SOME FURTHER TOPICS IN INTEGRATION

PETE L. CLARK

1. Improper integrals

Our entire theory of integrals has been set up for functions defined on intervals of finite length. This does not suffice for all applications: for instance, in probability and statistics one meets *probability density functions*, namely functions $f : \mathbb{R} \to [0, \infty)$ with the property that $\int_a^b f$ is to be interpreted as the probability that $a \leq f(x) \leq b$. But then we want the probability that $-\infty < f(x) < +\infty$ to be equal to 1, in other words we seek to give a meaning to $\int_{-\infty}^{+\infty} f$.

Needless to say, such an integral is not directly compassed by the Riemann/Darboux theory: if by $\int_{-\infty}^{+\infty} f$ we mean "the area under the curve $y = f(x)$ as x ranges over the entire real line," then because of the fact that for certain curves this area will be infinite – e.g. $f(x) \equiv 1$ – we cannot expect this new kind of integral always to exist.

The way out of this, as is familiar from calculus, is to define the integral as a limit over integrals over finite intervals. Then we say that the integral *converges* if the limit exists, and diverges otherwise. This is the – quite simple – idea behind **improper integrals**.

Here is the definition in the simplest case: suppose $f : [a, \infty) \to \mathbb{R}$ is a function which is integrable[1] on every closed subinterval $[a, b]$. Then we define

$$\int_a^\infty f := \lim_{b \to +\infty} \int_a^b f,$$

provided of course that the limit exists. (More specifically, in order to say that the integral is convergent, we require the limit to be a finite real number. If f takes both positive and negative values, it is not necessarily the case that the integral either converges or diverges to $+\pm\infty$: the limit may fail to exist due to oscillation.) We make an entirely similar definition for improper integrals of the form $\int_{-\infty}^b f$, i.e., as $\lim_{a \to -\infty} \int_a^b f$.

However, things get more subtle when we try to define the symbol $\int_{-\infty}^{\infty} f$. Roughly speaking, this should mean the limit of $\int_a^b f$ as $a \to -\infty$ and $b \to +\infty$, but exactly how are we taking these two limits?

Let us indicate how we could go wrong: we might try to define the integral as $\lim_{N \to \infty} \int_{-N}^N f$. Considering the example $f(x) = x$, then since it is an odd function,

[1] By "integrable" we shall for the rest of the course mean Riemann/Darboux integrable.

we have $\int_{-N}^{N} x = 0$ for all N, and according to the proposed definition we would get that $\int_{-\infty}^{\infty} x = 0$. Now there is certainly some sense to this, since the region in question consists of one "infinite triangle" (the region bounded by $y = x, y = 0$ and with $x > 0$) and a congruent infinite triangle ($y = x,\ y = 0$ and with $x < 0$), the one region giving positive area and the other negative area. Thus we are tempted to say that the infinite positive area cancels out the infinite positive area cancels out the infinite negative area, giving zero as the correct answer.

This is not really wise, however, although it is subtle to say why. One answer is that the area of a region is a geometric quantity, so it should be independent of the coordinate system used to define the region. But in the integral $\int_{-\infty}^{\infty} xdx$, make the change of variables $u = x-3$. Then we'd get the integral $\int_{-\infty}^{\infty}(u+3)du$, and construing this again as $\lim_{N\to\infty}\int_{-N}^{N}(u+3)du$, we would now get $\lim_{N\to\infty} 6N = \infty$. In other words, if we took $\lim_{N\to\infty}\int_{-N-3}^{N-3} xdx$, we'd get a totally different answer.

Thus the philosophy of improper integration is the following: if we take several different limits in order to evaluate an improper integral, we require each one of the limits to exist separately; we do not allow the various improprieties to cancel each other out. The following definition enforces this:

Definition: Let $f : \mathbb{R} \to \mathbb{R}$ be a function which is integrable on every interval $[a, b]$. We say f is improperly integrable on $(-\infty, \infty)$ if there exists a number L such that: for all $\epsilon > 0$, there exists an interval $I = I_\epsilon$ such that if $J = [a, b]$ is any interval containing I, then $|\int_a^b f - L| < \epsilon$.

