal{V}_F{}^*$. The main theoretical tool for proving nearly everything and connecting a function with its indefinite integral is the Henstock criterion, now expressible simply as

$$\int_a^b |\Delta F - f\mathcal{L}| = 0.$$

14.1. Rudimentary properties of the integral

Having now a powerful set of characterizations of the integral we can establish many properties without a return to the definition. As exercises the reader is encouraged to apply the theory developed to the following. Remember, in addition to the usual measure properties, to use also the simple subadditive inequalities

$$V^*(h_1 + h_2, E) \leq V^*(h_1, E) + V^*(h_2, E)$$

and

$$\int_a^b |h_1 + h_2|\,dx \leq \int_a^b |h_1|\,dx + \int_a^b |h_2|\,dx.$$

Exercise 14.1 (Integrability on subintervals). If $f : [a,b] \to \mathbb{R}$ is integrable on $[a,b]$ then f is also integrable on any subinterval $[c,d]$.

EXERCISE 14.2 (Linear combinations). Linear combinations of integrable functions are also integrable, i.e., if $f_1,\ f_2 : [a,b] \to \mathbb{R}$ are integrable on $[a,b]$ then so too is the linear combination $rf_1 + sf_2$ for any real r and s and

$$\int_a^b [rf_1(x) + sf_2(x)]\,dx = r\int_a^b f_1(x)\,dx + s\int_a^b f_2(x)\,dx.$$

EXERCISE 14.3 (Change of variable). Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing, differentiable function. If $f : \mathbb{R} \to \mathbb{R}$ is integrable on $[\phi(a), \phi(b)]$ then

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\phi'(t)\,dt.$$

EXERCISE 14.4 (Adjacent interval extension). Suppose that f is integrable on $[a,b]$ and also on $[b,c]$. Then f is integrable on $[a,c]$.

Hint: Let $F(z) = 0$ for $z \le a$, $F(z) = \int_a^z f(x)\,dx$ for $a < z \le b$,

$$F(z) = \int_a^b f(x)\,dx + \int_b^z f(x)\,dx$$

for $b < z \le c$, and finally $F(z) = F(c)$ for $z > c$. Check that

$$V^*(\Delta F - f\mathcal{L}, [a,b]]) = V^*(\Delta F - f\mathcal{L}, [[b,c]]) = 0$$

and from this deduce that

$$V^*(\Delta F - f\mathcal{L}, [a,c]]) = 0.$$

EXERCISE 14.5 (Cauchy extension property). Let $F : \mathbb{R} \to \mathbb{R}$ be continuous and suppose that $f : \mathbb{R} \to \mathbb{R}$ is integrable on all intervals $[c,d]$ with $a < c < d < b$ and that

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

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

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

Hint: Obtain $V^*(\Delta F - f\mathcal{L}, [c,d]) = 0$ for any $a < c < d < b$. Take $c_n \searrow a$ and $d_n \nearrow b$ and check that

$$V^*(\Delta F - f\mathcal{L}, (a,b)) \le \sum_{i=1}^{\infty} V^*(\Delta F - f\mathcal{L}, [c_n, d_n]) = 0.$$

Derive that, in fact, $V^*(\Delta F - f\mathcal{L}, [a,b]) = 0$.

EXERCISE 14.6 (Harnack extension property). Let $F : \mathbb{R} \to \mathbb{R}$, let E be a closed subset of $[a,b]$, and let $\{(a_i, b_i)\}$ be the sequence of intervals complementary to E in (a,b). Suppose that

(a) $f(x) = 0$ for all $x \in E$,
(b) f is integrable on all intervals $[a_i, b_i]$, and
(c) $\displaystyle\sum_{i=1}^{\infty} \sup_{a_i \le c_i < d_i \le b_i} \left| \int_{c_i}^{d_i} f(x)\,dx \right| < \infty.$

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

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

Hint: Define F appropriately, starting with

$$F(z) = \sum_{[a_i,b_i]\subset[a,z]} \int_{a_i}^{b_i} f(x)\,dx$$

for any $z \in E$ and, for $z \in (a_j, b_j)$, set

$$F(z) = F(a_j) + \int_{a_j}^{z} f(x)\,dx.$$

Obtain $V^*(\Delta F - f\mathcal{L}, [a,b]) = 0$ from $V^*(f\mathcal{L}, E) = 0$, $V^*(\Delta F, E) = 0$, and

$$V^*(\Delta F - f\mathcal{L}, [a,b] \setminus E) \le \sum_{i=1}^{\infty} V^*(\Delta F - f\mathcal{L}, [a_i, b_i]).$$

Exercise 14.7 (Lebesgue equivalence). For two integrable functions f,g : $[a,b] \to \mathbb{R}$ the following are equivalent:

(a) f and g are Lebesgue equivalent on $[a,b]$.
(b) $\int_a^b |f(x) - g(x)|\,dx = 0$.
(c) $\int_c^d f(x)\,dx = \int_c^d g(x)\,dx$ for all intervals $[c,d] \subset [a,b]$.

Hint: Recall that two functions f and g are said to be Lebesgue equivalent on a set E if $f(x) = g(x)$ for $\mathcal{L}$-almost every $x \in E$, i.e., if $\{x \in E : f(x) \neq g(x)\}$ is a null set. If $V^*([f-g]\mathcal{L}, E) = 0$, write $E_n = \{x \in E : |f(x) - g(x) > 1/n\}$ and show that $\mathcal{L}(E_n) \le nV^*(([f-g]\mathcal{L}, E_n)$. From this deduce that $\mathcal{L}(E) = 0$. For the final item note that, if F and G are the indefinite integrals of f and g, then $V^*(\Delta F - \Delta G, [a,b]) = 0$ if and only if $V^*(([f-g]\mathcal{L}, [a,b]) = 0$.

Exercise 14.8 (Integrable functions are almost continuous). A necessary condition for a function f to be integrable on an interval $[a,b]$ is that f is almost continuous.

Hint: Such a function is almost everywhere the derivative of a continuous function F. In particular $f(x)$ is the limit almost everywhere of the sequence of continuous functions $n[F(x+1/n) - F(x)]$. It follows from Egorov's theorem that f is almost continuous.

Exercise 14.9 (Integration by parts). Let F, G, f, $g : [a,b] \to \mathbb{R}$ with f and g integrable on a compact interval $[a,b]$, $F(t) = \int_a^t f(x)\,dx$, and $G(t) = \int_a^t g(x)\,dx$. Then $Fg + fG$ is integrable on $[a,b]$ and

$$\int_a^b [F(x)g(x) + f(x)G(x)]\,dx = F(b)G(b).$$

Hint: Write $M = \sup_{x\in[a,b]} \{(b-a) + |F(x)| + |G(x)|\}$. Define

$$\beta_1 = \{(I,x) : |f(x)\omega G(I)| < \epsilon/8M\},$$
$$\beta_1 = \{(I,x) : |g(x)\omega F(I)| < \epsilon/8M\},$$

and choose Cousin covers β_3 and β_4 of $[a,b]$ so that

$$V(\Delta F - f\mathcal{L}, \beta_3) + V(\Delta G - g\mathcal{L}, \beta_4) < \epsilon/2M.$$

Check that $\beta = \beta_1 \cap \beta_2 \cap \beta_3 \cap \beta_4$ is a Cousin cover of $[a,b]$. Take any partition

$$\pi = \{[x_{i-1}, x_i], t_i) : i = 1, 2, \ldots, n\}$$

of $[a,b]$ contained in β and verify that

$$\left| F(b)G(b) - \sum_{i=1}^{n} \left(F(t_i)g(t_i) + f(t_i)G(t_i)\right)(x_i - x_{i-1}) \right| < \epsilon.$$

The identity

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

might help in this task, as will a bit of patience.

14.2. Properties of the indefinite integral

What conditions on a function F will ensure that it is the indefinite integral of an integrable function on an interval $[a, b]$? To simplify we can assume throughout that $F,\ f : \mathbb{R} \to \mathbb{R}$, that $[a, b]$ is a compact interval and we require

$$F(x) = F(a) \text{ for } x < a \text{ and } F(x) = F(b) \text{ for } x > b. \tag{14.1}$$

THEOREM 14.10 (Necessary conditions). The following assertions are true whenever F is the indefinite integral of an integrable function f on a compact interval $[a, b]$:

(a) F is continuous,
(b) F is $\mathcal{L}$-absolutely continuous,
(c) F has σ-finite variation on $[a, b]$,
(d) $F'(x) = f(x)$ for almost every $x \in [a, b]$,
(e) F has the Vitali property on $[a, b]$, and
(f) for every set $E \subset [a, b]$,

$$\mathcal{V}_F{}^*(E) = \mathcal{V}_{F*}(E) = \int_E |f(x)|\, dx.$$

PROOF. All of these follow from the identity $V^*(\Delta F - f\mathcal{L}, [a, b]) = 0$.

14.2.1. Characterization by $\mathcal{L}$-absolute continuity.

The indefinite integral is exactly characterized by the notion of $\mathcal{L}$-absolute continuity.

> [Warning: This is not the same as absolute continuity in the sense of Vitali. Recall that $\mathcal{L}$-absolute continuity on a compact interval $[a, b]$ forces a function to have σ-finite variation only. Absolute continuity in the sense of Vitali on a compact interval, however, requires that the function have bounded Jordan variation there.]

THEOREM 14.11 (Necessary and sufficient condition). In order for F to be the indefinite integral of an integrable function on a compact interval $[a, b]$ it is necessary and sufficient that F be $\mathcal{L}$-absolutely continuous on $[a, b]$.

PROOF. We know the necessity. If F is $\mathcal{L}$-absolutely continuous on $[a, b]$ then it has σ-finite variation on $[a, b]$. From that we deduce that $F'(x)$ exists almost everywhere. Thus F is an indefinite integral for F' (or more precisely to any function f that is almost everywhere equal to F').

14.2.2. Integrability of derivatives.

Let $F,\ f : \mathbb{R} \to \mathbb{R}$. The basic condition asserting the integrability of the derivative is the following:

> *If $F'(x) = f(x)$ at every point of a compact interval $[a, b]$ excepting a point N and $\mathcal{L}(N) = \mathcal{V}_F{}^*(N) = 0$ then F is an indefinite integral of f on $[a, b]$.*

For Dini derivatives there is a weaker integration statement.

THEOREM 14.12. If $F : \mathbb{R} \to \mathbb{R}$ is continuous, $f(x) = \overline{D}^+F(x)$ is everywhere finite on a compact interval $[a, b]$ and integrable on $[a, b]$, then

$$F(b) - F(a) = \int_a^b \overline{D}^+F(x)\, dx.$$

This has been proved earlier in Section 10.9. Note that the integrability of $\overline{D}^+F(x)$ cannot be deduced; here it must be assumed. As an exercise the student may wish to check that the finiteness of the Dini derivate can be relaxed to allow a countable set of points x where $\overline{D}^+F(x) = \pm\infty$.

14.3. McShane's criterion and the absolute integral

A function $f : [a,b] \to \mathbb{R}$ has been defined (Definition 9.5) to be absolutely integrable on $[a,b]$ if it satisfies McShane's criterion there. We say that f is nonabsolutely integrable on $[a,b]$ if it is integrable but fails to satisfy McShane's criterion. Recall the criterion:

> **[McShane's criterion]** A function $f : [a,b] \to \mathbb{R}$ is *absolutely integrable* on $[a,b]$ if for every $\epsilon > 0$ a Cousin cover β of $[a,b]$ can be found so that
> $$\sum_{(I,z\in\pi} \sum_{(I',z')\in\pi'} |f(z) - f(z')|\,\mathcal{L}(I\cap I') < \epsilon$$
> for all partitions π, π' of $[a,b]$ contained in β.

We now show that a function f is absolutely integrable if and only if both f and $|f|$ are integrable. Thus f is nonabsolutely integrable if f is integrable but $|f|$ is not. We recall that the language of series required $\sum_{k=1}^{\infty} a_k$ to be absolutely convergent if $\sum_{k=1}^{\infty} |a_k|$ is convergent. The series $\sum_{k=1}^{\infty} a_k$ is nonabsolutely convergent if it converges but $\sum_{k=1}^{\infty} |a_k| = \infty$. Thus our language for absolute integrability is very close to the same language that is usually used for series.

14.3.1. Characterization of the absolute integral. The class of absolutely integrable functions can be characterized in a number of ways.

THEOREM 14.13. A function f is absolutely integrable on a compact interval $[a,b]$ if and only any one of the following five equivalent conditions is satisfied:

(a) f satisfies McShane's criterion on $[a,b]$.
(b) both f and $|f|$ are integrable on $[a,b]$.
(c) f is integrable on $[a,b]$ and has an indefinite integral that is of bounded variation on $[a,b]$.
(d) f is the derivative almost everywhere in $[a,b]$ of a function that has bounded variation there.
(e) f is almost continuous and
$$\int_{[a,b]} |f(x)|\,dx < \infty.$$

Proof. The first statement (a) is the definition. We have elsewhere proved that (b) follows from (a). Assume (b) and let F be an indefinite integral for f. We can assume that F is defined everywhere by agreeing that $F(x) = F(a)$ for $x < a$ and $F(x) = F(b)$ for $x > b$. We know, then, that
$$V^*(\Delta F, [a,b]) = \int_{[a,b]} |f(x)|\,dx = \int_a^b |f(x)|\,dx < \infty.$$
In particular F must have bounded variation on $[a,b]$. Thus (c) holds. If (c) holds then certainly (d) holds.

Assume that (d) holds, i.e., that $G'(x) = f(x)$ everywhere in $[a, b]$ except for a null set N for some function G of bounded variation. Verify that any such function must be almost continuous (see Exercise 14.8). Then observe that

$$V^*(f\mathcal{L}, [a, b]) = V^*(f\mathcal{L}, [a, b] \setminus N) = V^*(\Delta G, [a, b] \setminus N) \leq V^*(\Delta G, [a, b]) < \infty.$$

This proves (e).

Finally, for the last step, suppose that (e) holds. We shall verify directly that f must satisfy McShane's criterion, verifying that (e) implies (a).

Let $\epsilon > 0$. If f is bounded (say $|f(x)| < M$ for all x) then choose $\delta = \epsilon/(4M)$ and observe that for any open set G with $\mathcal{L}(G) < \delta$ we must have

$$V^*(f\mathcal{L}, G \cap [a, b]) < \epsilon/4.$$

If f is not bounded we must turn to Theorem 8.14 which gives the same property provided that f is almost continuous; thus in either case we have a positive number δ with this property.

