Abstract

We discuss and prove some fundamental results about differentiation, after which prove the fundamental theorem of calculus for Lebesgue integrals.

Preliminary

In our discussion of the Radon–Nikodym Theorem, we were able to define an abstract derivative of a (σ-finite) complex measure with respect to a different (σ-finite) measure. In Euclidean space, $\mathbb{R}^n$, we may consider trying to define a "pointwise" derivative by taking

$$F(x) = \lim_{r\to 0} \frac{\nu(U(x,r))}{m(U(x,r))},$$

where m is the Lebesgue measure, and ν is our given complex measure. If we take the Lebesgue–Radon–Nikodym decomposition

$$d\nu = d\lambda + f\,dm,$$

we would hope that $F = f$ almost everywhere. Indeed, we will show this to be the case, after which we may prove a stronger version of the fundamental theorem of calculus, this time for Lebesgue integrals.

Note that from now on, every measure-theoretic term (i.e., integrable, almost everywhere, etc.) is taken with respect to the Lebesgue measure on $\mathbb{R}^n$.

We start with a fundamental lemma in measure theory for Euclidean spaces.

Theorem (Vitali Covering Lemma)**:** Let $\mathcal{C}$ be a collection of open balls in $\mathbb{R}^n$, and let $U = \bigcup_{B\in\mathcal{C}} B$.

If $c < m(U)$, then there exist disjoint $B_1, \dots, B_k$ such that

$$3^{-n}c \leq \sum_{j=1}^{k} m(B_j).$$

Proof. By inner regularity, there is a compact $K \subseteq U$ such that $m(K) > c$; finitely many balls in $\mathcal{C}$, which we call $A_1, \dots, A_m$, cover K.

We proceed via exhaustion; select B_1 to be the largest of the A_j, B_2 to be the largest of the A_j disjoint from B_1, B_3 the largest of the A_j disjoint from B_2 and B_1, etc. According to this construction, if A_i is not among the B_j, then there is j such that $A_i \cap B_j \neq \emptyset$, and if j is the smallest such index, then the radius of A_i is at most that of B_j. Via some triangle inequality magic, we see that $A_i \subseteq B_j^*$, where B_j^* is defined to the ball with the same center as B_j and three times the radius.

Then, $K \subseteq \bigcup_{j=1}^{k} B_j^*$, so that

$$\begin{aligned} c &< m(K) \\ &\leq \sum_{j=1}^{k} m\left(B_j^*\right) \\ &= 3^n \sum_{j=1}^{k} m(B_j). \end{aligned}$$

□

The Lebesgue Differentiation Theorem

Definition: A function $f\colon \mathbb{R}^n \to \mathbb{C}$ is called *locally integrable* if $\int_K |f|\,dm < \infty$ for every bounded measurable $K \subseteq \mathbb{R}^n$.[I]

The space of locally integrable functions is denoted $L_{1,\text{loc}}$.

Definition: If $f \in L_{1,\text{loc}}$, and $x \in \mathbb{R}^n$, and $r > 0$, define

$$A_r f(x) = \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)\,dy$$

to be the *average* of f on $B(x,r)$.

Lemma: If $f \in L_{1,\text{loc}}$, then $A_r f$ is jointly continuous in r and x.

Proof. We know that $m(B(x,r)) = cr^n$, where $c = m(B(0,1))$, and $m(S(x,r)) = 0$, where $S(x,r) = \{y \mid |y-x| = r\}$.

Moreover, as $r \to r_0$ and $x \to x_0$, $\mathbb{1}_{B(x,r)} \to \mathbb{1}_{B(x_0,r_0)}$ pointwise on $\mathbb{R}^n \setminus S(x_0,r_0)$, so the convergence is pointwise almost everywhere. Furthermore, note that $\left|\mathbb{1}_{B(x,r)}\right| \leq \mathbb{1}_{B(x_0,r_0+1)}$ for $r < r_0 + 1/2$ and $|x - x_0| < 1/2$. Thus, by dominated convergence, it follows that $\int_{B(x,r)} f(y)\,dy$ is continuous in r and x, and so is $A_r f(x)$. □