Exercise 60: Show that this definition is equivalent to the following one which is more commonly seen in calculus: Let c be any real number. Then f is improperly integrable on $(-\infty, \infty)$ if $\int_{-\infty}^{c} f$ and $\int_{c}^{+\infty}$ *both* converge in the above sense. Conclude that the latter definition is independent of the choice of c.

There is another kind of improper integral which we have seen before: namely, suppose that $f : [a, b] \to \mathbb{R}$ is Riemann-integrable on $[a, c]$ for all $a < c < b$. We saw that if f is bounded on $[a, b]$, then this implies that it is integrable on all of $[a, b]$. But suppose f is unbounded (imagine it blowing up as $x \to b$). We know that f will *not* be integrable on $[a, b]$ in the ordinary sense, but nevertheless we can ask whether $\lim_{c\to b}\int_a^c f$ exists. If it does, we say again that f is improperly integrable.

We will just give a sequence of exercises that will develop the properties of improper integrals.

Exercise 61: Suppose that f is a *non-negative* function defined on all of $\mathbb{R}$.
a)Explain why $\int_{-\infty}^{\infty} f$, if not convergent, must approach $+\infty$.
b) Show: if $f \leq g$, then $\int_{-\infty}^{\infty} f \leq \int_{-\infty}^{\infty} g$.
c) Show that $\int_{-\infty}^{\infty} e^{-x^2}dx$ converges. Do you happen to know what the value is?

Exercise 62: Explain how every infinite series $\sum_{n=0}^{\infty} a_n$ can be viewed as an improper integral. Hint: the function f that you associate to the series need not be (and probably should not be) continuous.

Exercise 63*: Show that $\int_1^\infty \frac{\sin x}{x}$ is convergent but not absolutely convergent. (Hint for the convergence: try to adapt the proof of the Alternating Series Test.)

Exercise 64: We know, of course, that if a series $\sum_n a_n$ is convergent then $a_n \to 0$. Is it true that if $\int_0^\infty f$ is convergent then $\lim_{x\to\infty} f(x) = 0$?

The analogy between infinite series and improper integrals is especially strong when we have an integral of the form $\int_1^\infty f$, with f monotonically decreasing. In fact we get an important convergence test:

Theorem 1. *(Integral Test) Let $f : [1,\infty) \to [0,\infty)$ be a positive, monotonically decreasing function, and consider the series $\sum_{n=1}^\infty a_n$, where $a_n = f(n)$. Then*

$$\sum_{n=2}^\infty a_n \leq \int_1^\infty f \leq \sum_{n=1}^\infty a_n.$$

In particular, the convergence of the series is equivalent to the convergence of the integral.

Exercise 65: Prove the Integral Test. (Hint: view $\{1, 2, 3, \ldots\}$ as a partition of $[1,\infty)$ and draw a picture of the upper and lower sums of f with respect to this partition.)

Exercise 66: Use the integral test to give a new proof that the p-series $\sum_{n=1}^\infty \frac{1}{n^p}$ converges if and only if $p > 1$.

Exercise 67: In the notation of the Integral Test that $\int_1^\infty f = +\infty$. Show that $\sum_{n=1}^N a_n \sim \int_1^N f$. (Recall that this means: writing $LHS(N)$ and $RHS(N)$ for the left and right-hand sides as functions of N, we have $\lim_{N\to\infty} \frac{LHS(N)}{RHS(N)} = 1$.)

Exercise 67 says, for example, that the partial sum $\sum_{n=1}^N \frac{1}{n}$ grows as $\ln N$, which we stated without proof (but gave supporting computations for) at the beginning of the course.

2. Lebesgue's Theorem

The goal of this optional section is to give an idea of the answer to the following question: just which functions $f : [a,b] \to \mathbb{R}$ are Riemann-integrable? Recall that up to this point we have shown the following:

- Every Riemann-integrable function is bounded.
- Every continuous function is Riemann-integrable.
- Every bounded function with only finitely many discontinuities is Riemann-integrable.
- Every monotone function is Riemann-integrable.