By the almost continuity of f we may select an open set G with length smaller than δ and a continuous function $g : \mathbb{R} \to \mathbb{R}$ so that $f(x) = g(x)$ for all $x \in [a, b] \setminus G$. Choose a full cover β_1 of G so that

$$\beta_1 \subset \{(I, x) : x \in I \text{ and } I \subset G\}$$

in such a way that $V(f\mathcal{L}, \beta_1) < \epsilon/4$. Define

$$\beta_2 = \{(I, x) : x \in I \setminus G \text{ and } \omega g(I) < \epsilon/(4[b - a])\}.$$

Finally set $\beta = \beta_1 \cup \beta_2$, which we note is a Cousin cover of $[a, b]$.

Now we need only verify, for this choice of β, that the condition expressed by the McShane criterion holds. Let π, π' denote arbitrary partitions of $[a, b]$ contained in β. Observe that, when $(I, z) \in \pi$ and $(I', z') \in \pi'$ and $\mathcal{L}(I \cap I') > 0$, either $I \cap I' \subset G$ or else both z and z' are in $[a, b] \setminus G$. In the latter case

$$|f(z) - f(z')| = |g(z) - g(z')| < \epsilon/(2[b - a]).$$

From this one computes that

$$\sum_{(I,z)\in\pi} \sum_{(I',z')\in\pi'} |f(z) - f(z')| \, \mathcal{L}(I \cap I') < \epsilon$$

for all partitions π, π' of $[a, b]$ contained in β. This is exactly McShane's condition and so proves (a). This completes the proof since the chain of implications completes the circle. ■

14.3.2. Characterization by absolute continuity. In Section 14.2.1 we saw that the indefinite integral is exactly characterized by the notion of $\mathcal{L}$-absolute continuity. For the absolute integral more must be added: the variation must be finite. We recall that a function F is absolutely continuous in the sense of Vitali on a compact interval $[a, b]$ precisely when it is $\mathcal{L}$-absolute continuous and has finite variation. Thus Vitali's concept exactly captures the absolute integral.

THEOREM 14.14 (Necessary and sufficient condition). *In order for a function F to be the indefinite integral of an absolutely integrable function on a compact interval $[a, b]$ it is necessary and sufficient that F is $\mathcal{L}$-absolutely continuous and of bounded variation on $[a, b]$.*

PROOF. We already know that a necessary and sufficient condition for F to be the indefinite integral of an integrable function f on $[a, b]$ is that F is $\mathcal{L}$-absolutely continuous. In general then, since

$$V^*(\Delta F, [a, b]) = V^*(f\mathcal{L}, [a, b],$$

this is finite exactly in the situation in which F has bounded variation on $[a, b]$.

COROLLARY 14.15 (Necessary and sufficient condition). In order for a function F to be the indefinite integral of an absolutely integrable function on a compact interval $[a, b]$ it is necessary and sufficient that F is absolutely continuous in the sense of Vitali on $[a, b]$.

14.4. Local absolute integrability

While an integrable function need not be absolutely integrable there is a close relation between the two concepts, In the statement f_E is the function $f_E(x) = f(x)$ for $x \in E$ and $f_E(x) = 0$ otherwise. This theorem can be generalized by replacing the condition that f is integrable with the much weaker assumption that f is almost continuous and has finite upper and lower integrals. See Theorem 14.32.

THEOREM 14.16. Suppose that a function f is integrable on a compact interval $[a, b]$. Then every nonempty closed subset E of $[a, b]$ has a portion $E \cap (c, d)$ for which f_E is absolutely integrable on $[c, d]$.

Proof. This follows from Theorem 13.34. The indefinite integral F of f on $[a, b]$ has σ-finite variation on E and so, consequently, must have finite variation on some portion $E \cap (c, d)$ of E. But that means that f_E is an almost continuous function for which

$$\int_{[c,d]} f_E(x)\, dx = \mathcal{V}_F{}^*(E \cap [c, d]) < \infty.$$

Such a function is absolutely integrable on $[c, d]$. ■

14.5. Expression of the integral as a measure

If a function $f : [a, b] \to \mathbb{R}$ is absolutely integrable then the integral of f and of $|f|$ can be realized by measure assertions:

$$\int_a^b f(x)\, dx = \int_{[a,b]} f^+(x)\, dx - \int_{[a,b]} f^-(x)\, dx,$$

and

$$\int_a^b |f(x)|\, dx = \int_{[a,b]} f^+(x)\, dx + \int_{[a,b]} f^-(x)\, dx.$$

What is the situation for nonabsolutely integrable functions?

THEOREM 14.17. If f is nonabsolutely integrable on a compact interval $[a, b]$ then

$$\int_{[a,b]} f^+(x)\, dx = \int_{[a,b]} f^-(x)\, dx = \infty.$$

PROOF. As f is integrable, f must be almost continuous. It follows that both f^+ and f^- are also almost continuous so that the integrability of either f^+ and f^- can be determined by whether these measures are finite. If one of these is finite, say the first one, then f^+ must be integrable. But that would mean both f and f^+ are integrable; hence so too is $f^- = f - f^+$. It would follow that $|f| = f^+ + f^-$ is integrable and so f is absolutely integrable.

14.6. Riemann's criterion

In Section 9.7 we introduced the integrability criterion of Riemann, defining a narrow class of integrable functions. With some help from the measure theory we can give Lebesgue's characterization of this class.

THEOREM 14.18. Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then the following assertions are equivalent:

(a) For every $\epsilon > 0$ there is a partition π of $[a, b]$ for which
$$\sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I) < \epsilon.$$

(b)
$$\int_a^b \omega f \cdot \mathcal{L} = 0.$$

(c) f is continuous at almost every point of $[a, b]$.

Proof. Using the methods in Section 9.7 for proving the elementary Theorem 9.7.1, it is relatively straightforward to check that (a) and (b) are equivalent. Let us assume (a) and prove (c). Let
$$E_n = \{x \in [a, b] : \omega_f(x) \geq 1/n\}.$$
This is is a closed set containing only points of discontinuity of f. Indeed $\bigcup_{n=1}^{\infty} E_n$ is exactly the set of all points of discontinuity of f (see Theorem 4.34).

Fix n and $\epsilon > 0$. Choose a partition π of $[a, b]$ so that
$$\sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I) < \epsilon.$$
Let π' denote the subset of π containing just the elements $([x, y], z) \in \pi$ for which $(x, y) \cap E_n \neq \emptyset$. Note that each pair $([x, y], z)$ that belongs to π' will necessarily require
$$\omega f([x, y]) \geq 1/n.$$
Also the collection of such intervals $[x, y]$ will cover all but finitely many points of E_n. Hence we compute that
$$\mathcal{L}(E_n) \leq \sum_{(I,x)\in\pi'} \mathcal{L}(I) \leq \sum_{(I,x)\in\pi'} n\omega f(I)\mathcal{L}(I) < n\epsilon.$$
This can happen for all ϵ only if $\mathcal{L}(E_n) = 0$ from which it follows that $\bigcup_{n=1}^{\infty} E_n$ (the set of points of discontinuity of f) is also of $\mathcal{L}$-measure zero. This proves (c).

Finally let us assume (c) and show that this implies (b). Let $\eta > 0$ and choose M so that $|f(x)| < M$ for all x in $[a, b]$. Let
$$E_\eta = \{x \in [a, b] : \omega_f(x) \geq \eta\}.$$

This is a set of measure zero. There must be an open set G containing E_η so that $\mathcal{L}(G) < \eta/(2M)$. For each point x in $[a,b]$ that is not in G note that $\omega_f(x) < \eta$. Construct the covering relation

$$\beta_1 = \{(I,x) : x \in [a,b] \setminus G,\ x \in I, \text{ and } \omega f(I) < \eta\}.$$

This is a full cover of $[a,b] \setminus G$. Construct the covering relation

$$\beta_2 = \{(I,x) : x \in G,\ x \in I, \text{ and } I \subset G\}.$$

Observe that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a,b]$.

Let π be any partition of $[a,b]$ contained in β. Write $\pi' = \pi[G]$ and $\pi'' = \pi \setminus \pi[G]$. Then

$$\sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I) = \sum_{(I,x)\in\pi'} \omega f(I)\mathcal{L}(I) + \sum_{(I,x)\in\pi''} \omega f(I)\mathcal{L}(I)$$

$$\leq 2M\mathcal{L}(G) + \eta(b-a) < \eta(1+b-a).$$

Since η is an arbitrary positive number it follows that

$$\int_a^b \omega f \cdot \mathcal{L} = 0.$$

so that we have deduced (b) from (c). ■

14.6.1. Elementary computation of the integral. For the purposes of the next theorem we shall write

$$\|\pi\| = \sup\{\mathcal{L}(I) : (I,x) \in \pi\}$$

and refer to this notion as the *norm* of the partition π. We have an improved version of Theorem 9.7.1 that illustrates that, for this special class of functions, the integral can be computed in a simpler manner. Generally for absolutely integrable functions the only effective method we have is a lengthy measure-theoretic computation. For the small class of functions satisfying Riemann's criterion the value of the integral can be obtained in a simple sequence of steps.

Theorem 14.19. Let f be a function that satisfies Riemann's criterion on an interval $[a,b]$. Then f is absolutely integrable on $[a,b]$ and for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \int_a^b f(x)\,dx \right| < \epsilon \tag{14.2}$$

for every partition π of the interval $[a,b]$ for which $\|\pi\| < \delta$.

Proof. We have already shown in Theorem 9.7.1 that f would have to be absolutely integrable. Argue first that, if f satisfies the Riemann criterion, it is possible to select a $\delta > 0$ so that, whenever $\|\pi\| < \delta$, then

$$\sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I) < \epsilon/2. \tag{14.3}$$

For the remainder of the proof fix a partition π of the interval $[a,b]$ for which $\|\pi\| < \delta$.

Select a Cousin cover β of $[a,b]$ in such a way that

$$\left| \sum_{(I',x')\in\pi'} f(x')\mathcal{L}(I') - \int_a^b f(x)\,dx \right| < \epsilon/2 \tag{14.4}$$

whenever π' is a partition of $[a,b]$ from β. In particular we can select a partition π' from β (so that this inequality must hold) and also so that for any choice of $(I,x)\in\pi$ (the original partition fixed above), π' contains a partition of I. In that case we have

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \sum_{(I',x')\in\pi'} f(x')\mathcal{L}(I') \right| < \sum_{(I,x)\in\pi} \omega f(I)\mathcal{L}(I) < \epsilon/2. \tag{14.5}$$

Combining (14.4) and (14.5) gives (14.2) as required. ■

14.7. Freiling's criterion

We now give a criterion that is similar in a number of ways to Riemann's criterion, the second Cauchy criterion, and McShane's criterion. This expresses the condition that a function should be a derivative by means of a covering relation.

Lemma 14.20. Let F, $f:\mathbb{R}\to\mathbb{R}$. A necessary and sufficient condition in order that f be the derivative of F at each point is that for every $\epsilon>0$ there is a full cover β of the real line with the property that for every compact interval $[a,b]$ and every partition $\pi\subset\beta$ of $[a,b]$,

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon\mathcal{L}([a,b]). \tag{14.6}$$

Proof. Let $\epsilon>0$ and suppose that $F'(x)=f(x)$ at every point. Define

$$\beta = \{(I,x) : |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon\mathcal{L}(I)\}.$$

Check that β is a full cover of $\mathbb{R}$ and that it satisfies (14.6).

Conversely suppose that β has been chosen to be a full cover of $\mathbb{R}$ that satisfies (14.6). Fix $x\in\mathbb{R}$ and determine $\delta>0$ so that whenever (I,x) satisfies $x\in I$ and $\mathcal{L}(I)<\delta$ then necessarily $(I,x)\in\beta$. Note that if $([c,d],x)$ is any pair for which $c\le x\le d$ and $d-c<\delta$ then necessarily $([c,d],x)$ is in β and the set π containing only this one pair is itself also a partition of $[c,d]$. Consequently, using (14.6),

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

But this verifies that $F'(x)=f(x)$. ■

Theorem 14.21 (Freiling's criterion). Let $f:\mathbb{R}\to\mathbb{R}$. A necessary and sufficient condition in order that f be the derivative of some function F at each point is that for every $\epsilon>0$ there is a full cover β of the real line with the property that for every compact interval $[a,b]$ and every pair of partitions π_1, $\pi_2\subset\beta$ of $[a,b]$,

$$\left| \sum_{(I,z)\in\pi} \sum_{(I',z')\in\pi'} [f(z)-f(z')]\mathcal{L}(I\cap I') \right| < \epsilon\mathcal{L}([a,b]). \tag{14.7}$$

Proof. Suppose that $F'(x) = f(x)$ everywhere. Apply Lemma 14.20 to find a full cover β for which for every compact interval $[a, b]$ and every partition $\pi \subset \beta$ of $[a, b]$,

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon\mathcal{L}([a, b])/2. \tag{14.8}$$

Let π_1, $\pi_2 \subset \beta$ be partitions of $[a, b]$. Apply (14.8) to each of them and add to obtain that

$$\left| \sum_{(I,z)\in\pi} f(z)\mathcal{L}(I) - \sum_{(I',z')\in\pi'} f(z')\mathcal{L}(I') \right| < \epsilon\mathcal{L}([a, b]). \tag{14.9}$$

Now simply rearrange (14.9) to obtain (14.7).

Conversely suppose that the statement (14.7) in the theorem holds for ϵ and β. This is a stronger statement than the Cauchy second criterion and so f is integrable on every compact interval. Thus there is a function F that will serve as the indefinite integral for f on any interval. From (14.7) we deduce that

$$\left| \Delta F([a, b]) - \sum_{(I,z)\in\pi} f(z)\mathcal{L}(I) \right| < 2\epsilon\mathcal{L}([a, b]) \tag{14.10}$$

must be true for any partition π of $[a, b]$ from the cover β.