Definition: If $f \in L_{1,\text{loc}}$, we define the *Hardy–Littlewood Maximal Function*, Hf, by

$$\begin{aligned} Hf(x) &= \sup_{r>0} A_r|f|(x) \\ &= \sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\,dy. \end{aligned}$$

Theorem (The Maximal Theorem)**:** There is a constant $C > 0$ such that for all $f \in L_1$ and all $\alpha > 0$,

$$m(\{x \mid Hf(x) > \alpha\}) \leq \frac{C}{\alpha} \int_{\mathbb{R}^n} |f(x)|\,dx.$$

Proof. Let $E_\alpha = \{x \mid Hf(x) > \alpha\}$. For each $x \in E_\alpha$, we may find $r_x > 0$ such that $A_{r_x}|f|(x) > \alpha$. The balls $U(x,r_x)$ cover E_α, so by the Vitali Covering Lemma, if $c < m(E_\alpha)$, then there are $x_1, \dots, x_k$ such that $B_j = B\left(x_j, r_{x_j}\right)$ are disjoint and $\sum_{j=1}^{k} m(B_j) > 3^{-n}c$.

Then, we see that

$$\begin{aligned} c &< 3^n \sum_{j=1}^{k} m(B_j) \\ &\leq \frac{3^n}{\alpha} \sum_{j=1}^{k} \int_{B_j} |f(y)|\,dy \\ &\leq \frac{3^n}{\alpha} \int_{\mathbb{R}^n} |f(y)|\,dy. \end{aligned}$$

Thus, letting $c \to m(E_\alpha)$, we obtain our desired result. □

Exercise: A variant of the Hardy–Littlewood Maximal Function is defined by

$$H^* f(x) = \sup\left\{ \frac{1}{m(B)} \int_B |f(y)|\,dy \;\middle|\; B \text{ is a ball, } x \in B \right\}.$$

[I] Note that we still use the convention $0 \cdot \infty = 0$.

Show that $Hf \leq H^* f \leq 2^n Hf$.

Solution: We see that, necessarily,

$$\frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y)|\,dy \leq H^* f(x),$$

so that $Hf(x) \leq H^* f(x)$.

Now, if $r > 0$ is such that $x \in B(z,r)$, then clearly $B(z,r) \subseteq B(y,2r)$, so

$$\begin{aligned}\frac{1}{m(B(z,r))}\int_{B(z,r)} |f(y)|\,dy &\leq \frac{1}{m(B(z,r))}\int_{B(x,2r)} |f(y)|\,dy\\ &\leq \frac{2^n}{m(B(x,2r))}\int_{B(x,2r)} |f(y)|\,dy\\ &\leq 2^n Hf(x).\end{aligned}$$

Thus, by taking suprema, we see that $H^* f(x) \leq 2^n Hf(x)$.

Definition: If $f\colon \mathbb{R} \to \mathbb{C}$ is a function, then the limit superior as r approaches R is defined to be

$$\limsup_{r\to R} \phi(r) := \lim_{\varepsilon\to 0}\left(\sup_{0<|r-R|<\varepsilon} \phi(r)\right).$$

Remark: Note that

$$\lim_{r\to R} \phi(r) = C$$

if and only if

$$\limsup_{r\to R} |\phi(r) - c| = 0.$$

We will prove progressively stronger versions of the Lebesgue Differentiation Theorem.

Theorem: If $f \in L_{1,\text{loc}}$, then $\lim_{r\to 0} A_r f(x) = f(x)$ for a.e. $x \in \mathbb{R}^n$.

Proof. It suffices to show that for any $N \in \mathbb{N}$, $A_r f(x) \to f(x)$ for almost every x with $|x| \leq N$. Furthermore, we may replace f by $f\mathbb{1}_{B(0,N+1)}$ in this scenario, as $A_r f(x)$ only depends on the value $f(y)$ for $|y| \leq N+1$. Thus, we may assume $f \in L_1$.