In other words, continuous functions are integrable, and integrable functions are bounded, so the question somehow boils down to: how discontinuous can a bounded function be and still be integrable? Recall from Exercise 49 that for any countable subset $S \subset [a, b]$, there exists an increasing function $f : [a, b] \to \mathbb{R}$ which is discontinuous precisely at the elements of S. For instance we could take S to be the set of rational numbers in the interval $[a, b]$, and we would get an (in fact strictly) increasing integrable function which is discontinuous at every rational point and continuous at every irrational point. Thus functions can have "many" discontinuities and still be integrable.

The following definition and result give a further enlargement of the class of continuous functions.

Definition: A function $f : [a, b] \to \mathbb{R}$ is said to have a **simple discontinuity** at a point $c \in [a, b]$ if f is discontinuous at x but $\lim_{x \to c^-} f(x) = L_g$ and $\lim_{x \to c^+} f(x) = L_d$ both exist. (What this means is that at least one of L_g and L_d is not equal to $f(c)$; for instance, they may not be equal to each other.)

Linguistic Exercise: What is the explanation for the subscripted "g" and subscripted "d" in the definition?

A function $f : [a, b] \to \mathbb{R}$ is said to be **regulated** if its only discontinuities are simple discontinuities. The basic example of regulated functions are **step functions**: a step function is a function $f : [a, b] \to \mathbb{R}$ for which there exists a finite set of points $S \subset [a, b]$ such that $f|_{[a,b] \setminus S}$ is locally constant. In other words, there exists a partition P of $[a, b]$ such that f is constant on the interior of each subinterval of the partition. It is easy to see that a step function is bounded with only finitely many discontinuities (the possible discontinuities are contained in the finite set S of the definition). An arbitrary regulated function is much more complicated than a step function:

Exercise 68*: Show that monotone functions are regulated.[2]

Theorem 2. *Every regulated function $f : [a, b] \to \mathbb{R}$ is Riemann-integrable.*

Remark: It would be possible to give a direct proof of this result at this point, but it would take more time than we are willing to spare. What is in fact the case is that regulated functions are "well-approximated" by step functions in a certain sense that we will discuss later on in the course, and we will also have a result saying that a function which is "well-approximated" by a sequence of integrable functions is itself integrable, which will prove the result.

In fact regulated functions are, like monotone functions, "not very discontinuous" in the following sense:

Theorem 3. *The set of discontinuities of a regulated function $f : [a, b] \to \mathbb{R}$ is countable.*

Exercise 69*: Prove Theorem 3.[3]

[2]This is not an especially difficult exercise. However, since this is an optional section, all exercises in it must also be optional.

[3]This exercise really does deserve a $*$: it's not easy.

Based on these results, a reasonable guess might be that a bounded function is Riemann integrable if and only if its set of discontinuities is countable. On the other hand, it is plausible that the condition should somehow involve subintervals of $[a, b]$, because this is how the integrals (both Riemann and Darboux) are defined. Suppose that $|f(x)| \leq M$ for all $x \in [a, b]$, and suppose also – just for the sake of argument – that we knew that all of the discontinuities of f were contained in some subinterval $[c, d] \subset [a, b]$ with "very small" length δ. Then, when we look at the Darboux condition for integrability, we find that the upper sum of f on the subinterval $[c, d]$ is at most $M\delta$ and the lower sum is at most $-M\delta$. Thus, $\overline{\int}_c^d f - \underline{\int}_c^d f \leq 2M\delta$. Since f is continuous on the complement of $[c, d]$, it is integrable on both $[a, c]$ and $[d, b]$, so the upper and lower sums over both of these intervals can, by the Darboux integrability criterion, be made to differ by an arbitrarily small amount, so that, for any $\epsilon > 0$ we can find a a partition P of $[a, b]$ such that

$$\omega(f, P) = U(f, P) - L(f, P) \leq 2M\delta + \epsilon,$$

hence $\overline{\int}_a^b f - \underline{\int}_a^b f \leq 2M\delta$.