Fix $x \in \mathbb{R}$ and determine $\delta > 0$ so that whenever (I, x) satisfies $x \in I$ and $\mathcal{L}(I) < \delta$ then necessarily $(I, x) \in \beta$. Note that if $([c, d], x)$ is any pair for which $c \le x \le d$ and $d - c < \delta$ then necessarily $([c, d], x)$ is in β and the set π containing only this one pair is itself also a partition of $[c, d]$. Consequently, using (14.10),

$$|F(d) - F(c) - f(x)(d - c)| < 2\epsilon(d - c).$$

But this verifies that $F'(x) = f(x)$. ■

EXERCISE 14.22. Let $f : \mathbb{R} \to \mathbb{R}$. Characterize the following property: for every $\epsilon > 0$ there is a full cover β of the real line with the property that for every compact interval $[a, b]$ and every pair of partitions π_1, $\pi_2 \subset \beta$ of $[a, b]$,

$$\sum_{(I,z)\in\pi} \sum_{(I',z')\in\pi'} |f(z) - f(z')|\mathcal{L}(I \cap I') < \epsilon\mathcal{L}([a, b]).$$

14.8. Limits of integrable functions

If $\lim_{p\to\infty} f_p(x) = f(x)$ at almost every point x of an interval $[a, b]$ one would expect that, under some appropriate hypotheses, there would be a conclusion that

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

We already know from Corollary 6.20 that, if f is almost everywhere a sum of nonnegative integrable functions, i.e.,

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

for almost every x, then

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

Theorem 6.21 shows that, if f is the limit almost everywhere of a monotonic sequence of integrable functions, $\lim_{p\to\infty} f_p(x) = f(x)$, then

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

For most applications these will be enough. For theoretical reasons we now turn to uniformity conditions.

14.8.1. Equi-integrable sequences. As usual in investigations involving interchange of limit operations an assumption of uniformity proves useful.

DEFINITION 14.23. We say that $\{f_p\}$ are *equi-integrable* on $[a,b]$ if for all $\epsilon > 0$ a Cousin cover β of $[a,b]$ can be found so that

$$\left|\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'} [f_p(z) - f_p(z')]\mathcal{L}(I\cap I')\right| < \epsilon \tag{14.11}$$

for all partitions π, π' of $[a,b]$ contained in β.

EXERCISE 14.24. Let $\{f_p\}$ be a sequence of integrable functions converging uniformly to a function f on $[a,b]$. Then $\{f_p\}$ are equi-integrable on $[a,b]$.

THEOREM 14.25. Suppose that $\lim_{p\to\infty} f_p(x) = f(x)$ at almost every point of $[a,b]$ and that $\{f_p\}$ are equi-integrable on $[a,b]$. Then $\lim_{p\to\infty}\int_a^b f_p(x)\,dx$ exists, f is integrable and

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

Proof. Let N be the set of points in $[a,b]$ at which $\lim_{p\to\infty} f_p(x) = f(x)$ may fail. This is a null set and we can agree that $f_p(x) = f(x) = 0$ for every p and every $x\in N$. This changes none of the assertions but makes the proof transparent.

In equation (14.11) simply let $p\to\infty$. Then we obtain for that β that

$$\left|\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'} [f(z) - f(z')]\mathcal{L}(I\cap I')\right| \le \epsilon \tag{14.12}$$

for all partitions π, π' of $[a,b]$ contained in β. From this it follows that f is integrable.

From (14.11) again, deduce by our usual theory, that

$$\left|\sum_{(I,z)\in\pi} f_p(z)\mathcal{L}(I) - \int_a^b f_p(x)\,dx\right| < 2\epsilon$$

for all partitions π of $[a,b]$ contained in β. Once again let $p\to\infty$. ∎

14.8.2. Equi-absolutely integrable sequences. An identical theory can be used for absolutely integrable functions.

DEFINITION 14.26. We say that $\{f_p\}$ are *equi-absolutely integrable* if for all $\epsilon > 0$ a Cousin cover β of $[a,b]$ can be found so that for all p,

$$\sum_{(I,x)\in\pi}\sum_{(I',x')\in\pi'} |f_p(x) - f_p(x')|\mathcal{L}(I\cap I') < \epsilon$$

for all partitions π, π' of $[a,b]$ contained in β.

EXERCISE 14.27. Let $\{f_p\}$ be a sequence of absolutely integrable functions converging uniformly to a function f on $[a,b]$. Then $\{f_p\}$ are equi-absolutely integrable on $[a,b]$.

THEOREM 14.28. Suppose that $\lim_{p\to\infty} f_p(x) = f(x)$ at almost every point of $[a,b]$, and that $\{f_p\}$ are equi-absolutely integrable on $[a,b]$. Then $\lim_p \int_a^b f_p(x)\,dx$ exists, f is absolutely integrable and

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

14.8.3. Convergence criteria. The simplest condition that imposes equi-absolutely integrability on a sequence is uniform convergence. By using the techniques of Section 7.11.1 we can relax this to allow the removal of a set of small measure and obtain, in this way, a number of useful characterizations of the concept. The proofs are left to the student, using the usual properties of the variation.

THEOREM 14.29. Suppose that $\lim_{p\to\infty} f_p(x) = f(x)$ at almost every point of $[a,b]$, and that each f_p is absolutely integrable on $[a,b]$ with indefinite integral F_p. Then the following are equivalent:

(a) $\{f_p\}$ are equi-absolutely integrable on $[a,b]$,
(b) $\{|f_p|\}$, $\{[f_p]^+\}$ and $\{[f_p]^-\}$ are equi-absolutely integrable on $[a,b]$,
(c)
$$\lim_{p\to\infty}\overline{\int_a^b}|f_p(x) - f(x)|\,dx = 0.$$
(d)
$$\lim_{p\to\infty}\sup_{k\geq p}\overline{\int_a^b}|f_p(x) - f_k(x)|\,dx = 0.$$
(e)
$$\lim_{p\to\infty}\sup_{k\geq p}\overline{\int_a^b}|\Delta F_p - \Delta F_k| = 0.$$
(f) For all $\epsilon > 0$ there is a $\delta > 0$ so that, if $G\subset[a,b]$ is open, and $\mathcal{L}(G) < \delta$ then
$$\int_G |f_p(x)|\,dx < \epsilon$$
for all p.
(g) For all $\epsilon > 0$ there is a $\delta > 0$ so that, if $G\subset[a,b]$ is open, and $\mathcal{L}(G) < \delta$ then $V^*(\Delta F_p, G) < \epsilon$ for all p.

Should any of these hold then

$$\lim_{p\to\infty}\int_a^b |f_p(x)-f(x)|\,dx = 0,$$

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

and

$$\lim_{p\to\infty}\int_a^b |f_p(x)|\,dx = \int_a^b |f(x)|\,dx.$$

As corollaries the following offer sufficient conditions.

Exercise 14.30 (Dominated convergence). Suppose that $\lim_{p\to\infty} f_p(x) = f(x)$ at almost every point of $[a,b]$ and that each f_p is absolutely integrable on $[a,b]$. Suppose that g is absolutely integrable on $[a,b]$ and that $|f_p(x)| \leq |g(x)|$ for all x and p. Then $\{f_p\}$ are equi-absolutely integrable on $[a,b]$.

Exercise 14.31 (Monotone convergence). Suppose that

$$0 \leq f_1(x) \leq f_2(x) \leq f_3(x) \leq \cdots \leq f_p(x) \to f(x)$$

at each point of $[a,b]$ and that each f_p is absolutely integrable on $[a,b]$. Then provided $\int_{[a,b]} |f(x)|\,dx < \infty$, the sequence $\{f_p\}$ is equi-absolutely integrable on $[a,b]$.

14.9. Local absolute integrability conditions

An almost continuous function f is absolutely integrable on an interval $[a,b]$ provided that the variation $\overline{\int_a^b}|f(x)|\,dx$ is finite. If the variation is not finite then f cannot be absolutely integrable on $[a,b]$. But need it be absolutely integrable on some subinterval? The theorem we now prove gives a sufficient condition in order for an almost continuous functions to have a local integrability property. In the theorem we use the following notation for a function f and a closed set E: the function f_E is defined as $f_E(x) = f(x)$ whenever $x \in E$ and $f_E(x) = 0$ otherwise.

Theorem 14.32. Let E be a nonempty closed subset of $[a,b]$ and f an almost continuous function. Suppose that

$$-\infty < \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx < \infty.$$

Then E contains a portion $E \cap (c,d)$ so that f_E is absolutely integrable on $[c,d]$.

Proof. We make a simplifying assumption that allows a small technical detail later. We remove from the set E all points that are isolated on either the right side or the left side or both sides. There are only countably many such points and that does not influence either measure or integration statements. While the resulting set is not closed, it is a set of type $\mathcal{G}_\delta$ so that we may still apply the Baire-Osgood theorem to it.

Choose t so that

$$-t < \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx < t$$

and a Cousin cover β of $[a,b]$ so that[1]

$$\left|\sum_{\pi} f\mathcal{L}\right| < t$$

for all partitions $\pi \subset \beta$ of $[a,b]$. Let $[c,d]$ be any subinterval and let $\pi \subset \beta$ be a partition of $[c,d]$. Choose $\pi' \subset \beta$ so that it consists of a partition of $[a,c]$ and $[d,b]$. Then

$$\left|\sum_{\pi\cup\pi'} f\mathcal{L}\right| < t$$

so that

$$\left|\sum_{\pi} f\mathcal{L}\right| \leq t + \left|\sum_{\pi'} f\mathcal{L}\right|$$

In particular we can write

$$T(c,d) = \sup\left\{\left|\sum_{\pi} f\mathcal{L}\right| : \ \pi \subset \beta \text{ is a partition of } [c,d]\right\} < \infty.$$

We need a decomposition argument for β similar to that of the decomposition Lemma 11.11. Choose $\delta(x) > 0$ so that $x \in I \subset [a,b]$ and $\mathcal{L}(I) < 2\delta(x)$ requires $(I,x) \in \beta$. Define

$$E_n^+ = \{x \in E : \delta(x) > 1/n,\ 0 \leq f(x) \leq n\}$$

and

$$E_n^- = \{x \in E : \delta(x) > 1/n,\ 0 \geq f(x) \geq -n\}.$$

This sequence of sets exhausts the set E so that, by the Baire-Osgood theorem, there must be a portion of E so that one of the sets is dense there. Thus we are able to choose an integer m and a subinterval $[c,d]$ so that $d-c < 1/m$ and so that E_m^+ is dense in the nonempty portion $E \cap (c,d)$.

We shall investigate the absolute integrability of f_E on $[c,d]$. For that, let π be an arbitrary partition of $[c,d]$ chosen from β. We shall estimate

$$\sum_{\pi} f_E^+\mathcal{L} \quad \text{and} \quad \sum_{\pi} f_E^-\mathcal{L}$$

(where, as usual, f_E^+ and f_E^- denote the positive and negative parts of f_E).

Define $\pi_1 = \pi[E]$ and $\pi_2 = \pi \setminus \pi_1$. We alter π_1 in two different ways. The first alteration denoted as π_1' will replace each $(I,x) \in \pi_1$ by (I,x') where $x' \in E_m^+$. Since $x \in E$ and is not isolated on either side in E, and since E_m^+ is dense in this portion of E, such points are available. For any such point x' we see that the pair $(I,x') \in \beta$ because $\mathcal{L}(I) < 1/m < \delta(x')$. The second alteration denoted as π_1'' will replace each $(I,x) \in \pi_1$ for which $f(x) < 0$ by (I,x'') where $x' \in E_m^+$. For the same reasons as before, the pair $(I,x'') \in \beta$. We will make use of the fact that, for the adjusted points x' and x'', we have the inequalities $0 \leq f(x') \leq m$ and $f(x) < 0 \leq f(x'')$.

[1] We are using $\sum_{\pi} f\mathcal{L}$ to denote the sum $\sum_{(I,x)\in\pi} f(x)\mathcal{L}(I)$ in this proof as many such sums will be considered.

Now we do our computations:

$$\left|\sum_{\pi_1\cup\pi_2} f\mathcal{L}\right| \leq T(c,d) \tag{14.13}$$

$$\left|\sum_{\pi_1'\cup\pi_2} f\mathcal{L}\right| \leq T(c,d) \tag{14.14}$$

$$\left|\sum_{\pi_1'} f\mathcal{L}\right| \leq m(d-c) \leq 1. \tag{14.15}$$

Combining (14.14) and (14.15) we see that

$$\left|\sum_{\pi_2} f\mathcal{L}\right| \leq T(c,d)+1 \tag{14.16}$$

Thus we can estimate

$$\sum_{\pi} f_E^+\mathcal{L} = \sum_{\pi_1} f_E^+\mathcal{L} = \sum_{\pi_1''} f_E^+\mathcal{L} \leq \sum_{\pi_1''} f\mathcal{L}$$

$$\leq \sum_{\pi_1''} f\mathcal{L} + \left[\sum_{\pi_2} f\mathcal{L} + T(c,d)+1\right] = \sum_{\pi_1''\cup\pi_2} f\mathcal{L} + T(c,d)+1 \leq 2T(c,d)+1.$$

As such sums have this upper bound we can conclude, from Theorem 14.13, that f_E^+ is absolutely integrable on $[c,d]$.

Now we show that f_E^- is also absolutely integrable on $[c,d]$. Since

$$f_E^-(x) = f_E^+(x) - f(x)$$

for every $x \in E$, we find that

$$\sum_{\pi} f_E^-\mathcal{L} = \sum_{\pi_1} f_E^-\mathcal{L} = \sum_{\pi_1} f_E^+\mathcal{L} - \sum_{\pi_1} f\mathcal{L}$$

$$= \left[\sum_{\pi_1} f_E^+\mathcal{L} + \sum_{\pi_2} f_E^+\mathcal{L}\right] - \sum_{\pi_1} f\mathcal{L}$$

$$\leq [2T(c,d)+1] - \sum_{\pi_1} f\mathcal{L} - \left[\sum_{\pi_2} f\mathcal{L} - T(c,d) - 1\right]$$

$$= [3T(c,d)+2] - \sum_{\pi} f\mathcal{L} \leq 4T(c,d)+2.$$

Once again such sums have this upper bound we can conclude that f_E^- is absolutely integrable on $[c,d]$. Finally then $f_E = f_E^+ + f_E^-$ too must be absolutely integrable on $[c,d]$. This gives us our portion $E \cap (c,d)$ and completes the proof. ■

14.10. Continuity of upper and lower integrals

The indefinite integral of an integrable function is continuous. We can express this by saying that, if f is integrable on a compact interval $[a, b]$, then for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$-\epsilon < \int_c^d f(x)\,dx < \epsilon$$

for every subinterval $[c, d] \subset [a, b]$ for which $\mathcal{L}([c, d]) < \delta$. We wish a version of this that does not assume integrability and that can be used for a characterization.

