I. *Almost-Everywhere Convergence . . . Done Right*

Joel H. Shapiro

October 13, 2017

These notes develop the modern proof—via the Hardy-Littlewood Maximal Theorem—of the Lebesgue Differentiation Theorem. They close with a discussion of the connection between Lebesgue's theorem and a famous 1926 paper [1] of Banach that lays out the "cosmic truth" behind a.e. convergence.

1 *The Lebesgue Differentiation Theorem*

SUPPOSE the real-valued function f is continuous on the real line, and has compact support, in short: $f \in C_c(\mathbb{R})$. Then the *First Fundamental Theorem of Integral Calculus* tells us that

$$(1) \qquad \frac{d}{dx}\int_{-\infty}^{x} f(t)\,dt = f(x) \qquad \forall\ x \in \mathbb{R}.$$

If, more generally, we take f to be measurable and (absolutely) integral on $\mathbb{R}$ with respect to Lebesgue measure, in short: $f \in L^1(\mathbb{R})$, then equation (1) still holds, but now only for almost-every $x \in \mathbb{R}$. This is really a result about "almost-everywhere convergence," it asserts that if $f \in L^1(\mathbb{R})$ then

Don't forget: A "function" in $L^1(\mathbb{R})$ is really an *equivalence class* of functions, any two of which differ on at most a set of measure zero.

$$(2) \qquad \lim_{r\to 0}\frac{1}{r}\int_{x}^{x+r} f(t)\,dt = f(x) \quad \text{for a.e. } x \in \mathbb{R}.$$

The *Lebesgue Differentiation Theorem* makes a stronger statement in a more general setting:

Theorem 1.1 (The Lebesgue Differentiation Theorem)**.** *Suppose $d \in \mathbb{N}$ and $f \in L^1(\mathbb{R}^d)$. Then for a.e. $x \in \mathbb{R}^d$,*

Here: $\mathbb{N}$ denotes the set of positive integers, $B_r(x)$ denotes the open ball in $\mathbb{R}^d$ of radius $r > 0$ and center x, and m denotes Lebesgue measure on $\mathbb{R}^d$.

$$(3) \qquad \lim_{r\to 0+}\frac{1}{m(B_r(x))}\int_{B_r(x)} |f - f(x)|\,dm = 0$$

Each point $x \in \mathbb{R}^d$ for which (3) holds is called a *Lebesgue point* of f; the set of all such points is the *Lebesgue set of f*. I leave it to you to check that, back in our original one-dimensional setting, this version of Lebesgue's theorem implies that (2) holds for each $f \in L^1(\mathbb{R})$ and each x in the Lebesgue set of f.

Remarks. (a) Theorem 1.1 shows more; it's not difficult to check that for each $f \in L^1(\mathbb{R}^d)$, and each Lebesgue point x of f, equation (3) continues to hold if, for each x, the family of open balls centered at x is replaced by:

(i) The family of open cubes centered at x, or more generally

(ii) The family of open "rectangles" centered at x with largest side having length no more than some fixed multiple of the length of the shortest side.

For more on this see [4], §7.9–7.12, pp. 140–141.

(b) Note also that, once its proved, Theorem 1.1 "self-improves" in that its conclusion holds for every $f \in L^1_{loc}(\mathbb{R}^d)$, the space of Lebesgue measurable functions on $\mathbb{R}^d$ that are integrable on every set of finite measure.

Proposition 1.2. *Equation* (3) *is true for each* $f \in C_c(\mathbb{R}^d)$ *and each* $x \in \mathbb{R}^d$.

Proof. Fix $f \in C_c(\mathbb{R}^d)$. It's uniformly continuous on $\mathbb{R}^d$, so given $\varepsilon > 0$ we can find $\delta > 0$ such that $|f(t) - f(x)| < \varepsilon$ whenever $|t - x| < \varepsilon$. Thus if $0 < r < \delta$,

$$\frac{1}{m\{B_r(x)\}} \int_{B_r(x)} |f - f(x)|\, dm < \varepsilon,$$

which establishes (3). □

To prove Theorem 1.1 we'll use the time-honored strategy of observing that it's true for a dense subset of $L^1(\mathbb{R}^d)$, namely $C_c(\mathbb{R}^d)$, and then extending the result for the dense subset to the whole space via some kind of uniform estimate. But what kind of uniform estimate? To get an idea of what we need, let's rewrite the desired conclusion. For $f \in L^1(\mathbb{R}^d)$, let