The same argument works if we can confine the discontinuities of f to a union of finitely many intervals $[c_1, d_1] \cup [c_2, d_2] \cup \ldots \cup [c_N, d_N]$ of lengths, say, $\delta_1, \ldots, \delta_N$. We get in this case that $\overline{\int}_a^b f - \int_a^b f \leq 2M(\sum_{i=1}^N \delta_i)$.

Now the key idea is the following: if for our fixed function f (with fixed bound of M), we could, for any $\epsilon > 0$, confine the discontinuities to a finite union of intervals as above with *total length* at most ϵ, then the above argument gives that the difference between the upper and lower integral is at most $2M\epsilon$, and since ϵ was arbitrary, we get that the difference can be made arbitrarily small. In other words, any bounded function whose discontinuities can be covered by finitely many intervals whose total length is arbitrarily small is necessarily integrable.

There is a name for a subset S of $[a, b]$ which can, for any $\epsilon > 0$, be covered by finitely many subintervals of total length at most ϵ: such a set is said to have **content zero**. Unfortunately the condition that a bounded function should have a set of discontinuities of content zero is not quite the correct one: not every countably infinite subset of $[a, b]$ has content zero: for instance, if you try to cover the set of rational numbers in $[0, 1]$ by finitely many subintervals, you will find that the minimum possible total length of the subintervals is 1.

On the other hand, it is not hard to see that any countable subset $S \subset [a, b]$ can, for any $\epsilon > 0$, be covered by a sequence of subintervals $I_i = ([c_i, d_i])$ such that $\sum_i d_i - c_i < \epsilon$ provided we allow the sequence to be **infinite**. Namely, we write $S = (x_i)_{i=1}^\infty$ and take I_i to be $(x_i - \frac{\epsilon}{2^{i+1}}, x_i + \frac{\epsilon}{2^{i+1}})$, i.e., the open interval centered at x_i and of length $\frac{\epsilon}{2^i}$. Thus the total length of the intervals is just

$$\sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon \sum_{i=1}^{\infty} 2^{-i} = \epsilon \cdot 1 = \epsilon.$$

This suggests the following definition: A subset $S \subset [a, b]$ has **measure zero** if: for every $\epsilon > 0$, there exists a sequence of subintervals $I_i \subset [a, b]$ such that $S \subset \bigcup_{i=1}^\infty I_i$

– this just means that every element of S is contained in at least one of the subintervals – and such that $\sum_{i=1}^{\infty} \ell(I_i) < \epsilon$, where for an interval I, $\ell(I)$ just denotes its length.

So our argument directly preceding the definition showed: every countable subset S of $[a, b]$ has measure zero. That this is the right definition is spectacularly confirmed by the following result, which must be reckoned as one of the great mathematical theorems of all time.

Theorem 4. *(Lebesgue's criterion for Riemann-integrability) Let $f : [a, b] \to \mathbb{R}$ be any function. The following are equivalent:*
a) f is Riemann(/Darboux) integrable.
b) f is bounded and the set $S \subset [a, b]$ of discontinuities of f has measure zero.

We should remark that although the generalization from finitely many to infinitely many intervals is plausible given our hope that functions which are discontinuous at only countably many points are integrable, this apparently minor change makes the proof of Lebesgue's Theorem significantly more complicated than the analogous result for "content zero." We do not discuss the proof here; if you are interested, I can recommend some places to look for it.

So in particular we now know that any bounded function which is discontinuous at only countably many points is integrable, because countable sets have measure zero. The converse is not true: a set can be uncountable but still have measure zero. The traditional (and important) example is the **Cantor set**. The Cantor set is formed as follows: it is $\int_{i=0}^{\infty} C_i$, where $C_0 = [0, 1]$ and C_{i+1} is obtained from C_i by a simple process: namely, C_1 is obtained from C_0 just by removing the open interval $(\frac{1}{3}, \frac{2}{3})$, i.e., $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. To get C_2 from C_1 we remove the (open) middle third of each of the two subintervals in C_1, getting $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. We then continue in this way: in general C_i is a union of 2^i closed intervals each of length 3^{-i}. By definition, the Cantor set C is the set of points that are left after we repeat this process infinitely many times, which sounds mysterious but really just means points that are in C_i for all i.

