REVIEW OF INTEGRATION

KEITH CONRAD

Let (X, μ) be a measure space.

Definition 1. A function $f\colon X \to \mathbf{C}$ is called *measurable* when $f^{-1}(U)$ is a measurable set for every Borel set $U \subset \mathbf{C}$.

Example 2. If X is a topological space, equipped with the σ-algebra of Borel sets, then any continuous function $X \to \mathbf{C}$ is measurable.

Theorem 3. *If $f_n\colon X \to \mathbf{C}$ is a sequence of measurable functions which converges pointwise to f, then f is a measurable function.*

In particular, since continuity implies measurability (on topological spaces) a pointwise limit of continuous functions is not necessarily continuous but is always measurable.

Example 4. Let $X = [0,1]$ and $f_n(x) = x^n$. The f_n's are continuous and converge pointwise, but the limit function is not continuous. It is 0 on $[0,1)$ and 1 at $x = 1$. This pointwise limit is measurable.

Definition 5. For a subset $A \subset X$, let ξ_A be the characteristic function of A:

$$\xi_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Definition 6. A *simple map* $X \to \mathbf{C}$ is a finite sum $\sum_{i=1}^n c_i \xi_{A_i}$, with $c_i \in \mathbf{C}$.

Definition 7. A *step map* $X \to \mathbf{C}$ is a finite sum $\sum_{i=1}^n c_i \xi_{A_i}$ where $c_i \in \mathbf{C}$ and $\mu(A_i) < \infty$ for all i.

Simple maps and step maps each take only a finite number of values. The difference between a simple map and a step map is that a step map is required to take each nonzero value on a set of finite measure.

Example 8. If $\mu(A) = \infty$ then ξ_A is *not* a step map.

When f is a simple map or a step map, $|f|$ is also simple map or step map with the same expression as f except each c_i is replaced by $|c_i|$. In the expression of a simple map or step map as a linear combination of characteristic functions ξ_{A_i}, we can always break up the supports of the characteristic functions so the A_i's are disjoint.

Integration with respect to μ is defined on step maps, not simple maps: for a step map $f = \sum_{i=1}^n c_i \xi_{A_i}$, define

$$\int_X f \, \mathrm{d}\mu = \sum_{i=1}^n c_i \mu(A_i).$$

From the properties of measures, this sum is well-defined (independent of the way f is written as a linear combination of characteristic functions of finite-measure subsets of X) and integration is linear on step maps.

The following two theorems illustrate the difference between pointwise limits of simple maps and step maps.

Theorem 9. *For a function $f\colon X \to \mathbf{C}$, the following are equivalent:*

1. *f is a pointwise limit of simple maps,*
2. *f is a measurable function.*

Proof. See [1, pp. 118,119]. □

Notice the conditions in Theorem 9 don't involve the measure μ at all.

Theorem 10. *For a function $f\colon X \to \mathbf{C}$, the following are equivalent:*

1. *f is a pointwise limit of step maps almost everywhere,*
2. *f is a measurable function almost everywhere and vanishes outside a σ-finite set.*

Proof. See [1, pp. 124], taking E there to be $\mathbf{C}$. □

Notice the three concepts step map, almost everywhere, and σ-finite in Theorem 10 all depend on the measure μ. Call a function satisfying the conditions of Theorem 10 *μ-measurable.*

The point of Theorem 10 is that the first property is what leads to the possibility f could be integrable (it might not be!), while the second property is something you might be able to verify in practice with an individual function. When X itself is σ-finite (like $\mathbf{R}^n$ with Lesbesgue measure), the second property in Theorem 10 just says f is measurable almost everywhere, so in this case the distinction between pointwise limits of simple maps and step maps is negligible.

For step maps f and g, their difference $f - g$ is a step map and we call the number $\int_X |f-g| \, \mathrm{d}\mu$ their L^1-distance. We want to complete the space of step maps in the L^1-sense.

Here is the basic result that gets integration working, dubbed the "fundamental lemma of integration" in [1].

Theorem 11. *If $\{f_n\}$ is an L^1-Cauchy sequence of step maps, it converges pointwise almost everywhere and converges uniformly off a set of arbitrarily small measure.*

Proof. See [1, p. 129]. □

Definition 12. Set $L^1(\mu) = L^1(\mu, \mathbf{C})$ to be the functions $f\colon X \to \mathbf{C}$ which are the pointwise limit almost everywhere of an L^1-Cauchy sequence of step maps. The function f is μ-measurable by Theorem 10.