To prove $C_c(\mathbb{R}^d)$ dense in $L^1(\mathbb{R}^d)$ one only need show that the L^1-closure of $C_c(\mathbb{R}^d)$ captures the characteristic function of each measurable subset of $\mathbb{R}^d$ of finite measure. By the regularity of Lebesgue measure it's enough to consider only characteristic functions of compact subsets, whereupon the desired result follows readily from Urysohn's Lemma (easy to prove for euclidean spaces; see e.g., §1(f) of [3]).

$$(4) \qquad \Omega f(x) := \limsup_{r \to 0+} \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} |f - f(x)|\, dm \qquad (x \in \mathbb{R}^d),$$

which allows the conclusion of Lebesgue's Theorem, equation (3), to be rewritten as:

$$(5) \qquad \Omega f(x) = 0 \quad \text{for a.e. } x \in \mathbb{R}^d.$$

Equation (5) suggests that, in our search for something to estimate "uniformly," we study the quantity

$$(6) \qquad Mf(x) := \sup_{r>0} \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} |f|\, dm \qquad (x \in \mathbb{R}^d).$$

Mf is the *Hardy-Littlewood Maximal Function* of f. It's clear that if $f \in L^1(\mathbb{R}^d)$ is essentially bounded, then so is Mf. But in general we don't even know if Mf is finite a.e. for each $f \in L^1(\mathbb{R}^d)$. In the next section we'll discuss the uniform estimate on the Maximal Function which establishes not only this, but in fact the entire Lebesgue Differentiation Theorem. Then in the following section we'll prove this uniform estimate: the famous "Hardy-Littlewood Maximal Theorem."

2 *The Maximal Function & The Maximal Theorem*

HERE'S THE KEY to all that follows.

Theorem 2.1 (The Hardy-Littlewood Maximal Theorem)**.** *For each $f \in L^1(\mathbb{R}^d)$ and $\lambda > 0$:*

$$m\{Mf > \lambda\} \leq \frac{3^d}{\lambda}\|f\|_1 \,. \tag{7}$$

Notation: $\{g > \lambda\} =: \{x \in \mathbb{R}^d : g(x) > \lambda\}$

Corollary 2.2. *$Mf < \infty$ a.e. for each $f \in L^1(\mathbb{R}^d)$.*

Proof. Fix $f \in L^1(\mathbb{R}^d)$. From inequality (7) we know that for each $n \in \mathbb{N}$ the set $\{Mf > n\}$ has measure $\leq 3^d\|f\|_1 \cdot \frac{1}{n}$. Now $\{Mf = \infty\}$ is the decreasing intersection of the sets $\{Mf > n\}$, and each of these sets lies in $\{Mf > 1\}$ which, by inequality (7), has finite measure, it follows from the "continuity of measure" that

$$m\{Mf = \infty\} = \lim_n m\{Mf > n\} = 0. \qquad \square$$

INEQUALITY (7) calls to mind:

Chebyshev's Inequality. *For each $f \in L^1(\mathbb{R}^d)$ and $\lambda > 0$:*

$$m\{|f| > \lambda\} \leq \frac{1}{\lambda}\|f\|_1 \,.$$

Also known as "Markov's Inequality," or "The Chebyshev-Markov Inequality." *Proof.* For $f \in L^1(\mathbb{R}^d)$:

$$\lambda m\{|f| > \lambda\} \leq \int_{\{|f|>\lambda\}} |f|\, dm \leq \|f\|_1.$$

According to this inequality, if only Mf belonged to $L^1(\mathbb{R}^d)$ for every $f \in L^1(\mathbb{R}^d)$, then we'd have a Hardy-Littlewood-like inequality, with the constant 3^d on the right-hand side improved to 1, but $\|Mf\|_1$ on the right-hand side instead of $\|f\|_1$. Sadly:

$Mf \notin L^1(\mathbb{R}^d)$ for any $f \in L^1(\mathbb{R}^d)$ that is not a.e. equal to 0!

Here's a special case that (I leave it to you to check) implies the general one. Suppose $f = \chi_K$, the characteristic function of a compact subset K of $\mathbb{R}^d$ that has positive measure. Let δ be the radius of the smallest closed origin-centered ball that contains K. Fix $x \in \mathbb{R}^d \backslash K$. Then the closed ball of radius $r = |x| + \delta$, centered at x, contains K, so

$$(M\chi_K)(x) \geq \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} \chi_K \, dm = \frac{m(K)}{r^d} = \frac{m(K)}{(|x|+\delta)^d}.$$

Here we use the fact that an open ball and its closure have the same Lebesgue measure, as well as the scaling property of Lebesgue measure, and our normalization that balls of radius 1 have measure 1.

