Chapter 4
Differentiation

Does there exist a Lebesgue measurable set that fills up exactly half of each interval? To get a feeling for this question, consider the set $E = [0, \frac{1}{8}] \cup [\frac{1}{4}, \frac{3}{8}] \cup [\frac{1}{2}, \frac{5}{8}] \cup [\frac{3}{4}, \frac{7}{8}]$. This set E has the property that

$$|E \cap [0, b]| = \frac{b}{2}$$

for $b = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$. Does there exist a Lebesgue measurable set $E \subset [0, 1]$, perhaps constructed in a fashion similar to the Cantor set, such that the equation above holds for all $b \in [0, 1]$?

In this chapter we see how to answer this question by considering differentiation issues. We begin by developing a powerful tool called the Hardy–Littlewood maximal inequality. This tool is used to prove an almost everywhere version of the Fundamental Theorem of Calculus. These results lead us to an important theorem about the density of Lebesgue measurable sets.

Trinity College at the University of Cambridge in England. G. H. Hardy (1877–1947) and John Littlewood (1885–1977) were students and later faculty members here. If you have not already done so, you should read Hardy's remarkable book A Mathematician's Apology *(do not skip the fascinating Foreword by C. P. Snow) and see the movie* The Man Who Knew Infinity, *which focuses on Hardy, Littlewood, and Srinivasa Ramanujan (1887–1920).*

S. Axler, *Measure, Integration & Real Analysis*, Graduate Texts in Mathematics 282, https://doi.org/10.1007/978-3-030-33143-6_4

4A Hardy–Littlewood Maximal Function

Markov's Inequality

The following result, called Markov's inequality, has a sweet, short proof. We will make good use of this result later in this chapter (see the proof of 4.10). Markov's inequality also leads to Chebyshev's inequality (see Exercise 2 in this section).

4.1 *Markov's inequality*

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $h \in \mathcal{L}^1(\mu)$. Then

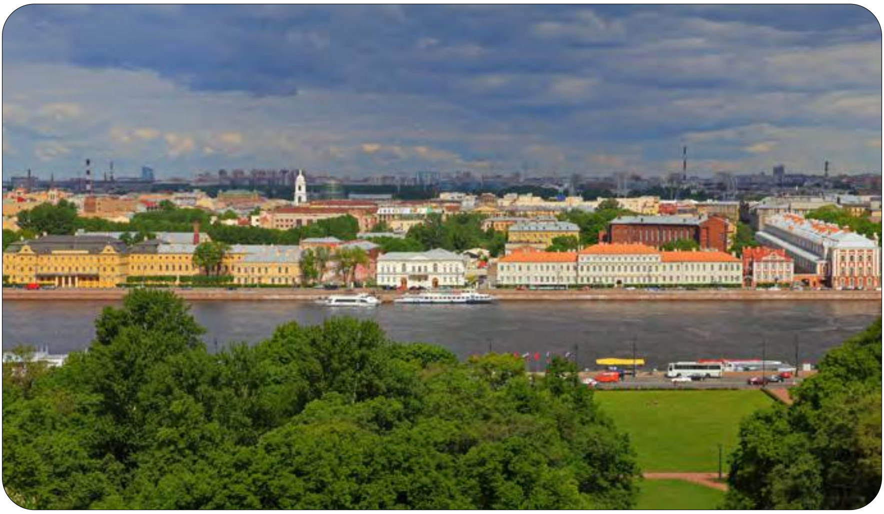
$$\mu(\{x \in X : |h(x)| \geq c\}) \leq \frac{1}{c}\|h\|_1$$

for every $c > 0$.

Proof Suppose $c > 0$. Then

$$\begin{aligned} \mu(\{x \in X : |h(x)| \geq c\}) &= \frac{1}{c}\int_{\{x \in X : |h(x)| \geq c\}} c \, d\mu \\ &\leq \frac{1}{c}\int_{\{x \in X : |h(x)| \geq c\}} |h| \, d\mu \\ &\leq \frac{1}{c}\|h\|_1, \end{aligned}$$

as desired. ■



St. Petersburg University along the Neva River in St. Petersburg, Russia. Andrei Markov (1856–1922) was a student and then a faculty member here.

Vitali Covering Lemma

4.2 Definition *3 times a bounded nonempty open interval*

Suppose I is a bounded nonempty open interval of $\mathbf{R}$. Then $3 * I$ denotes the open interval with the same center as I and three times the length of I.

4.3 Example *3 times an interval*

If $I = (0, 10)$, then $3 * I = (-10, 20)$.

The next result is a key tool in the proof of the Hardy–Littlewood maximal inequality (4.8).