Theorem 13. *If $f \in L^1(\mu)$ is the pointwise limit almost everywhere of the L^1-Cauchy sequence of step maps $\{f_n\}$, then the sequence of integrals*

$$\int_X f_n \, \mathrm{d}\mu$$

converges in $\mathbf{C}$ and the limit is independent of the choice of $\{f_n\}$.

Proof. See [1, p. 130]. □

The limit of integrals of step maps in Theorem 13, which depends only on f and not on the f_n's, is denoted $\int_X f \, \mathrm{d}\mu$. Integration with respect to μ is a linear map $L^1(\mu) \to \mathbf{C}$.

When f is a pointwise limit almost everywhere of an L^1-Cauchy sequence of step maps, the function $|f|$ is also such a limit and $|\int_X f\,\mathrm{d}\mu| \le \int_X |f|\,\mathrm{d}\mu$. One sets

$$|f|_1 = \int_X |f|\,\mathrm{d}\mu$$

and $|f-g|_1$ has the properties of a metric except it might be zero without $f = g$ everywhere.

Example 14. Let $X = [0,1]$ and write down a sequence of closed intervals in $[0,1]$ as follows:

$$[0,1], [0,1/2], [1/2,1], [0,1/3], [1/3,2/3], [2/3,1], [0,1/4], [1/4,2/4], [2/4,3/4], [3/4,1] \ldots$$

and let f_n be the characteristic function of the nth such interval. Then, using Lebesgue measure $\mathrm{d}x$, $\int_X |f_n(x)|\,\mathrm{d}x = \int_X f_n(x)\,\mathrm{d}x \to 0$, so f_n is L^1-convergent to 0 (draw a picture so you can see what's going on!), but for no x does the sequence of numbers $f_n(x)$ tend to 0: $f_n(x)$ oscillates from 0 to 1 and back as $n \to \infty$, being 0 more and more often as n grows). So L^1-convergence of a sequence of functions does not have to imply pointwise convergence of the sequence of functions *anywhere.*

Theorem 15. *If $f \in L^1(\mu)$ is the L^1-limit of a sequence of functions $f_n \in L^1(\mu)$ then f is the pointwise limit almost everywhere of some subsequence of the f_n's.*

Proof. Replacing $\{f_n\}$ by a subsequence, we may assume $|f_n - f_{n+1}|_1 \le 1/2^n$ for all n. Then it can be proved such a sequence converges pointwise almost everywhere. See [1, p. 138]. The bound $1/2^n$ can be replaced by c^n for any fixed $c \in (0,1)$. □

Example 16. In Example 14, the subsequence of f_n's which are characteristic functions of the intervals $[0,1/m]$ for $m \to \infty$ tend pointwise to 0 everywhere except at $x = 0$.

Corollary 17. *If f is μ-measurable and $\int_X |f|\,\mathrm{d}\mu = 0$ then $f = 0$ almost everywhere.*

Proof. Let $f_n = 0$ for all n, so f is an L^1-limit of the sequence $\{f_n\}$ for all n. Thus a subsequence of the f_n's converges pointwise to f almost everywhere, so f is 0 almost everywhere. □

Theorem 15 tells us that L^1-convergence implies pointwise convergence if we pass to a suitable subsequence. The following theorem is something of a converse, giving conditions under which pointwise convergence implies L^1-convergence. It is the most important basic result about limits of integrals.

Theorem 18 (Dominated Convergence Theorem)**.** *If $f_n \in L^1(\mu)$, $f_n \to f$ pointwise almost everywhere, and $|f_n(x)| \le g(x)$ almost everywhere for some $g \in L^1(\mu, \mathbf{R})$, then $f \in L^1(\mu)$ and $|f_n - f|_1 \to 0$, so in particular $\int_X f\,\mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n\,\mathrm{d}\mu$.*

Proof. See [1, p. 141]. □

Example 19. We will use the dominated convergence theorem to prove the intuitively reasonable result that integrating an integrable function over small subsets of X should give small values:

$$f \in L^1(\mu) \Longrightarrow \lim_{\mu(A)\to 0} \int_A f\,\mathrm{d}\mu = 0.$$

That is, for each $\varepsilon > 0$ we will show there is a $\delta > 0$ such that any A with $\mu(A) < \delta$ has $|\int_A f\,\mathrm{d}\mu| < \varepsilon$. (Here f is fixed.)