Given $\varepsilon > 0$, there is a compactly supported continuous function g such that $\|g - f\|_{L_1} < \varepsilon$. Since g is continuous, for any $x \in \mathbb{R}^n$ and $\delta > 0$, there is $r > 0$ such that $|g(y) - g(x)| < \delta$ whenever $|y - x| < r$. Thus,

$$\begin{aligned}|A_r g(x) - g(x)| &= \frac{1}{m(B(x,r))}\left|\int_{B(x,r)} g(y) - g(x)\,dy\right|\\ &< \delta,\end{aligned}$$

meaning $A_r g(x) \to g(x)$ as $r \to 0$ for every x. Thus,

$$\begin{aligned}\limsup_{r\to 0} |A_r f(x) - f(x)| &= \limsup_{r\to 0} |A_r(f-g)(x) + (A_r g - g)(x) + (g-f)(x)|\\ &\leq H(f-g)(x) + |f-g|(x).\end{aligned}$$

Now, if we set

$$E_\alpha = \left\{x \;\middle|\; \limsup_{r\to 0} |A_r f(x) - f(x)| > \alpha\right\}$$

$$F_\alpha = \{x \mid |f-g|(x) > \alpha\},$$

then

$$E_\alpha \subseteq F_{\alpha/2} \cup \{x \mid H(f-g)(x) > \alpha/2\}$$

Now, we see that

$$\begin{aligned} m\left(F_{\alpha/2}\right) &\leq \frac{2}{\alpha}\int_{F_{\alpha/2}} |f(x)-g(x)|\,dx \\ &< \varepsilon, \end{aligned}$$

so by the Maximal Theorem,

$$m(E_\alpha) \leq \frac{2}{\alpha}\varepsilon + \frac{2C}{\alpha}\varepsilon,$$

and since ε is arbitrary, $m(E_\alpha) = 0$. Thus, $\lim_{r\to 0} A_r f(x) = f(x)$ for all $x \notin \bigcup_{n=1}^{\infty} E_{1/n}$.

Thus, we find that if $f \in L_{1,\text{loc}}$, then

$$\lim_{r\to 0} \frac{1}{m(B(x,r))}\int_{B(x,r)} (f(y)-f(x))\,dy = 0$$

for almost every x. □

In fact, we can prove something stronger.

Definition: Let $f \in L_{1,\text{loc}}$, then we define

$$L_f = \left\{x \;\middle|\; \lim_{r\to 0}\frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y)-f(x)|\,dy = 0\right\}$$

to be the *Lebesgue set* of f.

Theorem: If $f \in L_{1,\text{loc}}$, then $m((L_f)^c) = 0$.

Proof. For each $c \in \mathbb{C}$, we may apply the previous theorem to the function $g_c(x) = |f(x)-c|$ to get that, except for a certain null set E_c,

$$\lim_{r\to 0}\frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y)-c|\,dy = |f(x)-c|.$$

Now, if D is a countable dense subset of $\mathbb{C}$, and $E = \bigcup_{c\in D} E_c$, then $m(E) = 0$, and if $x \notin E$, there is $c \in D$ with $|f(x)-c| < \varepsilon$, so that $|f(y)-f(x)| < |f(y)-c| + \varepsilon$, and

$$\begin{aligned} \limsup_{r\to 0}\frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y)-f(x)|\,dy &\leq |f(x)-c| + \varepsilon \\ &< 2\varepsilon. \end{aligned}$$

Since ε is arbitrary, our desired result follows. □

Definition: We say a family of subsets $\{E_r\}_{r>0}$ of Borel subsets of $\mathbb{R}^n$ *shrinks nicely* to $x \in \mathbb{R}^n$ if

- $E_r \subseteq B(x,r)$ for each r;
- there is $\alpha > 0$ independent of r such that $m(E_r) > \alpha m(B(x,r))$.

Remark: The sets E_r need not contain x.

Theorem (Lebesgue Differentiation Theorem)**:** If $f \in L_{1,\text{loc}}$, then for every $x \in L_f$,

$$\begin{aligned}\lim_{r\rightarrow 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy = 0\\ \lim_{r\rightarrow 0} \frac{1}{m(E_r)} \int_{E_r} f(y)\, dy = f(x)\end{aligned}$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to zero.