4.4 ***Vitali Covering Lemma***

Suppose $I_1, \dots, I_n$ is a list of bounded nonempty open intervals of $\mathbf{R}$. Then there exists a disjoint sublist $I_{k_1}, \dots, I_{k_m}$ such that

$$I_1 \cup \cdots \cup I_n \subset (3 * I_{k_1}) \cup \cdots \cup (3 * I_{k_m}).$$

4.5 Example *Vitali Covering Lemma*

Suppose $n = 4$ and

$$I_1 = (0, 10), \quad I_2 = (9, 15), \quad I_3 = (14, 22), \quad I_4 = (21, 31).$$

Then

$$3 * I_1 = (-10, 20), \quad 3 * I_2 = (3, 21), \quad 3 * I_3 = (6, 30), \quad 3 * I_4 = (11, 41).$$

Thus

$$I_1 \cup I_2 \cup I_3 \cup I_4 \subset (3 * I_1) \cup (3 * I_4).$$

In this example, I_1, I_4 is the only sublist of I_1, I_2, I_3, I_4 that produces the conclusion of the Vitali Covering Lemma.

Proof of 4.4 Let k_1 be such that

$$|I_{k_1}| = \max\{|I_1|, \dots, |I_n|\}.$$

Suppose $k_1, \dots, k_j$ have been chosen. Let k_{j+1} be such that $|I_{k_{j+1}}|$ is as large as possible subject to the condition that $I_{k_1}, \dots, I_{k_{j+1}}$ are disjoint. If there is no choice of k_{j+1} such that $I_{k_1}, \dots, I_{k_{j+1}}$ are disjoint, then the procedure terminates. Because we start with a finite list, the procedure must eventually terminate after some number m of choices.

The technique used here is called a greedy algorithm because at each stage we select the largest remaining interval that is disjoint from the previously selected intervals.

Suppose $j \in \{1, \ldots, n\}$. To complete the proof, we must show that

$$I_j \subset (3 * I_{k_1}) \cup \cdots \cup (3 * I_{k_m}).$$

If $j \in \{k_1, \ldots, k_m\}$, then the inclusion above obviously holds.

Thus assume that $j \notin \{k_1, \ldots, k_m\}$. Because the process terminated without selecting j, the interval I_j is not disjoint from all of $I_{k_1}, \ldots, I_{k_m}$. Let I_{k_L} be the first interval on this list not disjoint from I_j; thus I_j is disjoint from $I_{k_1}, \ldots, I_{k_{L-1}}$. Because j was not chosen in step L, we conclude that $|I_{k_L}| \geq |I_j|$. Because $I_{k_L} \cap I_j \neq \emptyset$, this last inequality implies (easy exercise) that $I_j \subset 3 * I_{k_L}$, completing the proof. ∎

Hardy–Littlewood Maximal Inequality

Now we come to a brilliant definition that turns out to be extraordinarily useful.

4.6 Definition ***Hardy–Littlewood maximal function***

Suppose $h\colon \mathbf{R} \to \mathbf{R}$ is a Lebesgue measurable function. Then the *Hardy–Littlewood maximal function* of h is the function $h^*\colon \mathbf{R} \to [0, \infty]$ defined by

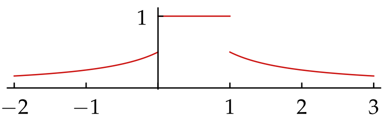
$$h^*(b) = \sup_{t>0} \frac{1}{2t} \int_{b-t}^{b+t} |h|.$$

In other words, $h^*(b)$ is the supremum over all bounded intervals centered at b of the average of $|h|$ on those intervals.

4.7 Example *Hardy–Littlewood maximal function of $\chi_{[0,1]}$*

As usual, let $\chi_{[0,1]}$ denote the characteristic function of the interval $[0, 1]$. Then

$$(\chi_{[0,1]})^*(b) = \begin{cases} \frac{1}{2(1-b)} & \text{if } b \leq 0, \\ 1 & \text{if } 0 < b < 1, \\ \frac{1}{2b} & \text{if } b \geq 1, \end{cases}$$



The graph of $(\chi_{[0,1]})^*$ on $[-2, 3]$.

as you should verify.

If $h\colon \mathbf{R} \to \mathbf{R}$ is Lebesgue measurable and $c \in \mathbf{R}$, then $\{b \in \mathbf{R} : h^*(b) > c\}$ is an open subset of $\mathbf{R}$, as you are asked to prove in Exercise 9 in this section. Thus h^* is a Borel measurable function.

Suppose $h \in \mathcal{L}^1(\mathbf{R})$ and $c > 0$. Markov's inequality (4.1) estimates the size of the set on which $|h|$ is larger than c. Our next result estimates the size of the set on which h^* is larger than c. The Hardy–Littlewood maximal inequality proved in the next result is a key ingredient in the proof of the Lebesgue Differentiation Theorem (4.10). Note that this next result is considerably deeper than Markov's inequality.

4.8 *Hardy–Littlewood maximal inequality*

Suppose $h \in \mathcal{L}^1(\mathbf{R})$. Then

$$|\{b \in \mathbf{R} : h^*(b) > c\}| \leq \frac{3}{c}\|h\|_1$$

for every $c > 0$.