Definition 14.33. A function f is said to have *continuous upper and lower integrals* on a compact interval $[a, b]$ if for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$-\epsilon < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < \epsilon$$

for every subinterval $[c, d] \subset [a, b]$ for which $\mathcal{L}([c, d]) < \delta$.

Lemma 14.34. Suppose that $f : [a, b] \to \mathbb{R}$ has continuous upper and lower integrals on a compact interval $[a, b]$. Then

$$-\infty < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < \infty$$

for every subinterval $[c, d] \subset [a, b]$.

Proof. There must be a $\delta > 0$ so that

$$-1 < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < 1$$

for every subinterval $[c, d] \subset [a, b]$ for which $\mathcal{L}([c, d]) < \delta$. Subdivide

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b$$

in such a way that each $a_i - a_{i-1} < \delta$. Then compute, using Exercise 6.7, that

$$\overline{\int_a^b} f(x)\,dx = \sum_{i=1}^{n} \overline{\int_{a_{i-1}}^{a_i}} f(x)\,dx \leq n < \infty.$$

A similar argument handles the lower integral. ■

Exercise 14.35 (Cauchy extension property). Let f be integrable on every subinterval $[c, d] \subset (a, b)$. Show that f is integrable on $[a, b]$ if and only if if f has continuous upper and lower integrals on $[a, b]$.

Hint: Compare this with Exercise 14.5.

14.11. A characterization of the integral

We have characterized the absolute integral using the concept of almost continuity and a requirement of finite variation. Here we show that this continuity condition on the upper and lower integrals, together with almost continuity, characterizes the integral in general.

Theorem 14.36. A function f is integrable on $[a, b]$ if and only if f is almost continuous and f has continuous upper and lower integrals on $[a, b]$.

Proof. We already know that an integrable function has these properties. Conversely suppose that f is almost continuous and that f has continuous upper and lower integrals on $[a,b]$. An open interval $(s,t) \subset (a,b)$ will be called "accepted" if f is integrable on every $[c,d] \subset (s,t)$. Let G be the union of all accepted intervals. This is an open subset of (a,b). Note that, if $[c,d] \subset G$, then by the Heine-Borel property (Theorem 3.23) $[c,d]$ can be written as the union of a finite collection of intervals $\{[c_i,d_i]\}$ each of which is inside an accepted interval. It follows that f is integrable on $[c,d]$ too.

Let

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

displaying G as a union of its component intervals. We claim first that f must be integrable on each of the compact intervals $[a_i,b_i]$. This follows directly from the Cauchy extension property (Exercise 14.35) using the hypothesis that f has continuous upper and lower integrals. We shall use a single function F to represent the indefinite integral of f on each of these intervals, but we are cautioned not to use F outside of the intervals.

In particular if $G = (a,b)$ then the proof is completed since then f must be integrable on $[a,b]$ as required. Suppose not, i.e., that the theorem fails and $G \neq (a,b)$. Then $E = [a,b] \setminus G$ is a nonempty closed set. Note that E can have no isolated points. Indeed if $c \in E$ is isolated then $(c-t,c) \subset G$ and $(c,c+t) \subset G$ for some $t > 0$ and another application of the Cauchy extension property would show that $(c-t,c+t)$ is accepted so that $(c-t,c+t) \subset G$ which is not possible.

The goal of the proof now will be to obtain a portion $E \cap (c',d')$ of E with the property that (c',d') is accepted, which would be impossible. Portions cannot be empty and no point of E would be allowed to belong to an accepted interval. The local integrability Theorem 14.32 and the Harnack extension property (Exercise 14.6) will play key roles.

The assumption that f satisfies the continuity condition in Definition 14.33 together with Lemma 14.34 shows that the upper and lower integrals of f are finite. Thus, we can apply Theorem 14.32 to find a portion $E \cap [c,d]$ so that f_E is absolutely integrable on $[c,d]$.

Since f has continuous upper and lower integrals on $[c,d]$ it follows from Lemma 14.34 that

$$-\infty < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < \infty.$$

Since f_E is absolutely integrable on $[c,d]$ it follows that

$$\int_c^d |f_E(x)|\,dx < \infty.$$

Thus we can select a real number $M > 0$ and a Cousin cover β of $[c,d]$ so that for any partition π of $[c,d]$ from β both

$$\left|\sum_{\pi} f\mathcal{L}\right| < M$$

and

$$\sum_{\pi} |f_E|\mathcal{L} < M.$$

We need a decomposition argument for β similar to that of the decomposition Lemma 11.11. Choose $\delta(x) > 0$ so that $x \in I \subset [a,b]$ and $\mathcal{L}(I) < 2\delta(x)$ requires $(I,x) \in \beta$. Define

$$E_n = \{x \in E \cap [c,d] : \delta(x) > 1/n\}.$$

This sequence of sets exhausts the set $E \cap [c,d]$ so that, by the Baire-Osgood theorem, there must be a portion of so that one of the sets is dense there. Let us agree that E_m is dense in $E \cap (c',d')$ and that $[c',d']$ is smaller in length than $1/m$. Let $\{(c_i,d_i)\}$ denote the component intervals of $(c',d') \setminus E$. There must be infinitely many such component intervals since otherwise it would follow that f is integrable on $[c',d']$. We claim that

$$\sum_{i=1}^{\infty} \omega F([c_i,d_i] = \infty. \tag{14.17}$$

For, if not, then the Harnack extension property (Exercise 14.6) shows that $f - f_E$ must be integrable on $[c',d']$ and hence f is integrable there. But that contradicts the fact that $[c',d']$ must contain points of E.

From the continuity of F we know that

$$\omega F([c_i,d_i] = |F(s) - F(t)| \tag{14.18}$$

for some subinterval $[s,t] \subset [c_i,d_i]$. Consequently we may choose a sequence of intervals $\{[s_k,t_k]\}$, chosen from different component intervals $[c_i,d_i]$ in such a way that either

$$0 \le \sum_{k=1}^{\infty} F(t_k) - F(s_k) = \infty \tag{14.19}$$

or

$$0 \ge \sum_{k=1}^{\infty} F(t_k) - F(s_k) = -\infty. \tag{14.20}$$

Let us assume the former. If (14.20) holds instead the same argument with a slight adjustment in the inequalities will work.

Now we fix an integer p and carefully construct a partition π of the interval $[c,d]$ from β. The first step is to choose π' from β to form a partition of $[c,c']$, then π'' from β to form a partition of $[d',d]$. For each of the intervals $\{[s_k,t_k]\}$ for $k = 1,2,3\dots,p$ we select a partition π_k of $[s_k,t_k]$ in such a way that

$$|F(t_k) - F(s_k) - \sum_{\pi_k} f\mathcal{L}| < 2^{-k}. \tag{14.21}$$

This is possible since f is integrable on each such interval and F is an indefinite integral. To complete the partition we take the remaining intervals, not yet covered by

$$\pi' \cup \pi'' \cup \bigcup_{k=1}^{p} \pi_k.$$

There are only finitely many of these intervals, say $I_1, I_2, \dots, I_q$. Each is a subinterval of $[c',d']$ and each one contains many points of E; thus each one also contains a point of E_m. Select a point x_i from $E_m \cap I_i$ $(i = 1, 2, \ldots, q)$ and note that (I_i, x_i) belongs to β. Thus if we set

$$\pi''' = \{(I_i, x_i) : i = 1, 2, \ldots, q\}$$

then we have obtained a partition

$$\pi = \pi' \cup \pi'' \cup \pi''' \cup \bigcup_{k=1}^{p} \pi_k$$

of the interval $[c, d]$ that is contained in β.

Consequently, by the way in which we chose M and β,

$$\left| \sum_{\pi} f\mathcal{L} \right| \leq M.$$

We know too that

$$\left| \sum_{\pi'''} f\mathcal{L} \right| \leq \sum_{\pi'''} |f_E|\mathcal{L} \leq M.$$

We combine these inequalities with (14.21) and the simple inequality

$$\sum_{k=1}^{p} 2^{-k} \leq 1$$

to obtain

$$\sum_{k=1}^{p} F(t_k) - F(s_k) \leq \left| \sum_{\pi' \cup \pi''} f\mathcal{L} \right| + 2M + 1.$$

This is true for any p and conflicts with our assumption that the inequality (14.19) holds.

Since neither inequality (14.19) nor (14.20) can hold it follows that inequality (14.18) also fails, thus f is integrable on $[c', d']$. In other words (c', d') is accepted, which would be impossible. This completes the proof. ■

CHAPTER 15

The Stieltjes Integral

The integral
$$\int_a^b f(x)\,dG(x)$$
denotes an integral obtained as a limit of Cauchy sums of the form
$$\sum_{(I,x)\in\pi} f(x)\Delta G(I).$$
This is the integral that might be expressed by the notation
$$\int_a^b f\Delta G.$$
Such an integral is called a Stieltjes integral.

EXERCISE 15.1. Let $G : [a, b] \to \mathbb{R}$ be any function. Show that
$$\int_a^b dG(x) = G(b) - G(a).$$

EXERCISE 15.2. Let $G : \mathbb{R} \to \mathbb{R}$ be defined by $G(x) = r$ for $x < 0$, $G(0) = s$, and $G(x) = t$ for $x > 0$. Determine
$$\int_a^b f(x)\,dG(x).$$

15.1. Reduction theorem

The Stieltjes integral reduces to an ordinary integral in a number of interpretations. When the integrating function G is an indefinite integral the whole theory reduces to ordinary integration.

THEOREM 15.3. Suppose that G, f, $g : \mathbb{R} \to \mathbb{R}$ and that g is integrable on a compact interval $[a, b]$ with an indefinite integral
$$G(d) - G(c) = \int_c^d g(x)\,dx \quad (a \le c < d \le b).$$
Then the Stieltjes integral
$$\int_a^b f(x)\,dG(x)$$
exists if and only if fg is integrable on $[a, b]$, in which case
$$\int_a^b f(x)\,dG(x) = \int_a^b f(x)g(x)\,dx.$$

Proof. To simplify, assume that $F(x) = F(a)$ and $G(x) = G(a)$ for all $x < a$ while $F(x) = F(b)$ and $G(x) = G(b)$ for all $x > b$. The assumption in the statement of the theorem on G and g requires that

$$\int_a^b |\Delta G - g\mathcal{L}| = V^*(\Delta G - g\mathcal{L}, [a,b]) = 0.$$

From that it is trivial to deduce (using our measure methods) that

$$V^*(f\Delta G - fg\mathcal{L}, [a,b]) = 0$$

or, equivalently, that

$$\int_a^b |f\Delta G - fg\mathcal{L}| = 0.$$

Then the requirement that H be an indefinite integral of fg would read as

$$V^*(\Delta H - fg\mathcal{L}, [a,b]) = 0.$$

But the requirement that H be a Stieltjes integral for $f\Delta G$ would be

$$V^*(\Delta H - f\Delta G, [a,b]) = 0.$$

These two identities are equivalent because of the equivalence relation

$$V^*(f\Delta G - fg\mathcal{L}, [a,b]) = 0.$$

■

15.2. Variational properties

Let us assume throughout this section that f, F, $G : \mathbb{R} \to \mathbb{R}$, $[a,b]$ is a compact interval and that

$$F(x) = F(a), \quad G(x) = G(a)$$

for all $x < a$ while

$$F(x) = F(b), \quad G(x) = G(b)$$

for all $x > b$. We require that the Stieltjes integral

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

exists and that F is an indefinite integral, i.e., that

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

Then we know from the basic theory that

$$V^*(\Delta F - f\Delta G, [a,b]) = 0$$

exactly characterizes this situation. From that we can immediately deduce the following facts by applying our general methods:

LEMMA 15.4. The indefinite integral F is continuous at every point at which G is continuous.

LEMMA 15.5. The indefinite integral F has zero variation on any subset E of $[a,b]$ on which G has zero variation.

LEMMA 15.6. The indefinite integral F has σ-finite variation on any subset E of $[a,b]$ on which G has finite variation.

LEMMA 15.7. If f is bounded on $[a,b]$ then the indefinite integral F has finite variation on any subset E of $[a,b]$ on which G has finite variation.

LEMMA 15.8. If G is a saltus function on $[a,b]$ then so too is the indefinite integral F and

$$\mathcal{V}_F{}^*([a,b]) = \sum_{x\in[a,b]} |f(x)|\mathcal{V}_G{}^*(\{x\}).$$

15.3. Derivative of the integral

We continue the assumptions on F, G and f given in the preceding section and examine, from the identity

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

the relation between the derivatives of F and G.

LEMMA 15.9. For almost every point x in $[a,b]$

$$\lim_{(I,x)\Rightarrow x} \frac{\Delta F(I) - f(x)\Delta G(I)}{\mathcal{L}(I)} = 0.$$

In particular $F'(x) = f(x)G'(x)$ at almost every point x at which either F or G is differentiable.

Proof. This follows from the fundamental theorem of the calculus (Theorem 11.31 from Section 11.11). ■

LEMMA 15.10. For every point x in $[a,b]$

$$\lim_{(I,x)\Rightarrow x} \frac{\Delta F(I)}{\Delta G(I)} = f(x)$$

except for points x in a set N for which $\mathcal{V}_{G*}(N) = 0$.

Proof. This is similar to the proof of the preceding lemma, but requires a direct application of the limsup limit Lemma 11.28. ■

EXERCISE 15.11. In Lemma 15.9 show that, at almost every point x,

$$\overline{D}F(x) = f(x)\overline{D}G(x) \quad \text{and} \quad \underline{D}F(x) = f(x)\underline{D}G(x)$$

if $f(x) \geq 0$ while

$$\overline{D}F(x) = f(x)\underline{D}G(x) \quad \text{and} \quad \underline{D}F(x) = f(x)\overline{D}G(x)$$

if $f(x) \leq 0$. In particular $F'(x) = 0$ at almost every point x where $f(x) = 0$.