Proof. For some $\alpha > 0$, the definition of $\{E_r\}_{r>0}$ allows us to take

$$\begin{aligned}\frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy &\leq \frac{1}{m(E_r)} \int_{B(x,r)} |f(y) - f(x)|\, dy\\ &\leq \frac{1}{\alpha m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\, dy.\end{aligned}$$

□

Definition: A Borel measure ν on $\mathbb{R}^n$ is called *regular* if

(i) $\nu(K) < \infty$ for all compact sets K;

(ii) for all $E \in \mathcal{B}_{\mathbb{R}^n}$,

$$\nu(E) = \inf\{\nu(U) \mid U \text{ open, } E \subseteq U\}.$$

If ν is a signed measure, then we say ν is regular if $|\nu|$ is regular.

Proposition: If λ and μ are positive, mutually singular, and $\lambda + \mu$ is regular, then λ and μ are regular.

Proof. Let $A \subseteq \mathbb{R}^n$ be such that A is μ-null and A^c is λ-null.

We see that condition (i) in the definition of regularity holds necessarily, so we show condition (ii). Now, let $\varepsilon > 0$ and let $E \subseteq A$ be Borel. Since $\lambda + \mu$ is regular, and λ is concentrated on A, there is an open $U \in \mathcal{B}_{\mathbb{R}^n}$ such that

$$\begin{aligned}(\lambda + \mu)(U) &< (\lambda + \mu)(E) + \varepsilon\\ &= \lambda(E) + \varepsilon,\end{aligned}$$

meaning that

$$\lambda(U) < \lambda(E) + \varepsilon,$$

so condition (ii) for λ, and similarly for μ (by taking $E \subseteq A^c$). □

Proposition: The measure $f\, dm$ is regular if and only if $f \in L_{1,\text{loc}}$.

Proof. The condition $f \in L_{1,\text{loc}}$ is equivalent to $f\, dm$ being finite on compact sets, so condition (i) holds.

Now, if E is a bounded Borel set, then given $\delta > 0$, there is a bounded open $U \supseteq E$ such that $m(U) < m(E) + \delta$, meaning $m(U \setminus E) < \delta$. At the same time, given $\varepsilon > 0$, there is an open $U \supseteq E$ such that $\int_{U\setminus E} f\, dm < \varepsilon$, meaning $\int_U f\, dm < \int_E f\, dm + \varepsilon$ with $m(U \setminus E) < \delta$.

If E is unbounded, then we write $E = \bigcup_{j=1}^{\infty} E_j$ as a union of bounded Borel sets, and finding $U_j \supseteq E_j$ with $\int_{U_j \setminus E_j} f\, dm < \varepsilon 2^{-j}$. □

Theorem: Let ν be a regular signed or complex Borel measure on $\mathbb{R}^n$, and let $d\nu = d\lambda + f\ dm$ be the Lebesgue–Radon–Nikodym representation. Then, m-a.e. $x \in \mathbb{R}^n$,

$$\lim_{r\to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

Proof. Note that $d|\nu| = d|\lambda| + |f|\ dm$, so regularity of ν means that both λ and $f\ dm$ are regular.

Since $f \in L_{1,\text{loc}}$, it suffices to show that if λ is regular and $\lambda \perp m$, then for m-a.e. x,

$$\lim_{r\to 0} \frac{\lambda(E_r)}{m(E_r)} = 0$$

whenever E_r shrinks nicely to x.

It also suffices to take $E_r = B(x, r)$, and assume λ is positive, since for some $\alpha > 0$,

$$\begin{aligned}\left|\frac{\lambda(E_r)}{m(E_r)}\right| &\leq \frac{|\lambda|(E_r)}{m(E_r)}\\ &\leq \frac{|\lambda|(B(x,r))}{m(E_r)}\\ &\leq \frac{|\lambda|(B(x,r))}{\alpha m(B(x,r))}.\end{aligned}$$

Thus, let A be a Borel set such that $\lambda(A) = m(A^c) = 0$. Let

$$F_k = \left\{x \in A \;\middle|\; \limsup_{r\to 0} \frac{\lambda(B(x,r))}{m(B(x,r))} > \frac{1}{k}\right\}.$$

We will show that $m(F_k) = 0$ for all k.

