Lecture 6-2: Riemann integration, concluded

June 2, 2025

I conclude the course by stating an important criterion for a bounded function f on a closed bounded interval $[a, b]$ to be integrable. Roughly speaking, this holds if and only if f is not too badly discontinuous on $[a, b]$..

To be more precise, I first need to define a property of subsets of $\mathbb{R}$.

Definition 7.6.1, p. 240

A set $A \subset \mathbb{R}$ has *measure 0* if given any $\epsilon > 0$ there is a countable collection $\{O_n = (a_n, b_n)\}$ of open intervals such that $A \subset \cup O_n$ and $\sum_{n=1}^{\infty}(b_n - a_n) < \epsilon$.

Note that the difference $b_n - a_n$ is a natural measure of the length of the interval (a_n, b_n); thus this definition says that A has total length 0 in some sense, For example, any countable set $\{x_i, i \in \mathbb{N}\}$ has measure 0, since we can take $O_i = (x_i - \epsilon/3^i, x_i + \epsilon/3^i)$ in that case. More interestingly, as I observed when I constructed the Cantor set C, this set is uncountable but also has measure 0, since for any n it is contained in the union of 2^n intervals, each of length $1/3^n$, and we have $2^n/3^n < \epsilon$ if n is large enough.

More strongly, we say that $A \subset \mathbb{R}$ has *content 0* if given any $\epsilon > 0$ there is a *finite* collection $(a_1, b_1), \ldots, (a_n, b_n)$ of open intervals such that A is contained in the union U of this collection and $\sum_{i=1}^{n}(b_i - a_i) < \epsilon$. Clearly any finite set has content 0; but the intersection $S = \mathbb{Q} \cap [0, 1]$ has measure 0 but not content 0. For if there were a finite set $(a_1, b_1), \ldots, (a_n, b_n)$ whose union contains S and whose total length $\sum(b_i - a_i)$ is less than $1/2$, then the union of the *closed* intervals $[a_i, b_i]$ would also contain S, whence it would contain the unit interval $[0, 1]$ by the density of S in $[0, 1]$; but it is easy to see check that no finite union of closed intervals of total length less than $b - a$ can contain a closed interval $[a, b]$. In general, however, a *closed* set has measure 0 if and only if it has content 0.

The criterion for integrability is then

Theorem 7.6.5, p. 242

The (bounded) function f is integrable on $[a, b]$ if and only if the set D of points $x \in [a, b]$ such that f is discontinuous at x has measure 0.

A function f satisfying this criterion is said to be continuous *almost everywhere*; more generally, a property of real numbers which holds except on a set of measure 0 is said to hold almost everywhere. A proof of the theorem is given in the text on pp. 242-3; it is set up as a sequence of exercises and uses some additional facts about compact sets from Chapter 3. One first writes the set D as the countable union $\sum D_n$, where D_n consists of the points $x \in D$ for which the oscillation $\omega_x f > 1/n$; recall that $\omega_x f$ is defined as the limit of $M_n - m_n$ as $n \to \infty$, where M_n is the supremum of f on $[-1/n, x + 1/n]$ and m_n is the infimum of f on the same interval.

Then it turns out that D has measure 0 if and only if each D_n has content 0. If this holds, then by arguing as in the proof that Thomae's function is integrable one produces for each $\epsilon > 0$ a partition P of $[a, b]$ such that $U(f, P) - L(f, P) < \epsilon$, whence f is integrable. Conversely, if f is integrable and the partition P is chosen to make $U(f, P) - L(f, P) < \epsilon/n$, then one can show directly that each D_n has content 0, whence D has measure 0.

I conclude with an example of a continuous increasing function f on the unit interval $[0, 1]$ such that $f'(x) = 0$ almost everywhere, $f(0) = 0$, and $f(1) = 1$. The derivative $f'(x)$ is Riemann integrable, being continuous almost everywhere, but its integral from 0 to 1 is 0 rather thatn $1 = f(1) - f(0)$. Recall the definition of the Cantor set C from p. 86 of the text: we have $C = \cap_{n=0}^{\infty} C_n$, where C_n is the disjoint union of 2^n subintervals of $C_0 = [0, 1]$, each of length 3^{-n}. We construct the subintervals of C_n by removing the open middle third of each subinterval of C_{n-1}, thereby replacing it by the union of two subintervals, each $1/3$ the length of the original subinterval.

Define the function g_n on C_0 by $g_n(x) = (3/2)^n$ if $x \in C_n, g_n(x) = 0$ otherwise, and set $f_n(x) = \int_0^x g_n(t)\,dt$. Then $f_n(0) = 0, f_n(1) = 1$, and each f_n is a (weakly) increasing function which is constant on each subinterval in the complement of C_n. If I is one of the 2^n intervals whose union is C_n, then $\int_I g_n(t)\,dt = \int_I g_{n+1}(t)\,dt = 2^{-n}$. It follows that $f_{n+1}(x) = f_n(x)$ if $x \notin C_n$ and $|f_{n+1}(x) - f_n(x)| < \int_I |g_n - g_{n+1}|\,dt < 2^{-n+1}$ if $x \in E_n$. Hence the sequence f_n converges uniformly to a continuous increasing function f with $f(0) = 0, f(1) = 1$, and $f'(x) = 0$ for all $x \notin C$ (since any such x fails to lie in C_n for n sufficiently large). Since C has measure 0, this function f has the desired properties.

Such a function f is called *singular*. There is a more general notion of the measure $m(A)$ of a subset A of $\mathbb{R}$; this need not be 0, and indeed $m[a, b] = b - a$ for every closed bounded interval $[a, b]$. Then $m(A)$ is not defined for all subsets A of $\mathbb{R}$, but it is defined for many such subsets, including all open and closed subsets and all countable unions or intersections of such sets. A real-valued function f on $\mathbb{R}$ is said to be *measurable* if $V = f^{-1}(U)$ is measurable (i.e. $m(V)$ is defined) whenever U is. A generalization of the Riemann integral called the *Lebesgue integral* is then defined for all bounded measurable functions f on $[a, b]$. The Lebesgue integral of any function f equal to 0 almost everywhere is 0.

There is a brief discussion of the Lebesgue integral on p. 247 of the text, followed in Chapter 8 by the definition of further generalization of the Riemann integral that goes beyond even the Lebesgue integral.

The remaining lectures this week will be devoted to review for the final.