15.3.1. Existence of the integral from derivative statements. The existence of the integral

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

can be deduced from a variety of differentiation statements. For example:

LEMMA 15.12. If at every point x of a compact interval $[a,b]$

$$\lim_{(I,x)\Rightarrow x} \frac{\Delta F(I) - f(x)\Delta G(I)}{\mathcal{L}(I)} = 0$$

except for points x in a set N for which $\mathcal{V}_G{}^*(N) = \mathcal{V}_F{}^*(N) = 0$ then

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

Proof. This is a direct application of the fundamental theorem of the calculus (Theorem 11.32 from Section 11.11). ■

15.4. Existence of the Stieltjes integral

THEOREM 15.13. Let f, $G : \mathbb{R} \to \mathbb{R}$ and suppose that f is continuous on a compact interval $[a,b]$ and that G is monotonic nondecreasing throughout that interval. Then the Stieltjes integral exists and

$$\left| \int_a^b f(x)\,dG(x) \right| \le \left(\sup_{t\in[a,b]} |f(t)| \right) [G(b) - G(a)].$$

Proof. The methods of Theorem 9.3.1 can be used here with only minor modifications to show that the integral exists. The inequality is easy since, if π is any partition of $[a,b]$, then

$$\left| \sum_{(I,x)\in\pi} f(x)\Delta G(I) \right| \le \left(\sup_{t\in[a,b]} |f(t)| \right) \sum_{(I,x)\in\pi} \Delta G(I) = \left(\sup_{t\in[a,b]} |f(t)| \right) [G(b) - G(a)].$$

■

THEOREM 15.14. Let f, $G : \mathbb{R} \to \mathbb{R}$ and suppose that f is continuous on a compact interval $[a,b]$ and that G has bounded variation throughout that interval. Then the Stieltjes integral exists and

$$\left| \int_a^b f(x)\,dG(x) \right| \le \left(\sup_{t\in[a,b]} |f(t)| \right) \overline{\int_a^b} |\Delta G|.$$

Proof. Again the methods of Theorem 9.3.1 can be used here with only minor modifications to show that the integral exists. The inequality is easy since, if π is any partition of $[a,b]$, then

$$\left| \sum_{(I,x)\in\pi} f(x)\Delta G(I) \right| \le \left(\sup_{t\in[a,b]} |f(t)| \right) \sum_{(I,x)\in\pi} |\Delta G(I)| \le \left(\sup_{t\in[a,b]} |f(t)| \right) \overline{\int_a^b} |\Delta G|.$$

■

15.5. Helley's first theorem

The study of the Stieltjes integral $\int_a^b f(x)dG(x)$ finds some of it most interesting features when f is assumed to be continuous and G of bounded variation. In that setting two theorems of Helley prove most useful.

THEOREM 15.15 (Helley). Let $\{G_n\}$ be a sequence of functions of bounded variation on a compact interval $[a, b]$ so that $G_n(a) = 0$ for all n and the total variations are uniformly bounded, i.e., so that

$$\sup_n \overline{\int_a^b} |\Delta G_n| = K < \infty.$$

Then there must be a subsequence $\{G_{n_k}\}$ so that

$$\lim_{k\to\infty} G_{n_k}(x) = G(x)$$

exists at every point x in $[a, b]$ and

$$\overline{\int_a^b} |\Delta G| \leq K.$$

Proof. As usual we can reduce the theorem, using the Jordan decomposition, to the following case: assume that $G_n(a) = 0$, G_n is nondecreasing, and $G_n(b) \leq K < \infty$ for all n.

Step 1. Take a single point x in $[a, b]$. Then certainly, since $\{G_{n_k}(x)\}$ is a bounded sequence of real numbers, there is a subsequence so that

$$\lim_{k\to\infty} G_{n_k}(x) = G(x)$$

at that one point.

Step 2. Take a sequence of points $\{x_p\}$ in $[a, b]$. Then for each point x_1, x_2, ... in turn we can select a subsequence and a further subsequence of that etc. so that the subsequence converges at the point in consideration. A diagonal argument produces a single subsequence with the property that

$$\lim_{k\to\infty} G_{n_k}(x_p) = G(x_p)$$

exists for each p (but perhaps not any other point in the interval).

Step 3. Let us assume then that we have selected by now a subsequence $\{G_{n_k}\}$ so that

$$\lim_{k\to\infty} G_{n_k}(x) = F(x)$$

exists for $x = a$, $x = b$, and every x in $[a, b]$ that is rational. This defines F at these points, but not yet at any irrational number in (a, b). F is nondecreasing, relative to the points where we have defined it.

For each irrational x in (a, b) define

$$F(x) = \sup\{G(r) : \ r < x \text{ and } r \text{ rational}\}.$$

It is easy to check that

$$\lim_{k\to\infty} G_{n_k}(x) = F(x)$$

provided we can be assured that x is a point of continuity of all these functions. There may be, of course, points x that are irrational and not a point of continuity for the functions: but there are only countably many such points. Thus there is a set C, possibly, that is countable and we have so far verified that

$$\lim_{k\to\infty} G_{n_k}(x) = F(x)$$

for all x not in C.

Final step. Since the set C is countable we can arrange it into a sequence of points and apply Step 2 to obtain a further subsequence of $G_{n_k}(x)$ that is, indeed, convergent at every point. Define $G(x)$ to be the limit of this final subsequence of $\{G_n(x)\}$. ■

15.6. Helley's second theorem

THEOREM 15.16 (Helley). Let $\{f_n\}$, $\{G_n\}$ be sequences of real functions on a compact interval $[a,b]$ and suppose that

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{and} \quad \lim_{n\to\infty} G_n(x) = G(x)$$

for each $x \in [a,b]$. Suppose that $\{f_n\}$ are equicontinuous on $[a,b]$ and that

$$\overline{\int_a^b} |\Delta G_n| < K < \infty$$

for all n. Then

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dG_n(x) = \int_a^b f(x)\, dG(x).$$

Proof. The methods of Theorem 9.3.1 can be can be used (cf. Theorem 15.13) to show that the functions $f_n \Delta G_n$ are equi-integrable on $[a,b]$. Then our standard methods from Theorem 14.25 quickly obtain the limit identity of the theorem. ■

15.7. Linear functionals

The structure of the Stieltjes integral as it applies to continuous functions is revealed in Theorems 15.13 and 15.14. A deeper look at that structure uncovers an interesting characterization of the integral itself.

For any linear combination of continuous functions $f_1,\ f_2 : [a,b] \to \mathbb{R}$ we note that

$$\int_a^b [rf_1(x) + sf_2(x)]\, dG(x) = r\int_a^b f_1(x)\, dG(x) + s\int_a^b f_2(x)\, dG(x)$$

would be true for any function G of bounded variation. Let us explore this linear structure.

15.7.1. The space of continuous functions. Let $\mathcal{C}[a,b]$ denote the collection of all continuous functions defined on a compact interval $[a,b]$. Since linear combinations of continuous functions are again continuous this has the structure of a linear space.

We commonly study this space in combination with the following metric that measures the size of functions:

$$\|f\| = \sup\{|f(t)| : t \in [a,b]\}$$

which will assume a finite, nonnegative value for every $f \in \mathcal{C}[a,b]$.

EXERCISE 15.17. Let $f,\ f_1,\ f_2,\ \ldots$ be a sequence of elements of $\mathcal{C}[a,b]$. Show that $\{f_n\}$ converges uniformly to f on the interval $[a,b]$ if and only if

$$\lim_{n\to\infty} \|f_n - f\| = 0.$$

15.7.2. Linear functionals.

DEFINITION 15.18. A mapping $\Gamma : \mathcal{C}[a,b] \to \mathbb{R}$ is a *linear functional* if

$$\Gamma(rf_1 + sf_2) = r\Gamma(f_1) + r\Gamma(f_2)$$

for all f_1, f_2 in $\mathcal{C}[a,b]$ and all real numbers r and s.

DEFINITION 15.19. A linear functional Γ on $\mathcal{C}[a,b]$ is *positive* if

$$\Gamma(f) \geq 0$$

for all nonnegative functions f in $\mathcal{C}[a,b]$.

DEFINITION 15.20. A linear functional Γ on $\mathcal{C}[a,b]$ is *bounded* if

$$\|\Gamma\| = \sup\left\{\frac{|\Gamma(f)|}{\|f\|} : \ f \text{ in } \mathcal{C}[a,b] \text{ and } \|f\| \neq 0\right\} < \infty.$$

EXERCISE 15.21. Let $G : [a,b] \to \mathbb{R}$ be a monotonic nondecreasing function. Show that

$$\Gamma_G(f) = \int_a^b f(x)\, dG(x)$$

defines a positive linear functional on $\mathcal{C}[a,b]$ for which

$$\|\Gamma_G\| = G(b) - G(a).$$

EXERCISE 15.22. Let $G : [a,b] \to \mathbb{R}$ be a function of bounded variation throughout $[a,b]$. Show that

$$\Gamma_G(f) = \int_a^b f(x)\, dG(x)$$

defines a bounded linear functional on $\mathcal{C}[a,b]$ for which

$$\|\Gamma_G\| = \overline{\int_a^b} |\Delta G|,$$

i.e., the variation of G on $[a,b]$.

EXERCISE 15.23. Let f, f_1, f_2, ... be a sequence of elements of $\mathcal{C}[a,b]$. Show that if $\{f_n\}$ converges uniformly to f on the interval $[a,b]$ and Γ is a bounded linear functional on $\mathcal{C}[a,b]$ then

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

Hint: Use Exercise 15.17.

15.7.3. Basic structural properties. The structure of linear functionals in this kind of setting can be studied by formal algebraic means, without knowing many of the details of the space itself.

LEMMA 15.24. Every positive linear functional Γ on $\mathcal{C}[a,b]$ is bounded.

Proof. Let $g(x) = 1$ for all x and let f be any function in $\mathcal{C}[a,b]$. Write $c = \|f\|$ and note that

$$-cg(x) \leq f(x) \leq cg(x).$$

By the linearity and positivity of Γ

$$-c\Gamma(g) \leq \Gamma(f) \leq c\Gamma(g).$$

Thus

$$\|\Gamma\| \leq \Gamma(g).$$

(In fact a little further thought will reveal that $\|\Gamma\| = \Gamma(g)$.) ∎

LEMMA 15.25. Every bounded linear functional Γ on $\mathcal{C}[a,b]$ can be expressed as the difference of two positive linear functionals

$$\Gamma = \Gamma_1 - \Gamma_2.$$

Proof. Write for a nonnegative f in $\mathcal{C}[a,b]$,

$$\Gamma_+(f) = \sup_{0 \leq g \leq f} \Gamma(g).$$

Verify that Γ_+ satisfies

$$\Gamma_+(cf) = c\Gamma_+(f)$$

for all such f and $c \geq 0$. Verify that Γ_+ satisfies

$$\Gamma_+(f_1 + f_2) = \Gamma_+(f_1) + \Gamma_+(f_2)$$

for all nonnegative f_1 and f_2 in $\mathcal{C}[a,b]$.

To extend Γ_+ to all f in $\mathcal{C}[a,b]$, consider any two numbers r and s larger than $\|f\|$. Then $f + r$ and $f + s$ are nonnegative continuous functions. Check that

$$\Gamma_+(f + r + s) = \Gamma_+(f + r) + \Gamma_+(s) = \Gamma_+(f + s) + \Gamma_+(r)$$

showing that

$$\Gamma_+(f + r) - \Gamma_+(r) = \Gamma_+(f + s) - \Gamma_+(s).$$

Thus we can choose to define

$$\Gamma_+(f) = \Gamma_+(f + r) - \Gamma_+(r)$$

for any choice of $r > \|f\|$. Verify that Γ_+ so defined is a positive linear functional on $\mathcal{C}[a,b]$.

The proof is completed by checking that

$$\Gamma_- = \Gamma_+ - \Gamma$$

is also a positive linear functional. ∎

15.8. Representation of positive linear functionals on $\mathcal{C}[a,b]$

THEOREM 15.26. Let Γ be a positive linear functional on $\mathcal{C}[a,b]$. Then there must exist a monotonic nondecreasing function $G : [a,b] \to \mathbb{R}$ such that

$$\Gamma(f) = \int_a^b f(x)\, dG(x)$$

and for which

$$\|\Gamma\| = G(b) - G(a).$$

Proof. We can translate the theorem to the interval $[0,1]$ by finding an affine transformation $\phi : [a,b] \to [0,1]$ and using ϕ and ϕ^{-1} to translate back and forth between functions on the two compact intervals $[a,b]$ and $[0,1]$.

Fix an integer n. Recall that

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1$$

for all x in $[0,1]$ and that each of the terms in the sum is a positive polynomial on $[0,1]$. It follows that

$$\Gamma(1) = \sum_{k=0}^{n} \Gamma\left(\left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k} \right)$$

and each term in this sum is a nonnegative real number.

Define a function G_n by writing $G_n(0) = 0$, $G_n(1) = \Gamma(1)$ and on each of the intervals $(0 < x < 1/n)$, $(1/n \leq x < 2/n)$, $\dots ((n-1)/n \leq x < 1)$ G_n assumes a different constant value. These n constant values are defined by $c_0 = 0$, and for $k = 1, 2, \dots, n$

$$c_k = c_{k-1} + \Gamma\left(\left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k} \right).$$

Note that the finite sequence of values c_k is nondecreasing. Thus G_n is a nondecreasing "step" function whose values at the endpoints of $[0,1]$ are 0 and $\Gamma(1)$ and whose value $G_n(x)$ for $(k-1)/n \leq x < k/n$ is the constant c_k. Moreover all satisfy the inequality

$$0 \leq G_n(x) \leq \Gamma(1).$$

Check by direct computation that, for any continuous function f on $[0,1]$,

$$\int_a^b f(x)\, dG_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) c_k = \sum_{k=0}^{n} \Gamma\left(f\left(\frac{k}{n}\right) \left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k} \right).$$

Recall from Theorem 4.16.1 that we write the Bernstein polynomial for f as

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k}.$$

Thus what we have verified is that

$$\int_0^1 f(x)\, dG_n(x) = \Gamma(B_n(x)).$$

We will be able then to take advantage of the fact that B_n converges uniformly to f on $[0,1]$, which implies that

$$\lim_{n\to\infty} \Gamma(B_n(x)) = \Gamma(f).$$

Notice that, if the sequence of functions $\{G_n\}$ happened to converge to a function G at every point, then we would be done since we would know that

$$\Gamma(f) = \lim_{n\to\infty} \Gamma(B_n(x)) = \lim_{n\to\infty} \int_0^1 f(x)\, dG_n(x) = \int_0^1 f(x)\, dG(x).$$

While that may be too much to hope for, we can achieve the same result by passing to a subsequence.