Exercise 70*: a) Show that if S_1 and S_2 are two subsets of $[a, b]$ each of measure zero, then their union $S_1 \cup S_2$ has measure zero. In fact, show that if $\{S_i\}_{i=1}^{\infty}$ is any (countable!) sequence of subsets of $[a, b]$ such that each S_i has meaure zero, then their union $\bigcup_{i=1}^{\infty} S_i$ has measure zero.
b) Show that the first statement of part a) is still true for content zero, but not the second. (This is the problem with content zero.)

Exercise 71*:
a) Show that the Cantor set C has measure zero. In fact, show that it has content zero. (Hint: the C_i's themselves will do the trick.)
b) Show that the Cantor set is uncountable by characterizing it as follows: it is the set of all real numbers in $[0, 1]$ whose base three expansion $0.a_1a_2a_3 \ldots a_n \ldots$ does not contain the digit "1" but only the digits "0" and "2". (Recall that there is a slight ambiguity in decimal expansions which is still there in base three expansions; for instance $\frac{1}{3} = 0.1000 \ldots = 0.022222 \ldots$. Which expansion must be taken to make the statement true?)

c) Deduce that the Cantor set is uncountable. (Hint: relate the Cantor set to binary expansions of arbitrary real numbers.)

Exercise 72*: Construct a bounded function $f : [0,1] \to \mathbb{R}$ whose set of discontinuities is precisely the Cantor set C.

Remark: In the late eighties and early nineties, almost every popular math book one could find was about chaos and/or fractals (the reason being, I think, that then-recent popularization of computers made it possible to try to interest people in mathematics by showing them that mathematics could be used to generate computer printouts that were striking and even psychedelic). The Cantor set is an example of a fractal: one of the ways to interpret this is to remark that it is **self-similar**: the Cantor set can be viewed as a union of two copies of itself, each scaled down by a fraction of $\frac{1}{3}$. Now of course the unit interval $[0,1]$ is also self-similar, but it is the union of *three* copies of itself, each scaled down by a factor of 3: this equality of the number of pieces and the scale factor is characteristic of one-dimensional objects. Compare with the unit square in the plane, which is the union of $9 = 3^2$ copies of itself, each scaled down by a factor of 3. The fact that it is two-dimensional can be captured in the exponent of the following equation: (number of copies) = (scale factor)^2. Similarly, for the unit cube one would find that (number of copies) = (scale factor)^3, so in general it seems reasonable to say that a self-similar object has **dimension d** if the equation (number of copies) = (scale factor)^d holds. For the Cantor set, this yields the value $d = \log_3 2 = \frac{\log 2}{\log 3}$, which is, of course, strictly between zero and one. ("Fractal" came about as an abbreviation for "fractional dimension" although, ironically, $\frac{\log 2}{\log 3}$ is not a fraction, i.e., not rational: rather it is a transcendental number.)

Now it is almost 15 years later and the math books full of pretty pictures have all but vanished from the bookstores. Nevertheless, the Cantor set is a supremely important object in modern analysis (and geometry and even number theory). By contemplating it one gets a sense of the richness of even the subsets of the real line that arise when one studies real functions (and in fact this increased attention to the subsets themselves is characteristic of twentieth-century analysis).

You might want to go back and see how many of the other integrability theorems are implied by Lebesgue's Theorem: it really is "the answer." In other words, instead of assuming that something is integrable and showing that something else is integrable, you can assume that something is bounded and continuous except on a set of measure zero and show that something else is bounded and continuous on a set of measure zero. Since the notion of sets of measure zero is a very robust one – cf. Exercise 70 – this can be quite easy. For instance, it follows from Lebesgue's Theorem and Exercise 70 that if f_1 and f_2 are integrable, so is $f_1 f_2$. Can you see how?