Since the right-hand side of this inequality is not integrable over $\mathbb{R}^d \backslash K$, neither is $M\chi_K$. *Conclusion*: $M\chi_K \notin L^1(\mathbb{R}^d)$. $\square$

3 *Proof that: Maximal Theorem ⇒ Differentiation Theorem*

The proofs here and in the next section are standard; cf. Rudin [4], Chapter 7, for example.

Fix $f \in L^1(\mathbb{R}^d)$. We wish to show that for a.e. $x \in \mathbb{R}^d$:

$$\Omega f(x) := \limsup_{r \to 0+} \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} |f - f(x)| \, dm = 0.$$

Now for $x \in \mathbb{R}^d$:

$$\begin{aligned}\Omega f(x) &\le \sup_{r>0} \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} |f - f(x)|\, dm \\ &\le \sup_{r>0} \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} (|f| + |f(x)|)\, dm \\ &= \sup_{r>0} \frac{1}{m\{B_r(x)\}} \int_{B_r(x)} |f|\, dm + |f(x)|\,,\end{aligned}$$

i.e.,

$$\Omega f(x) \le Mf(x) + |f(x)| \quad \forall\, x \in \mathbb{R}^d. \tag{8}$$

Now recall that the Lebesgue Differentiation Theorem holds for every $g \in C_c(\mathbb{R}^d)$. Thus $\Omega g \equiv 0$ for each $g \in C_c(\mathbb{R}^d)$, and we can write $f = g + h$ where $g \in C_c(\mathbb{R}^d)$ can be chosen so that $h \in L^1(\mathbb{R}^d)$ has norm as small as we wish. Upon checking that the "Ω-operator" is subadditive, then appealing to inequality (8), we see that for each such decomposition $f = g + h$ and each $x \in \mathbb{R}^d$:

$$\Omega f(x) \le \underbrace{\Omega g(x)}_{=0} + \Omega h(x) = \underbrace{\Omega h(x) \le Mh(x) + |h(x)|}_{by\ (8)} \qquad \forall\, x \in \mathbb{R}^d.$$

Thus for each $\lambda > 0$,

$$\{\Omega f > \lambda\} \subset \{Mh + |h| > \lambda\} \subset \Big\{Mh > \frac{\lambda}{2}\Big\} + \Big\{|h| > \frac{\lambda}{2}\Big\},$$

from which it follows that

$$\begin{aligned}m\{\Omega f > \lambda\} &\le m\{\Omega h > \lambda\} \\ &\le m\Big\{Mh > \frac{\lambda}{2}\Big\} + m\Big\{|h| > \frac{\lambda}{2}\Big\} \\ &\le 2\frac{3^d}{\lambda}\|h\|_1 + 2\frac{\|h\|_1}{\lambda}\end{aligned}$$

where in the last line: the first term on the right-hand side of the inequality comes from the Hardy-Littlewood Maximal Theorem, and the second one from Chebyshev's inequality.

In summary: if $f \in L^1(\mathbb{R}^d)$ is decomposed as $f = g + h$ with $g \in C_c(\mathbb{R}^d)$ and $h \in L^1(\mathbb{R}^d)$, then for every $\lambda > 0$,

$$m\{\Omega f > \lambda\} \le \frac{C_d}{\lambda}\|h\|_1\,, \tag{9}$$

where $C_d = 2(3^d + 1)$.

Now fix $\lambda > 0$ and $\varepsilon > 0$, and use the denseness of $C_c(\mathbb{R}^d)$ in $L^1(\mathbb{R}^d)$ to choose $g \in C_c(\mathbb{R}^d)$ such that $\|f - g\|_1 \le \varepsilon\lambda / C_d$. Upon setting $h = f - g$ we see from equation (9) that

$$m\{\Omega f > \lambda\} \le \varepsilon\,,$$

so, because ε was an arbitrary positive number, $m\{\Omega f > \lambda\} = 0$, and this is true *for each* $\lambda > 0$. Now

$$\{\Omega f > 0\} = \bigcup_{n \in \mathbb{N}} \left\{\Omega f > \frac{1}{n}\right\},$$

which exhibits the set $\{\Omega f > 0\}$ as a countable union of sets of measure zero, so $m\{\Omega f > 0\} = 0$.