The proof is by contradiction. Assume there is no such δ, so we can find a sequence of subsets $Y_n \subset X$ with $\mu(Y_n) \to 0$ and $|\int_{Y_n} f \, \mathrm{d}\mu| \geq \varepsilon$. For concreteness, we can select the Y_n's so that $\mu(Y_n) \leq 1/2^n$. Let $g_n = f\xi_{Y_n}$, *i.e.*, g_n is f on Y_n and 0 elsewhere. Set

$$Z = \{x \in X : x \text{ is in } Y_n \text{ for infinitely many } n\}.$$

For any n, $Z \subset Y_n \cup Y_{n+1} \cup Y_{n+2} \cup \cdots$, so $\mu(Z) \leq \mu(Y_n) + \mu(Y_{n+1}) + \cdots = 1/2^{n-1}$. Thus Z has measure 0. When $x \notin Z$, $x \notin Y_n$ for all $n \gg 0$, so $g_n(x) = 0$. Hence $g_n \to 0$ almost everywhere on X (namely on $X - Z$). Since $|g_n(x)| \leq |f(x)|$ for all x and $f \in L^1(\mu)$, by the dominated convergence theorem we have $\int_X g_n \, \mathrm{d}\mu \to \int_X 0 \, \mathrm{d}\mu = 0$. But $\int_X g_n \, \mathrm{d}\mu = \int_{Y_n} f \, \mathrm{d}\mu$, so $|\int_X g_n \, \mathrm{d}\mu| \geq \varepsilon$ for all n. Taking n large enough, we get a contradiction.

Remark 20. By Example 19, if $f \in L^1(\mu)$ and $A \subset X$ satisfies $\mu(A) = 0$ then $\int_A f \, \mathrm{d}\mu = 0$. It follows (with some work) that the set-function $A \mapsto \int_A f \, \mathrm{d}\mu$ (fixed f, varying A) is a finite measure on X. The converse is also true if μ is σ-finite: if ν is a finite Borel measure on X such that $\nu(A) = 0$ whenever $\mu(A) = 0$ then there is an $f \in L^1(\mu)$ such that $\nu(A) = \int_A f \, \mathrm{d}\mu$, and such an f is unique as an L^1-function. This is a special case of the Radon–Nikodym theorem. (Without σ-finiteness this converse can fail. Try $X = [0,1]$ with μ being counting measure and ν being Lebesgue measure.)

Example 21. The dominated convergence theorem is only generally valid for sequences of functions, not for nets of functions (*i.e.*, not for functions indexed by directed sets other than the positive integers). For example, partially order the finite subsets $A \subset [0,1]$ by inclusion and let f_A be the characteristic function of A. Then $0 \leq f_A \leq 1$ and the net $\{f_A\}$ tends to the integrable function 1 pointwise. However, $\int_0^1 f_A(x) \, \mathrm{d}x = 0$ for all finite A while $\int_0^1 1 \, \mathrm{d}x = 1$. There are situations in real analysis where one wants to deal with a net of integrable functions rather than a sequence of integrable functions (*e.g.*, integrals of the Newton quotients $(f(x+h) - f(x))/h$ as $h \to 0$), and a special argument is needed to justify applying the dominated convergence theorem to such a family of functions.

Corollary 22. *For a μ-measurable function f, $f \in L^1(\mu)$ if and only if $|f| \in L^1(\mu, \mathbf{R})$.*

Proof. See [1, p. 142]. □

Note that Corollary 22 is the *first* time that something special has appeared for positive functions. Most treatments of integration place an unduly heavy emphasis on positive functions right at the start, leading to an *unnnatural* method of defining the integral, first defining it on nonnegative (measurable) functions, then on real functions using positive and negative parts, and then on complex functions using real and imaginary parts. Here we have described the passage from step maps to general integrable functions without starting from positive functions.

Corollary 23. *If $f \in L^1(\mu)$ and $\varphi \colon X \to \mathbf{C}$ is bounded and measurable then $\varphi f \in L^1(\mu)$.*

Proof. See [1, p. 143]. □

Example 24. We will use Corollary 23 and the dominated convergence theorem to prove a "negative" result. Let $C[0,1]$ be the space of continuous functions $[0,1] \to \mathbf{C}$. For $f \in C[0,1]$ and $g \in L^1([0,1])$, the product fg is integrable since f is bounded, so $\int_0^1 fg \, \mathrm{d}x$ makes sense. (Note f and g here play the respective roles of φ and f from Corollary 23.) In particular, sending f to $\int_0^1 fg \, \mathrm{d}x$ is an example of a linear function $C[0,1] \to \mathbf{C}$. So integration against the product of a fixed function in $L^1([0,1])$ provides many examples of continuous linear maps $C[0,1] \to \mathbf{C}$, where $C[0,1]$ is given the sup-norm.