Apply Helley's first theorem to the sequence of monotonic functions $\{G_n\}$: there is then a subsequence $\{G_{n_k}\}$ so that

$$G(x) = \lim_{k\to\infty} G_{n_k}(x)$$

at each point x of $[0,1]$. In particular we can deduce that

$$\lim_{k\to\infty} \int_0^1 f(x)\, dG_{n_k}(x) = \lim_{k\to\infty} \Gamma(B_{n_k}(x)) = \Gamma(f).$$

Helley's second theorem assures us that

$$\lim_{k\to\infty}\int_0^1 f(x)\,dG_{n_k}(x)=\int_0^1 f(x)\,dG(x).$$

It follows that

$$\Gamma(f)=\int_0^1 f(x)\,dG(x)$$

as we require. ■

15.9. Representation of bounded linear functionals on $\mathcal{C}[a,b]$

THEOREM 15.27. Let Γ be a bounded linear functional on $\mathcal{C}[a,b]$. Then there must exist a function $G:[a,b]\to\mathbb{R}$ that has bounded variation throughout $[a,b]$ and so that

$$\Gamma(f)=\int_a^b f(x)\,dG(x)$$

and

$$\|\Gamma\|=\overline{\int_a^b}|\Delta G|,$$

i.e., the variation of G on $[a,b]$.

Proof. Lemma 15.25 shows that every bounded linear functional on $\mathcal{C}[a,b]$ can be expressed as a difference of positive linear functionals Γ_+ and Γ_-. Apply Theorem 15.26 to obtain two monotonic functions G_1 and G_2 representing Γ_+ and Γ_- and verify that $G=G_1-G_2$ works here. ■

15.10. Hellinger integrals

Let F, $G:[a,b]\to\mathbb{R}$ where $[a,b]$ is a compact interval. If the interval function

$$\frac{\Delta F\Delta G}{\mathcal{L}}$$

is integrable on $[a,b]$ then for its integral we will employ the suggestive notation

$$\int_a^b \frac{dF(x)\,dG(x)}{dx}.$$

This is called a *Hellinger integral* after Ernst Hellinger (1883–1950). For the simplest cases the theory is easy to develop. Usually the Hellinger integral can be represented by Stieltjes or ordinary integrals.

15.10.1. Differentiation of Hellinger integrals.

THEOREM 15.28. Let F, $G:[a,b]\to\mathbb{R}$ where $[a,b]$ is a compact interval. If the interval function $\Delta F\Delta G/\mathcal{L}$ is integrable on $[a,b]$ with indefinite integral

$$H(t)=\int_a^t \frac{dF(x)\,dG(x)}{dx}\qquad (a<t\le b)$$

then $H'(x)=F'(x)G'(x)$ at almost every point x in $[a,b]$ at which both F and G are differentiable.

Proof. In order for H to be an indefinite integral we must have

$$\int_a^b |\Delta H - \Delta F \Delta G/\mathcal{L}| = 0.$$

We apply the fundamental theorem of the calculus (Theorem 11.31 from Section 11.11). This provides that, for almost every point x in $[a, b]$,

$$\lim_{(I,x)\Rightarrow x} \left|\Delta H(I) - \frac{\Delta F(I)\Delta G(I)}{\mathcal{L}(I)}\right| \frac{1}{\mathcal{L}(I)} = 0.$$

From that we deduce the assertion of the theorem. ■

15.10.2. Reduction theorem.

THEOREM 15.29. Let F, $G : [a, b] \to \mathbb{R}$ where $[a, b]$ is a compact interval. Suppose that F is absolutely continuous in the sense of Vitali on $[a, b]$ and that G is a Lipschitz function. Then the interval function $\Delta F \Delta G/\mathcal{L}$ is integrable on $[a, b]$ and

$$\int_a^t \frac{dF(x)\,dG(x)}{dx} = \int_a^b F'(x)dG(x) = \int_a^b F'(x)G'(x)\,dx.$$

Proof. As usual we interpret the functions F' and G' (which exist only almost everywhere) in the integrals by replacing them with Lebesgue equivalent functions. Specifically let $f(x) = F'(x)$ and $g(x) = G'(x)$ for x where these exist and set $f(x) = g(x) = 0$ elsewhere. We know that f and g are integrable on $[a, b]$ and that

$$\int_a^b |\Delta F - f\mathcal{L}| = \int_a^b |\Delta G - g\mathcal{L}| = 0.$$

Let M be a Lipschitz constant for G; then

$$\left|\frac{\Delta G(I)}{\mathcal{L}(I)}\right| \leq M$$

for all subintervals I of $[a, b]$.

Now we merely compute, for any interval-point pair (I, x), that

$$\left|\frac{\Delta F(I)\Delta G(I)}{\mathcal{L}(I)} - f(x)\Delta G(I)\right| = \left|\frac{\Delta G(I)}{\mathcal{L}(I)}\right| \cdot |\Delta F(I) - f(x)\mathcal{L}(I)|$$
$$\leq M\,|\Delta F(I) \ - f(x)\mathcal{L}(I)|\,.$$

Integrating gives us

$$\int_a^b |\Delta F \Delta G/\mathcal{L} - f\Delta G| = 0.$$

From that we conclude that

$$\int_a^t \frac{dF(x)\,dG(x)}{dx} = \int_a^b f(x)dG(x)$$

provided one of the integrals exists (which we do not yet know).

From the fact that

$$\int_a^b |\Delta G - g\mathcal{L}| = 0$$

we can deduce using simple measure theory arguments, that multiplying by any function f, has the effect that

$$\int_a^b |f\Delta G - fg\mathcal{L}| = 0.$$

Thus we conclude that

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

provided one of the integrals exists (which we still do not know).

To complete the proof then it suffices to show the existence of one of the integrals. Since F is absolutely continuous in the sense of Vitali it follows that f is absolutely integrable. Thus f is almost continuous. Multiplying by the bounded function g produces another almost continuous function fg that is also almost continuous. It is also absolutely integrable: this is because

$$\int_{[a,b]} |f(x)g(x)|\,dx \leq M \int_{[a,b]} |f(x)|\,dx < \infty.$$

In particular the integral

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

exists and the proof is complete. ■

As a project for the student we suggest the following weakening of the hypothesis for F and strengthening of the hypothesis for G in the theorem.

Exercise 15.30. Let F, $G : [a,b] \to \mathbb{R}$ where $[a,b]$ is a compact interval. Suppose that F is absolutely continuous on $[a,b]$ and that G is the indefinite integral of a function of bounded variation. Show that the interval function $\Delta F \Delta G/\mathcal{L}$ is integrable on $[a,b]$ and

$$\int_a^t \frac{dF(x)\,dG(x)}{dx} = \int_a^b F'(x)dG(x) = \int_a^b F'(x)G'(x)\,dx.$$

15.11. Bounded linear functionals on $\mathcal{AC}_0[a,b]$

The Stieltjes integral was used earlier in this chapter to provide an integral representation of bounded linear functionals. See Section 15.7. A similar theory is available in a variety of settings. Here is one, using both a Stieltjes integral and a Hellinger integral.

Definition 15.31. By the collection $\mathcal{AC}_0[a,b]$ we denote the linear space of all functions $F : [a,b] \to \mathbb{R}$ that are absolutely continuous in the sense of Vitali and for which $F(a) = 0$.

We commonly study this space in combination with the following metric that measures the size of functions:

$$\|F\| = \int_a^b \omega F$$

which will assume a finite, nonnegative value for every F in $\mathcal{AC}_0[a,b]$. (The norm used in the space $\mathcal{C}[a,b]$, we will recall, was quite different and the use of the same symbol should be tolerated by the student.)

DEFINITION 15.32. A linear functional Γ on $\mathcal{A}C_0[a, b]$ is *bounded* if

$$\|\Gamma\| = \sup\left\{\frac{|\Gamma(F)|}{\|F\|} : F \text{ in } \mathcal{A}C_0[a, b] \text{ and } \|F\| \neq 0\right\} < \infty.$$

EXERCISE 15.33. Let $G : [a, b] \to \mathbb{R}$ be a Lipschitz function on $[a, b]$. Show that

$$\Gamma_G(F) = \int_a^b \frac{dF(x)dG(x)}{dx}$$

defines a bounded linear functional on $\mathcal{A}C_0[a, b]$ for which

$$\|\Gamma_G\| = Lip_G[a, b],$$

i.e., the Lipschitz constant for G on $[a, b]$.

EXERCISE 15.34. Characterize the bounded linear functionals on $\mathcal{A}C_0[a, b]$.

EXERCISE 15.35. Characterize the bounded linear functionals on $\mathcal{A}CG_0[a, b]$, the linear space of all functions F that are absolutely continuous on $[a, b]$ and for which $F(a) = 0$.

Hint: For the norm in this space do not use $\|F\| = \int_a^b \omega F$ (which may be infinite) but instead just use $\|F\| = \omega F([a, b]$. The clue to the characterization would be in Exercise 15.30.

APPENDIX: Formal Theory of the Calculus

The fundamental program of the calculus connects the notions of measure, integral, and derivative. Each of these concepts has a simple realization that can be expressed by covering relations. The natural interconnections are mostly transparent since they can be interpreted as different aspects of the same structure.

In the formal theory of this chapter we express this in more systematic language. The reader with a taste for more abstract formal theories need not read anything else since essentially the calculus is contained in these formal considerations.

15.12. Covering Relations

We review some basic definitions.

Definition 15.36. A *covering relation* is a collection whose elements are pairs $([x, y], z)$ with $x < y$ and $z \in [x, y]$, consisting of a compact interval and a point belonging to that interval.

The family $\mathbb{H}$ denotes the set of *all* such pairs (I, x). This is itself a covering relation and all other covering relations are subsets of $\mathbb{H}$.

The following definitions are critical in handling and manipulating covering relations.

Definition 15.37. Given a covering relation β and a set E the following two subsets (sometimes called, colorfully, *prunings*) of β often play a role:

$$\beta[E] = \{([x, y], z) \in \beta : z \in E\}$$

and

$$\beta(E) = \{([x, y], z) \in \beta : [x, y] \subset E\}.$$

Definition 15.38. We say that β *ignores* a set E if $\beta[E] = \emptyset$.

Definition 15.39. We say that β *ignores a point* x if $\beta[\{x\}] = \emptyset$.

15.13. Functions defined on covering relations

We need a language for functions

$$h : \mathbb{H} \to \mathbb{R},$$

i.e., functions for which $h([x, y], z)$ assumes a real value.

Length of an interval-point pair: $\mathcal{L}$ (standing for "*length*") denotes the function defined by writing as

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

This extends the Lebesgue measure $\mathcal{L}$ to interval-point pairs. We retain, of course, the meaning of $\mathcal{L}$ for arbitrary sets.

Length with a weight: Let $f : \mathbb{R} \to \mathbb{R}$. We write

$$f\mathcal{L}([x,y],z) = f(z)(y-x).$$

The function f can be thought of as a weight multiplying the length.

Increment of a function: Let $F : \mathbb{R} \to \mathbb{R}$. Then ΔF denotes the function defined by

$$\Delta F([x,y],z) = \Delta F([x,y]) = F(y) - F(x)$$

called the increment of the function F.

Product with a function: If $h : \mathbb{H} \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ we denote by fh the function defined on $\mathbb{H}$ by the identity

$$fh(I,x) = f(x)h(I,x).$$

15.14. Partitions

Partitions and subpartitions play a critical role in defining the integral and the measure.

(partition): A finite covering relation π is a *partition* of $[a,b]$ if

(a) Whenever (I,x), (J,y) are two distinct elements of π the intervals I and J must not overlap, and

(b) $\bigcup_{(I,x)\in\pi} I = [a,b]$.

(subpartition): Any subset π of a partition is a *subpartition*.

15.15. Full and fine covers

Full covers contain *all sufficiently small* intervals near a point, while fine covers contain always *arbitrarily small* intervals near the point. The fine covers are defined as concept dual to the full covers, originating from work of Vitali in the early 20th century.

Definition 15.40. A covering relation β is a *full cover* for a set E if for every $z \in E$ there is a $\delta > 0$ so that every pair $([x,y],z)$ with $0 < y-x < \delta$ and $z \in [x,y]$ must belong to β.

Definition 15.41. A covering relation β is a *fine cover* for a set E if for every $z \in E$ and every $\delta > 0$ there must exist at least one pair $([x,y],z)$ with $0 < y-x < \delta$ and $z \in [x,y]$ that belongs to β.

15.16. Differentiation bases

We define two collections of covering relations:

$$\mathbf{B} = \{\beta : \ \beta \text{ is a full cover of } \mathbb{R}\}$$

and

$$\mathbf{B}^* = \{\beta : \ \beta \text{ is a fine cover of } \mathbb{R}\}.$$

We refer to $\mathbf{B}$ as *the basis of ordinary differentiation*. It could equally well be called the basis of *ordinary integration* since it can be used both to define and study derivatives or to define and study integrals. The collection $\mathbf{B}^*$ is called the *dual* of $\mathbf{B}$ for reasons that will become clear in the next section. Note that $\mathbf{B} \subset \mathbf{B}^*$.

15.17. Nature of the duality

The following lemmas, easily proved from the definitions, express the nature of the duality between the basis $\mathbf{B}$ and its dual $\mathbf{B}^*$ which is exploited in the theory.

LEMMA 15.42. Let β be a covering relation. Then β is a full cover of a set E if and only if, for every fine cover β_1 of E, the intersection $\beta \cap \beta_1$ ignores no point in E.

LEMMA 15.43. Let β be a covering relation. Then β is a fine cover of a set E if and only if, for every full cover β_1 of E, the intersection $\beta \cap \beta_1$ ignores no point in E.

In particular the dual basis $\mathbf{B}^*$ can be described as the collection of all covering relations β that have the property that $\beta \cap \beta_1$ ignores no point for all $\beta_1 \in \mathbf{B}$. The dual of $\mathbf{B}^*$ according to this definition would then be $\mathbf{B}$ itself. (In symbols $\mathbf{B}^{**} = \mathbf{B}$.)

15.18. Properties of full/fine covers

The differentiation basis $\mathbf{B}$ is a filter on $\mathbb{H}$. It is this property that allows limiting notions to be expressed. The basic limiting notion is used to express integrals, derivatives and measures. The dual basis $\mathbf{B}^*$ has a weaker filtering property.

LEMMA 15.44 (Full filtering property). Let β_1 and β_2 be full covers of a set E. Then if a covering relation β includes the intersection $\beta_1 \cap \beta_2$, β is also a full cover of E.

LEMMA 15.45 (Fine filtering property). Let β_1 be a full cover of a set E and let β_2 be a fine cover of E. Then if a covering relation β includes the intersection $\beta_1 \cap \beta_2$, β is a fine cover of E.