CONCLUSION: $\Omega f = 0$ a.e. for each $f \in L^1(\mathbb{R}^d)$. □

4 *Proof of the Maximal Theorem*

Fix $f \in L^1(\mathbb{R}^d)$ and $\lambda > 0$. We need to prove:

(7) $$m\{Mf > \lambda\} \leq \frac{3^d}{\lambda} \|f\|_1 .$$

By the regularity of Lebesgue measure, it's enough to show that: for any compact subset K of $\{Mf > \lambda\}$ inequality (7) holds with K replacing $\{Mf > \lambda\}$ on the left-hand side.

For each $x \in K$ the definition of $Mf(x)$ promises that there is an open ball B_x centered at x such that

$$\frac{1}{m\{B_x\}} \int_{B_x} f \, dm > \lambda \,,$$

i.e., that

(10) $$m\{B_x\} < \frac{1}{\lambda} \int_{B_x} f \, dm \,.$$

Since the collection of balls $\mathscr{B} := \{B_x : x \in K\}$ is an open cover of K, there exists a finite subcollection $\{B_{x_j}\}_1^n$ that still covers K.

Let's write B_j for B_{x_j}, and indulge in some (temporary) wishful thinking. Suppose our subcollection of B_j's is *pairwise disjoint*! Then we'd have:

$$\begin{aligned} m\{K\} &\leq m\Big\{\bigcup_{j=1}^n B_j\Big\} \leq \sum_{j=1}^n m\{B_j\} \\ &\leq \sum_{j=1}^n \frac{1}{\lambda} \int_{B_j} |f| \, dm \quad \text{[by (10)]} \\ &= \frac{1}{\lambda} \int_{\cup_j B_j} |f| \, dm \qquad \text{[by disjointness !]} \\ &\leq \frac{1}{\lambda} \int_{\mathbb{R}} |f| \, dm \\ &= \frac{1}{\lambda} \|f\|_1 \end{aligned}$$

which would prove (a better version of) the Maximal Theorem.

NEEDLESS TO SAY, the finite subcollection $\{B_j\}_1^n$ of K need *not* be pairwise disjoint. However all is not lost! Choose a ball $B_j \in \mathscr{B}$ of largest radius (if there are several, make an arbitrary choice). Call this ball L_1, and remove from $\mathscr{B}$: not only L_1, but also all the "satellite" balls in $\mathscr{B}$ that intersect it. Note that $3L_1$, the open ball having the same center as L_1 and three times its radius, contains every one of these satellites!

If L_1 and its satellites exhaust all of $\mathscr{B}$, stop. Otherwise do the same with the remaining balls in $\mathscr{B}$, i.e., choose the largest one—call it L_2, remove it and all its satellites, and note that L_1 is disjoint from L_1, and that $3L_2$ swallows up all of L_2's satellites. Continue if necessary; eventually this "greedy algorithm" must halt, resulting in a pairwise disjoint collection $\{L_k\}_1^\ell$ of open balls with centers in K such that $\{3L_k\}_1^\ell$ covers K. Thus:

It's "greedy" because at each stage the algorithm grabs the largest available piece of the pie.

$$m\{K\} \le m\Big\{\bigcup_{k=1}^{\ell} 3L_k\Big\} \le \underbrace{\sum_{k=1}^{\ell} m\{3L_k\}}_{\text{since } K\subset\bigcup_k 3L_k} = \underbrace{3^d \sum_{k=1}^{\ell} m\{L_k\}}_{\text{since } m\{3L_k\}=3^d m\{L_k\}} .$$

The estimate now proceeds as in the "wishful thinking" argument, with L_k's replacing the B_j's. □

5 a.e. Convergence vs. Maximal Functions . . . The "Cosmic Truth"

STANDING HYPOTHESES for this section: $(T, \mathscr{F}, \mu)$ is a finite measure space (think: $[0,1]$ with Lebesgue measure), and $L^0 = L^0(\mu)$ is the space of (μ-a.e. equivalence classes of) $\mathscr{F}$-measurable real-valued functions that are a.e. finite-valued. The definition

$$\|f\|_0 := \int \frac{|f|}{1+|f|}\, d\mu \qquad (f \in Ł^0)$$

induces a translation-metric

$$d(f,g) = \|f-g\|_0 \qquad (f,g \in L^0)$$

that makes L^0 into a metrizable topological vector space in which sequential convergence is just convergence in measure (easy exercise). Since "Cauchy in measure implies convergence in measure," L^0 is complete in this metric.