Not all continuous linear maps $C[0,1] \to \mathbf{C}$ have the form $f \mapsto \int_0^1 fg \, \mathrm{d}x$. Specifically, $f \mapsto f(0)$ is such a map and we will show there is no $g \in L^1([0,1])$ such that

$$f(0) = \int_0^1 fg \, \mathrm{d}x$$

for all $f \in C[0,1]$. That is, evaluation at 0 is not the same as multiplying by a fixed function and then integrating on $[0,1]$.

Assume otherwise: g exists. Let $f_n \in C[0,1]$ be the piecewise linear function whose graph is the line segment between the points $(0,1)$ and $(1/n, 0)$ and is 0 for $1/n \leq x \leq 1$. Then $|f_n(x)| \leq 1$ for all x, so $|f_n(x)g(x)| \leq |g(x)|$. Thus $|f_n g| \leq |g|$ pointwise. For each $x \neq 0$, $f_n(x)g(x) \to 0$ as $n \to \infty$ simply because $g(x)$ is staying put while (for fixed x) $f_n(x) = 0$ for $n \gg 0$. Therefore we can apply the dominated convergence theorem (the $|f_n g|$'s are dominated by g, assumed to be in L^1) to say $\int_0^1 f_n g \, \mathrm{d}x \to 0$ as $n \to \infty$. But, since each f_n is in $C[0,1]$, by hypothesis $\int_0^1 f_n g \, \mathrm{d}x = f_n(0) = 1$ for all n.

Measures and their integrals lead to linear maps on spaces of functions. We can also go the other way in important cases, starting with suitable linear maps and recovering a measure, as follows.

Theorem 25 (Riesz representation theorem)**.** *Let X be a locally compact Hausdorff space and let $C_c(X)$ be the continuous functions $X \to \mathbf{C}$ with compact support. Any linear map $L\colon C_c(X) \to \mathbf{C}$ such that $L(f) \geq 0$ when $f(x) \geq 0$ for all x has the form $L(f) = \int_X f \, \mathrm{d}\mu$ for a unique Borel measure μ on X with the following properties:*

1. *$\mu(K) < \infty$ for compact K,*
2. *for any Borel set A, $\mu(A) = \inf_{U \supset A} \mu(U)$ over the open U containing A,*
3. *for any open or σ-finite Borel set A, $\mu(A) = \sup_{K \subset A} \mu(K)$ over the compact K contained in A.*

Example 26. Take $X = [0,1]$, so Theorem 25 says any linear function $C[0,1] \to \mathbf{C}$ which sends nonnegative functions to nonnegative numbers is integration against a unique Borel measure. Consider the functional $L(f) = f(0)$ from Example 24. The corresponding measure is the Dirac measure δ_0 at 0: $\delta_0(A) = 1$ if $0 \in A$ and $\delta_0(A) = 0$ if $0 \notin A$, so $\int_0^1 f \, \mathrm{d}\delta_0 = f(0)$ for continuous f. The Dirac measure at 0 is not of the form $g \, \mathrm{d}x$ for any $g \in L^1([0,1])$ since $\int_A g \, \mathrm{d}x = 0$ when A has Lebesgue measure 0 (Example 19) but δ_0 doesn't have this property for all A (take $A = \{0\}$!). This gives a different solution to Example 24.

If we consider linear maps $L\colon C_c(X) \to \mathbf{C}$ which are continuous (with respect to the sup-norm on $C_c(X)$) rather than positive, then there is an analogue of the Riesz representation theorem: such L have the form $L(f) = \int_X f \, \mathrm{d}\mu$ where μ is a complex Borel measure. These two representation theorems for linear maps on $C_c(X)$ have an overlap, but neither is a special case of the other in general since positive (*i.e.*, ordinary) measures are not always complex measures: the former can assign the value ∞ to subsets while this is ruled out for complex measures.

REFERENCES

[1] S. Lang, "Real and Functional Analysis," 3rd ed., Springer–Verlag, New York, 1993.