LEMMA 15.46 (Pruning by open sets). Let $E \subset G$ where G is open. If β is a full [fine] cover of E then so too is $\beta(G)$.

LEMMA 15.47 (Assembling the pieces). Let $\{E_i\}$ be a sequence of sets with $E \subset \bigcup_{i=1}^{\infty} E_i$ and let β_i be a sequence of covering relations. Suppose that each β_i is a full [fine] cover of E_i. Then $\bigcup_{i=1}^{\infty} \beta_i$ is a full [fine] cover of E.

LEMMA 15.48 (Cousin's Lemma). If β is a full cover of an interval $[a, b]$ then $\beta([a, b])$ is a Cousin cover of the interval $[a, b]$ and, hence, contains a partition of $[a, b]$.

15.18.1. Decomposition of full covers.

There is a decomposition of full covers that is often useful.

LEMMA 15.49 (Decomposition Lemma). Let β be a full cover of a set E. Then there is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E_n so that if x is any point in E_n and I is any subinterval of I_{kn} that contains x then (I, x) belongs to β.

15.19. Variation

Let $h : \mathbb{H} \to \mathbb{R}$. Sums of the form

$$\sum_{(I,x)\in\pi} h(I,x)$$

where π is a partition or a subpartition are central to our concerns. This notion of variation is the fundamental concept on which the measure construction is based.

DEFINITION 15.50 (Variation over a subpartition). If π is a subpartition then

$$V(h,\pi) = \sum_{(I,x)\in\pi} |h(I,x)|$$

is called the *variation of h over the subpartition*.

DEFINITION 15.51 (Variation over a covering relation)). If β is a covering relation then

$$V(h,\beta) = \sup_{\pi\subset\beta} V(h,\pi),$$

where the supremum is taken over all subpartitions π contained in β, is called the *variation for h over β*.

DEFINITION 15.52 (Variation over a differentiation basis). Let $\mathbf{B}_0$ be any collection of covering relations. Then

$$V(h;\mathbf{B}_0) = \inf_{\beta\in\mathbf{B}_0} V(h,\beta),$$

where the supremum is taken over all subpartitions π contained in β, is called the *variation for h over the basis* $\mathbf{B}_0$.

15.20. The measures

DEFINITION 15.53. Let $h : \mathbb{H} \to \mathbb{R}$. We define two variational measures, a full measure and a fine measure by writing, for any real set E,

$$V^*(h,E) = \inf_{\beta\in\mathbf{B}} V(h,\beta[E])$$

and

$$V_*(h,E) = \inf_{\beta\in\mathbf{B}^*} V(h,\beta[E]).$$

These set functions are measures on the real line [or what some call *outer* measures]. For both measures Borel sets are measurable. We say h has *zero variation* [full or fine], *finite variation*, *σ-finite variation* on a set if the corresponding measure [$V_*(h,E)$ or $V^*(h,E)$] has this property. The function h is *continuous* if the full variational measure vanishes on all countable sets; it is *absolutely continuous* if the full variational measure vanishes on sets of Lebesgue measure zero.

The variation also has properties as a function of the h here:

$$V^*(rh,E) = rV^*(h,E) \quad \text{and} \quad V_*(rh,E) = rV_*(h,E).$$

$$V_*(h_1+h_2,E) \le V_*(h_1,E) + V^*(h_2,E)$$

and

$$V^*(h_1+h_2,E) \le V^*(h_1,E) + V^*(h_2,E).$$

15.21. Vitali property

In general the fine measure is smaller than the full measure:

$$V_*(h, E) \leq V^*(h, E).$$

The identity plays a role in the theory and acquires the following name;

DEFINITION 15.54. Let $h : \mathbb{H} \to \mathbb{R}$. Then we say that h has the *Vitali property* provided that

$$V_*(h, E) = V^*(h, E)$$

for all E.

The special cases are worthy of note and play a fundamental role in the calculus. Using $\mathcal{L}$ (classical Lebesgue measure) we find the identity

$$\mathcal{L}(E) = V^*(\mathcal{L}, E) = V_*(\mathcal{L}, E)$$

for all sets $E \subset \mathbb{R}$. This is the classical Vitali covering theorem. A particularly important generalization is this:

> Suppose $F : \mathbb{R} \to \mathbb{R}$. Then ΔF has the Vitali property if and only if ΔF is continuous and has σ-finite variation.

It is natural to refer to this too as the *Vitali covering theorem*.

15.22. Kolmogorov equivalence

The most important concept in the theory is Kolmogorov equivalence. If two functions are Kolmogorov equivalent then for most purposes they are essentially the same object within the calculus. For example the relation between a function and its indefinite integral or a function and its derivative is one expressible as an equivalence relation and the calculus program is best understood from this viewpoint. The notion originated with the Russian mathematician Kolmogorov, and was exploited in this context by Henstock.

DEFINITION 15.55. Let $h_1,\ h_2 : \mathbb{H} \to \mathbb{R}$. Then h_1 and h_2 are *Kolmogorov equivalent* on a set E if

$$V^*(h_1 - h_2, E) = 0.$$

15.22.1. Differentials. The equivalence classes for this equivalence relation are known as *Leader differentials*, after Solomon Leader. (Leader's original idea used a finer differentiation basis than **B**.)

For any function $h : \mathbb{H} \to \mathbb{R}$ we can define

$$[h] = \{k : \mathbb{H} \to \mathbb{R} : V^*(h - k, \mathbb{R}) = 0\}.$$

The collection of all such objects $[h]$ is the space of differentials and any property that is shared by all members of these classes can be defined for differentials. For example we can say that $[h]$ has the Vitali property if any element $k \in [h]$ has the Vitali property. The reader is invited to formulate a theory of differentials that would give meaning to the expression

$$dF(x) = f(x)\,dx$$

and, in particular, allow a reasonable interpretation of the notation inside the integral

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

For some historical periods this was interpreted within Liebnitz's quaint notion of infinitesmals and for other periods as asserting no more than $F'(x) = f(x)$.

15.23. The Integral

For a function $h : \mathbb{H} \to \mathbb{R}$ and a compact interval $[a, b]$ we define an *upper integral* by

$$\overline{\int_a^b} h = \inf_{\beta \in \mathbf{B}} \sup_{\pi \subset \beta} \sum_{(I,x) \in \pi} h(I, x)$$

where the supremum is taken over all partitions π of $[a, b]$ contained in β.

Similarly we define a *lower integral*:

$$\underline{\int_a^b} h = \sup_{\beta \in \mathbf{B}} \inf_{\pi \subset \beta} \sum_{(I,x) \in \pi} h(I, x).$$

If the upper and lower integrals are identical and finite then we say h is *integrable* and we write the common value as $\int_a^b h$ simply called the integral.

We have already, in Chapter 6, introduced the classical calculus integral as

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

by exactly the same definition applied to the function $h(I, x) = f(x)\mathcal{L}(I)$. We would expect much of the theory of the more general integral to be similar, and we can apply the more general theory certainly to the classical integral.

15.24. Integrability of subadditive, continuous functions

Here is a simple, but useful, integrability criterion for the special case of subadditive functions. We say a real-valued function h is a *pure interval function* if $h(I)$ is defined for all compact intervals I. Naturally we also write $h(I, x) = h(I)$ and thus consider that h also defines a function on $\mathbb{H}$.

Definition 15.56. Let h be a pure interval function. Then h is subadditive if, for any partition π of an interval J,

$$0 \le h(J) \le \sum_{(I,x) \in \pi} h(I).$$

Theorem 15.57. Suppose that h is a pure subadditive interval function, that is continuous . Then

$$\underline{\int_a^b} h = \overline{\int_a^b} h$$

and h is integrable on $[a, b]$ if and only if $\overline{\int_a^b} h < \infty$.

15.25. Two summation lemmas

LEMMA 15.58. Suppose that h, $h_n : \mathbb{H} \to \mathbb{R}$ for $n = 1, 2, 3, \ldots$ are nonnegative functions and $[a, b]$ is a compact interval. Suppose that

$$h(I, x) \geq \sum_{n=1}^{\infty} h_n(I, x)$$

for all pairs (I, x) then

$$\underline{\int_a^b} h \geq \sum_{n=1}^{\infty} \left(\underline{\int_a^b} h_n \right).$$

LEMMA 15.59. Suppose that h, $h_n : \mathbb{H} \to \mathbb{R}$ for $n = 1, 2, 3, \ldots$ are nonnegative functions and $[a, b]$ is a compact interval. Suppose that for each x and $t < 1$ there is an integer N so that

$$th(I, x) \leq \sum_{n=1}^{N} h_n(I, x) \leq \sum_{n=1}^{\infty} h_n(I, x)$$

for all pairs $(I, x) \in \mathbb{H}$. Then

$$\overline{\int_a^b} h \leq \sum_{n=1}^{\infty} \left(\overline{\int_a^b} h_n \right).$$

15.26. Metric characterizations of integral

There are the usual Cauchy criteria.

THEOREM 15.60 (Cauchy's first criterion). A function $h : \mathbb{H} \to \mathbb{R}$ is integrable on a compact interval $[a, b]$ if and only if there is a number c so that for all $\epsilon > 0$ an element $\beta \in \mathbf{B}$ can be found so that

$$\left| \sum_{(I,x)\in\pi} h(I, x) - c \right| < \epsilon$$

for all partitions π of $[a, b]$ contained in β.

THEOREM 15.61 (Cauchy's second criterion). A function $h : \mathbb{H} \to \mathbb{R}$ is integrable on a compact interval $[a, b]$ if and only if for all $\epsilon > 0$ an element $\beta \in \mathbf{B}$ can be found so that

$$\left| \sum_{(I,x)\in\pi} h(I, x) - \sum_{(I',x')\in\pi'} h(I', x') \right| < \epsilon$$

for all partitions π, π' of $[a, b]$ contained in β.

15.27. Rudimentary properties of the integral

THEOREM 15.62 (Integrability on subintervals). If a function $h : \mathbb{H} \to \mathbb{R}$ is integrable on a compact interval $[a, b]$ then it is also integrable on any subinterval $[c, d]$ and there is a real function $H : [a, b] \to \mathbb{R}$ (called an *indefinite integral of h*) for which

$$\int_c^d h = H(d) - H(c) \qquad (a \leq c < d \leq b).$$

THEOREM 15.63 (Linear combinations). Linear combinations of integrable functions are also integrable and

$$\int_a^b [rh_1(x) + sh_2(x)]\,dx = r\int_a^b h_1(x)\,dx + s\int_a^b h_2(x)\,dx.$$

THEOREM 15.64 (The vanishing integral). A necessary and sufficient condition for a function $h : \mathbb{H} \to \mathbb{R}$ to be integrable on a compact interval $[a,b]$ with $\int_c^d h = 0$ for all $[c,d] \subset [a,b]$ is that

$$\int_a^b |h| = 0.$$

15.28. Henstock's criterion

The most useful characterization of the relation that must hold between a function and its indefinite integral is the general version of the Henstock criterion from Chapter 6. Here both the statement and the proof are much more transparent. In order for

$$\int_c^d h = \Delta H([c,d])$$

when $[c,d] \subset [a,b]$, we must have

$$\int_c^d (h - \Delta H) = 0.$$

By the vanishing integral property of Theorem 15.64 this is equivalent to

$$\int_a^b |h - \Delta H| = 0.$$

THEOREM 15.65 (Henstock's criterion). A necessary and sufficient condition for the relation

$$\int_c^d h = H(d) - H(c) \quad [c,d] \subset [a,b]$$

to hold for a function $h : \mathbb{H} \to \mathbb{R}$ and its indefinite integral $H : [a,b] \to \mathbb{R}$ on a compact interval $[a,b]$ is that

$$\int_a^b |\Delta H - h| = 0.$$

Note that, in general, we must have

$$V^*(\Delta H - h, (a,b)) \leq \overline{\int_a^b} |\Delta H - h| \leq V^*(H - h, [a,b]).$$

Thus Henstock's criterion implies the Kolmogorov equivalence of ΔH and h on (a,b) and is in turn implied by the Kolmogorov equivalence of ΔH and h on $[a,b]$.

COROLLARY 15.66. If H is the indefinite integral of a function $h : \mathbb{H} \to \mathbb{R}$ on a compact interval $[a,b]$, then

$$V^*(h,E) = V^*(\Delta H, E)$$

and

$$V_*(h,E) = V_*(\Delta H, E)$$

for all $E \subset (a,b)$.

Corollary 15.67. Suppose that $h_1,\ h_2 : \mathbb{H} \to \mathbb{R}$ are Kolmogorov equivalent on a compact interval $[a,b]$. Then h_1 is integrable on $[a,b]$ if and only if h_2 is integrable on $[a,b]$, and necessarily

$$\int_c^d h_1 = \int_c^d h_2$$

for all $[c,d] \subset [a,b]$.

15.29. Limit properties of integrals

If $h,\ h_p : \mathbb{H} \to \mathbb{R}$ for $p = 1,2,3,\dots$ and

$$\lim_{p\to\infty} h_p(I,x) = h(I,x)$$

for each $(I,x) \in \mathbb{H}$, one would expect that, under some appropriate hypotheses, there might be a conclusion that

$$\lim_p \int_a^b h_p = \int_a^b h.$$

As usual in investigations involving interchange of limit operations an assumption of uniformity provides the answer.

Definition 15.68. Let $h_p : \mathbb{H}([a,b]) \to \mathbb{R}$ for $p = 1,2,3,\dots$. We say that $\{h_p\}$ are *equi-integrable* on $[a,b]$ if there are numbers c_p so that for all $\epsilon > 0$ an element $\beta \in \mathbf{B}$ can be found so that

$$\left| \sum_{(I,x)\in\pi} h_p(I,x) - c_p \right| < \epsilon \tag{15.1}$$

for all partitions π of $[a,b]$ contained in β.

Theorem 15.69. Let $h,\ h_p : \mathbb{H}([a,b]) \to \mathbb{R}$ for $p = 1,2,3,\dots$. Suppose that $\lim_{p\to\infty} h_p(I,x) = h(I,x)$ for each pair $(I,x) \in \mathbb{H}([a,b])$ and that $\{h_p\}$ are equi-integrable on $[a,b]$. Then $\lim_{p\to\infty} \int_a^b h_p$ exists, h is integrable on $[a,b]$ and

$$\lim_{p\to\infty} \int_a^b h_p(x)\,dx = \int_a^b h.$$

15.30. Limits

Suppose that $h : \mathbb{H} \to \mathbb{R}$ and x is a point in $\mathbb{R}$. Define limits

$$\limsup_{(I,x)\Rightarrow x} h(I,x) = \inf_{\beta\in\mathbf{B}} \left(\sup\{h(I,x) : (I,x) \in \beta\}\right)$$

and

$$\liminf_{(I,x)\Rightarrow x} h(I,x) = \sup_{\beta\in\mathbf{B}} \left(\inf\{h(I,x) : (I,x) \in \beta\}\right).$$

As usual, if the limsup and liminf assume the same value (including $\pm\infty$) then we write $\lim_{(I,x)\Rightarrow x} h(I,x)$ for the common value.