A *topological vector space* is a vector space V with a topology that makes addition $V \times V \to V$ and scalar multiplication $\mathbb{R} \times V \to \mathbb{R}$ continuous.

Let E denote a Banach space, and $T \colon E \to L^0$ a linear transformation that is continuous, i.e., for which

$$f_n \to f \text{ in } E \implies Tf_n \to Tf \text{ in } \mu\text{-measure on } X.$$

Now suppose we have linear transformations $T_n \colon E \to L^0$, with T_n continuous for each $n \in \mathbb{N}$, and

To say the limit of a sequence of real-valued functions "exists a.e." means it exists *finitely* a.e.

$$(*) \qquad \lim_n T_n f := Tf \text{ exists for } \mu\text{-a.e. } \forall f \in E.$$

Then clearly the mapping $T \colon E \to L^0$ is linear, and the "maximal function"

$$T^* f := \sup_n |T_n f|$$

is finite a.e. for every $f \in E$. In his 1926 paper [1], Stefan Banach told us much more.

Theorem 5.1 (Banach's a.e. convergence theorem). *For (T_n) and T^* as above, the "maximal map" $T^* \colon E \to L^0$ is continuous; consequently, so is the linear transformation T.*

See [1] Théoréme I, pp. 356-7 and Théoréme II, page 359.

In plain language, the conclusion on T^* asserts that for every pair of positive numbers ε and λ there exists $\delta > 0$ such that

$$(11) \qquad \|f\| < \delta \implies \mu\{T^* f > \lambda\} < \varepsilon\,.$$

In other words (at least for μ a finite measure): "pointwise a.e. convergence of a sequence (T_n) of continuous linear transformations $E \to L^0$ implies a maximal theorem for (T_n).

BANACH THEN PROVES the following converse, in just the way we used the Hardy-Littlewood maximal theorem to prove Lebesgue's Theorem:

Theorem 5.2 (Banach's a.e. converse). *Suppose (T_n) is a sequence of continuous linear maps $E \to L^0$ for which $T^* f$ is finite a.e. for each $f \in E$, and suppose* $\lim_n T_n f$ *exists a.e. for each f in a dense subset of E. Then the limit*

See [1], Théoréme III, page 359.

$$\lim_n T_n f := Tf$$

exists a.e. for each $f \in E$, *and the linear transformation T so defined is continuous $E \to L^0$.*

EXAMPLE. Take our measure space to be $[0, 1]$ with Lebesgue measure (and Lebesgue measurable sets as sigma-algebra), and take E to be the resulting L^1-space, i.e. $E = L^1([0, 1])$. Fix a sequence $1 \geq r_n \searrow 0$, and for each n define $T_n \colon L^1 \to L^0$ by

$$T_n f = \frac{1}{2r_n} \int_{-r_n}^{r_n} f(t)\, dt \qquad (f \in L^1).$$

Then the maximal function $T^* f$ is pointwise dominated, for each $f \in L^1$, by the Hardy-Littlewood maximal function Mf, so we have

$$m\{T^* f > \lambda\} \leq \frac{3}{\lambda} \|f\|_1$$

where m is Lebesgue measure (normalized so that $m\{[-1,1]\} = 1$). Thus we have the following special case of (11) above:

$$\|f\|_1 < \delta := \frac{\varepsilon\lambda}{3} \implies m\{T^* f > \lambda\} < \varepsilon\,.$$

Banach used a "condensation of singularities" argument to prove Theorem 5.1. For a more modern proof, using the Baire Category Theorem, see Garsia's monograph [2], Chapter 1, pp. 1–4.

References

1. Stefan Banach, *Sur la convergence presque partout de fonctionelles linéaires*, Bull. Sci. Math 50 (1926) 27–32 & 36–43.

2. Adriano M. Garsia, *Topics in Almost Everywhere Convergence*, Markham 1970.

3. Christopher Heil, *Lecture Notes for Real Analysis, Georgia Institute of Technology, 2008*. Freely downloadable at: `http://people.math.gatech.edu/~heil/6338/summer08/`

4. Walter Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill 1987.