Now, by regularity of λ, given $\varepsilon > 0$, there is $U_\varepsilon \supseteq A$ such that $\lambda(U_\varepsilon) < \varepsilon$. Each $x \in F_k$ is the center of an open ball $U_x \subseteq U_\varepsilon$ such that $\lambda(U_x) > \frac{1}{k} m(U_x)$ (by the properties of the limit superior). Now, if $V_\varepsilon = \bigcup_{x\in F_k} U_x$, and $c < m(V_\varepsilon)$, then by the Vitali Covering Lemma, there are $U_{x_1}, \dots, U_{x_J}$ that are disjoint such that

$$\begin{aligned}c &< 3^n \sum_{j=1}^{J} m\big(U_{x_j}\big)\\ &< 3^n k \sum_{j=1}^{J} \lambda\big(U_{x_j}\big)\\ &< 3^n k \lambda(V_\varepsilon)\\ &\leq 3^n k \lambda(U_\varepsilon)\\ &\leq 3^n k \varepsilon,\end{aligned}$$

meaning that $m(V_\varepsilon) \leq 3^n k\varepsilon$, and since $F_k \subseteq V_\varepsilon$ and ε is arbitrary, $m(F_k) = 0$. ☐

The Fundamental Theorem of Calculus for Lebesgue Integration

Recall from the construction of the Lebesgue measure that there is a one-to-one correspondence between increasing, right-continuous function on $\mathbb{R}$ and Borel measures μ_F determined by $\mu_F((a, b]) = F(b) - F(a)$. We will use this to help prove the almost-everywhere differentiability of increasing functions.

Theorem: Let $F\colon \mathbb{R} \to \mathbb{R}$ be increasing, and let $G(x) = F(x+)$.

(a) The set of points at which F is discontinuous is countable.

(b) The functions F and G are differentiable almost everywhere, and $F' = G'$ almost everywhere.

Proof.

(a) Since F is increasing, the intervals $(F(x-), F(x+))$ for each x are disjoint, and for $|x| < N$, they lie in the interval $(F(-N,), F(N))$. Thus,

$$\begin{aligned}\sum_{|x|<N} (F(x+) - F(x-)) &\leq F(N) - F(-N)\\ &< \infty,\end{aligned}$$

meaning that the set of all x in $(-N, N)$ such that $F(x+) \neq F(x-)$ is countable.

(b) Observe that G is increasing and right-continuous, and $G = F$ almost everywhere. Moreover, we see that

$$G(x+h) - G(x) = \begin{cases}\mu_G((x, x+h]) & h > 0\\ -\mu_G((x+h, x]) & h < 0,\end{cases}$$

and the families $\{(x - |h|, x]\}$ and $\{(x, x+|h|]\}$ shrink nicely to x as $|h| \to 0$. Applying the previous theorem, since μ_G is regular, we see that $G'(x)$ exists almost everywhere.

Finally, we show that if $H = G - F$, then H' is zero almost everywhere. Letting $\{x_j\}_{j=1}^{\infty}$ be an enumeration of points where $H \neq 0$, we see that $H(x_j) > 0$, and $\sum_{|x_j|<N} H(x_j) < \infty$ for any N.

Let δ_j be the point mass at x_j, and set $\mu = \sum_{j=1}^{\infty} H(x_j)\delta_j$. Note that μ is finite on compact sets, and μ is regular, and $\mu \perp m$ since $m\left(\{x_j\}_{j=1}^{\infty}\right) = \mu\left(\left(\{x_j\}_{j=1}^{\infty}\right)^c\right) = 0$.

Then,

$$\begin{aligned}\left|\frac{H(x+h) - H(x)}{h}\right| &\leq \frac{H(x+h) + H(x)}{h}\\ &\leq \frac{4\mu((x - 2|h|, x + 2|h|))}{4|h|},\end{aligned}$$

which tends to zero as $h \to 0$ for almost every x, meaning $H' = 0$ almost everywhere.

$\square$