15.30.1. Dual form of limits. Naturally the limit concept can be defined in terms of the dual basis, but notice that sup/infs turn into inf/sups.

THEOREM 15.70. Suppose that $h : \mathbb{H} \to \mathbb{R}$ and x is a point in $\mathbb{R}$. Then

$$\limsup_{(I,x)\Rightarrow x} h(I,x) = \sup_{\beta\in\mathbf{B}^*} \left(\inf\{h(I,x) : (I,x) \in \beta\}\right)$$

and

$$\liminf_{(I,x)\Rightarrow x} h(I,x) = \inf_{\beta\in\mathbf{B}^*} \left(\sup\{h(I,x) : (I,x) \in \beta\}\right).$$

15.30.2. Limits and covering relations.

LEMMA 15.71. Let $h : \mathbb{H} \to \mathbb{R}$ and $x \in \mathbb{R}$. The assertion

$$\limsup_{(I,x)\Rightarrow x} h(I,x) = c$$

is valid if and only if for all $r < c$ and all $c < s$ the covering relation

$$\beta_1 = \{(I,x) : h(I,x) < s\}$$

is full at x and the covering relation

$$\beta_2 = \{(I,x) : h(I,x) > r\}$$

is fine at x.

LEMMA 15.72. Let $h : \mathbb{H} \to \mathbb{R}$ and $x \in \mathbb{R}$. The assertion

$$\lim_{(I,x)\Rightarrow x} h(I,x) = c$$

if and only if for all $r < c < s$ the covering relation

$$\beta_1 = \{(I,x) : r < h(I,x) < s\}$$

is full at x.

15.30.3. Metric characterization of limits.

THEOREM 15.73 (Cauchy's first criterion). Let $h : \mathbb{H} \to \mathbb{R}$ and $x \in \mathbb{R}$. Then

$$\lim_{(I,x)\Rightarrow x} h(I,x) = c$$

for a finite number c if and only if, for all $\epsilon > 0$,

$$\beta = \{(I,t) : |h(I,t) - c| < \epsilon\}$$

is full at x.

THEOREM 15.74 (Cauchy's second criterion). Let $h : \mathbb{H} \to \mathbb{R}$ and $x \in \mathbb{R}$. Then

$$\lim_{(I,x)\Rightarrow x} h(I,x) = c$$

for some finite number c if and only if, for all $\epsilon > 0$, there is a covering relation β full at x such that, for all (I,x), (J,x) in β,

$$|h(I,t) - h(J,t)| < \epsilon$$

15.31. Limits and Measures

There are close connections between limit properties and measure properties. These lemmas illustrate the scope and nature of the methods. We shall be considering limits of expressions of the form

$$\left|\frac{h(I,x)}{k(I,x)}\right|.$$

Limits of such expressions allow a comparison between the growth properties of the two functions h and k. In special cases these would be called *derivatives*.

As is usually the case, we need to be concerned if the denominator is zero. To avoid this we might simply assume that $k(I,x) \neq 0$ for all pairs (I,x). Better, for these purposes, is to interpret $0/0$ as 0 and $c/0$ as ∞. The methods do not change and the results are then formally quite general.

15.31.1. Limsup comparison.

LEMMA 15.75. (Limsup Comparison) Let h, $k : \mathbb{H} \to \mathbb{R}$ and $x \in \mathbb{R}$. Suppose that

$$s < \limsup_{(I,x)\Rightarrow x} \left|\frac{h(I,x)}{k(I,x)}\right| < r$$

for every $x \in E$. Then

$$sV^*(k,E) \leq V^*(h,E) \leq rV^*(k,E)$$

and

$$V_*(h,E) \leq rV_*(k,E).$$

15.31.2. Liminf comparison. A similar theorem with a similar proof uses the limit inferior.

LEMMA 15.76. (Liminf Comparison) Let h, $k : \mathbb{H} \to \mathbb{R}$ and $x \in \mathbb{R}$. Suppose that

$$s < \liminf_{(I,x)\Rightarrow x} \left|\frac{h(I,x)}{k(I,x)}\right| < r$$

for every $x \in E$. Then

$$sV_*(k,E) \leq V_*(h,E) \leq rV^*(k,E)$$

and

$$sV^*(k,E) \leq V^*(h,E).$$

15.32. Kolmogorov equivalence from a limit

LEMMA 15.77 (Kolmogorov equivalence from limit). Suppose that h, $k : \mathbb{H} \to \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$, E is a set of real numbers and that $V^*(k,E) < \infty$. If

$$\lim_{(I,x)\Rightarrow x} \frac{h(I,x)}{k(I,x)} = f(x)$$

for every $x \in E$ then h and fk are Kolmogorov equivalent on E,

$$V^*(h - fk, E) = 0,$$

$$V_*(h,E) = V_*(fk,E),$$

and

$$V^*(h,E) = V^*(fk,E)..$$

15.33. Limit from a Kolmogorov equivalence

LEMMA 15.78. Suppose that h, $k : \mathbb{H} \to \mathbb{R}$ and that $V^*(h,E) = 0$. Then

$$\lim_{(I,x)\Rightarrow x} \frac{h(I,x)}{k(I,x)} = 0$$

at every point $x \in E$ except for those in a set N for which $V_*(k,N) = 0$.

COROLLARY 15.79 (Limit from Kolmogorov equivalence). Suppose that h_1, h_2, $k : \mathbb{H} \to \mathbb{R}$ and that h_1 and h_2 are Kolmogorov equivalent on a set E. Then

$$\lim_{(I,x)\Rightarrow x} \frac{h_1(I,x)}{k(I,x)} = \lim_{(I,x)\Rightarrow x} \frac{h_2(I,x)}{k(I,x)}$$

at every point $x \in E$ for which one at least of the limits exists, except for those x in a set N for which $V_*(k,N) = 0$.

15.34. The Lebesgue differentiation theorem

THEOREM 15.80 (Lebesgue). Let $h : \mathbb{H} \to \mathbb{R}$ satisfy the Vitali property. Then

$$\lim_{(I,x)\Rightarrow x} \left| \frac{h(I,x)}{\mathcal{L}(I)} \right|$$

exists (possibly $+\infty$) for all x except for those in a set N for which

$$\mathcal{L}(N) = V^*(h,N) = 0.$$

15.35. The fundamental theorem of the calculus

The fundamental theorem of the calculus asserts the relationship between a function and its indefinite integral. This relation is obtained easily as applications of the limit theorems of Section 15.31.

15.35.1. The derivative of the integral.

THEOREM 15.81. Let $h : \mathbb{H} \to \mathbb{R}$ and suppose that h is integrable on a compact interval $[a,b]$ with $H : [a,b] \to \mathbb{R}$ as an indefinite integral. Then

$$\lim_{(I,x)\Rightarrow x} \frac{\Delta H(I)}{h(I,x)} = 1$$

for every $x \in (a,b)$ excepting a set N of points for which $V_*(h,N) = V_*(H,N) = 0$. Moreover $V^*(h,E) = V^*(H,E)$, $V_*(h,E) = V_*(H,E)$ for all subsets E of (a,b), and H is continuous precisely at those points in (a,b) at which h is continuous.

15.35.2. The integral of the derivative. There is a partial converse, but note that the full measures are used here, whereas the fine measures were used in Theorem 15.81.

THEOREM 15.82. Let $h : \mathbb{H} \to \mathbb{R}$ and $H : \mathbb{R} \to \mathbb{R}$ Suppose that there is a sequence of sets $\{E_i\}$ covering $[a,b]$ so that

$$V^*(h,E_i) < \infty$$

for each i. Suppose that

$$\lim_{(I,x)\Rightarrow x} \frac{\Delta H(I)}{h(I,x)} = 1$$

for every $x \in [a,b]$ excepting a set N of points for which

$$V^*(h,N) = V^*(\Delta H, N) = 0.$$

Then h is integrable on $[a,b]$ and H is an indefinite integral.

15.35.3. Kolmogorov equivalence relation from measure relation.

THEOREM 15.83. Let $h,\, k : \mathbb{H} \to \mathbb{R}$ and suppose that for every set $E \subset \mathbb{R}$,

$$V^*(k,E) = V_*(k,E) = V^*(h,E) = V_*(h,E).$$

Suppose that there is a sequence of sets $\{E_i\}$ covering the real line so that

$$V^*(k,E_i) < \infty$$

for each i. Then

$$V^*(|h| - |k|, \mathbb{R}) = 0, \tag{15.2}$$

$$\lim_{(I,x)\Rightarrow x} \left| \frac{h(I,x)}{k(I,x)} \right| = 1 \tag{15.3}$$

for every real number x excepting those in a set N for which $V^*(k,N) = 0$, and

$$\int_a^b (|h| - |k|) = 0 \tag{15.4}$$

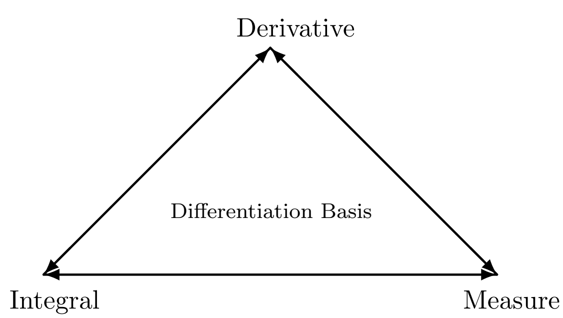
for every compact interval $[a,b]$.

Afterword

One night, after long hours of proofreading and checking details, I found myself in some anxiety over the reception of the manuscript. An earlier version had been met with some confusion (mostly by non native speakers of english) who had thought it a serious proposal that students should be taught this way. They had found nothing amusing or satirical in it. My intention was not obvious evidently. They thought it designed, misguidedly, for students. I aimed at those who continue to teach the Riemann integral, not knowing any better, or (worse) knowing better but continuing in the old traditions. It was aimed, too, at those who write about this better integration theory, but have chosen to present it in most unattractive ways. I hoped to provoke some thought among professional mathematicians as to the nature of the subject they address nearly every day and challenge the conventions of elementary calculus and introductory real analysis texts. (As an author of one of the latter I share the guilt.)

That night I had a vivid dream in which I found myself presenting these ideas to a panel of mathematicians, among whom I recognized Newton himself. He had a broad forehead and an expression of concentrated meditation. His nose was long, thin and prominent. His chin was square and broad. He was of middle height and stout. He had a lively and piercing eye, a comely and gracious aspect, with a fine head of white hair[1].

I reported to him that I considered his calculus program for functions of one real variable now complete. I sketched out the nature of covering relations, how they formed a filter describing a differentiation basis and that the integral, derivative, and measure were all different aspects of one structure, using a diagram that I found had been placed behind me.



I was about to introduce the dual notion of fine covers when I noticed, sitting with a group of Italians that included Dini and Ascoli, the incomparable Giuseppe Vitali. I muttered something about a "ricoprimento di Vitali" and quickly switched

[1]These phrases came to me in the dream, I know not wherefrom.

to the dual basis, calling it the basis of Vitali coverings. These two structures, the differentiation basis and the Vitali basis carry the geometry of the limiting process which is central to the calculus.

There were a few sour looks coming from the direction of the French mathematicians, Lebesgue, Borel, Baire, and especially Denjoy. I recovered by saying that this so far was merely formal and essentially trivial. The interesting task was to develop these formal notions in a constructive manner. Thus I went on to discuss the methods of Lebesgue, Borel and others to construct the values of the measures and the absolute integral. And finally, nodding to Denjoy who was the shortest and most visibly irritated of his companions, I sketched out a transfinite sequence of operations that would construct the value of the nonabsolute integral.

This ended my presentation and there was some feeble applause. I was about to ask for questions when the next speaker rose and I realized that the dream was not to be a short one.

This speaker began by expressing great interest in the material and praising it for its interesting narrative. But, he insisted, this was not the true calculus program of Newton, but merely an interesting cul-de-sac that stressed too much the wrong points.

> "The pointwise approach here is really off the mark. There is little interest in derivatives considered pointwise. The class of functions that are everywhere derivatives is both mysterious and unimportant. The real focus of the calculus should be on the class of all locally Lebesgue integrable functions. The class of integrable functions is larger, but merely a curiosity.
>
> The locally Lebesgue integrable functions are best studied with just the tools of measure theory. Properly speaking, the calculus program of Professor Newton is completed by the development of modern measure theory. It is precisely that body of knowledge that opened up all the other applications such as Fourier series as the eminent Lebesgue amply demonstrated.
>
> Indeed let me make my point by an example. Cantor was much applauded in his time for having solved the following problem: If a series
>
> $$a_0/2 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) = f(x)$$
>
> at every point x, then are the coefficients determined by the function? This question led to sets of uniqueness and attempts to solve the coefficient problem for such representations. Representations? Do you think?

Here he nodded in the direction of Joseph Fourier, who seemed not to notice. Cantor was noticing and oddly distressed.

> "Fourier's program is to find *representations* of functions, indeed. But pointwise representations are a distraction. A pointwise representation cannot be used in any meaningful way. Such questions belong in the same cul-de-sac as all this stuff of Denjoy and inverting derivatives and transfinite totals.

The dream, in typical dream fashion, continued in a series of disjointed scenes. Cantor and Denjoy both stormed from the room. Perron followed, screaming abuse at Denjoy. One member of the audience ran to the board, wrote the equation

$$\int_{-\infty}^{\infty} F'\phi = -\int_{-\infty}^{\infty} F\phi'$$

and said, in a loud voice, "There, there you have it. The derivative means only that! That's the whole calculus program in a nutshell."

There was much yelling. Leibnitz arrived late with a claque of nonstandard analysts and tried noisily to take over the meeting. Some intuitionists had set up an information table and were distributing pamphlets. In the midst of the turmoil I felt Newton's hand on my shoulder. As I turned, he said, "My own opinion in this is ..." and I awoke.

Index