Proof Suppose F is a closed bounded subset of $\{b \in \mathbf{R} : h^*(b) > c\}$. We will show that $|F| \leq \frac{3}{c}\int_{-\infty}^{\infty}|h|$, which implies our desired result [see Exercise 24(a) in Section 2D].

For each $b \in F$, there exists $t_b > 0$ such that

4.9
$$\frac{1}{2t_b}\int_{b-t_b}^{b+t_b}|h| > c.$$

Clearly

$$F \subset \bigcup_{b \in F}(b - t_b, b + t_b).$$

The Heine–Borel Theorem (2.12) tells us that this open cover of a closed bounded set has a finite subcover. In other words, there exist $b_1, \dots, b_n \in F$ such that

$$F \subset (b_1 - t_{b_1}, b_1 + t_{b_1}) \cup \dots \cup (b_n - t_{b_n}, b_n + t_{b_n}).$$

To make the notation cleaner, relabel the open intervals above as $I_1, \dots, I_n$.

Now apply the Vitali Covering Lemma (4.4) to the list $I_1, \dots, I_n$, producing a disjoint sublist $I_{k_1}, \dots, I_{k_m}$ such that

$$I_1 \cup \dots \cup I_n \subset (3 * I_{k_1}) \cup \dots \cup (3 * I_{k_m}).$$

Thus

$$\begin{aligned}
|F| &\leq |I_1 \cup \dots \cup I_n| \\
&\leq |(3 * I_{k_1}) \cup \dots \cup (3 * I_{k_m})| \\
&\leq |3 * I_{k_1}| + \dots + |3 * I_{k_m}| \\
&= 3(|I_{k_1}| + \dots + |I_{k_m}|) \\
&< \frac{3}{c}\Big(\int_{I_{k_1}}|h| + \dots + \int_{I_{k_m}}|h|\Big) \\
&\leq \frac{3}{c}\int_{-\infty}^{\infty}|h|,
\end{aligned}$$

where the second-to-last inequality above comes from 4.9 (note that $|I_{k_j}| = 2t_b$ for the choice of b corresponding to I_{k_j}) and the last inequality holds because $I_{k_1}, \dots, I_{k_m}$ are disjoint.

The last inequality completes the proof. ■

EXERCISES 4A

1 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $h\colon X \to \mathbf{R}$ is an $\mathcal{S}$-measurable function. Prove that

$$\mu(\{x \in X : |h(x)| \geq c\}) \leq \frac{1}{c^p} \int |h|^p \, d\mu$$

for all positive numbers c and p.

2 Suppose $(X, \mathcal{S}, \mu)$ is a measure space with $\mu(X) = 1$ and $h \in \mathcal{L}^1(\mu)$. Prove that

$$\mu\Bigl(\Bigl\{x \in X : \Bigl|h(x) - \int h \, d\mu\Bigr| \geq c\Bigr\}\Bigr) \leq \frac{1}{c^2}\Bigl(\int h^2 \, d\mu - \Bigl(\int h \, d\mu\Bigr)^2\Bigr)$$

for all $c > 0$.
[*The result above is called Chebyshev's inequality; it plays an important role in probability theory. Pafnuty Chebyshev* (*1821–1894*) *was Markov's thesis advisor.*]

3 Suppose $(X, \mathcal{S}, \mu)$ is a measure space. Suppose $h \in \mathcal{L}^1(\mu)$ and $\|h\|_1 > 0$. Prove that there is at most one number $c \in (0, \infty)$ such that

$$\mu(\{x \in X : |h(x)| \geq c\}) = \frac{1}{c}\|h\|_1.$$

4 Show that the constant 3 in the Vitali Covering Lemma (4.4) cannot be replaced by a smaller positive constant.

5 Prove the assertion left as an exercise in the last sentence of the proof of the Vitali Covering Lemma (4.4).

6 Verify the formula in Example 4.7 for the Hardy–Littlewood maximal function of $\chi_{[0,\,1]}$.

7 Find a formula for the Hardy–Littlewood maximal function of the characteristic function of $[0, 1] \cup [2, 3]$.

8 Find a formula for the Hardy–Littlewood maximal function of the function $h\colon \mathbf{R} \to [0, \infty)$ defined by

$$h(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

9 Suppose $h\colon \mathbf{R} \to \mathbf{R}$ is Lebesgue measurable. Prove that

$$\{b \in \mathbf{R} : h^*(b) > c\}$$

is an open subset of $\mathbf{R}$ for every $c \in \mathbf{R}$.

10 Prove or give a counterexample: If $h\colon \mathbf{R} \to [0,\infty)$ is an increasing function, then h^* is an increasing function.

11 Give an example of a Borel measurable function $h\colon \mathbf{R} \to [0,\infty)$ such that $h^*(b) < \infty$ for all $b \in \mathbf{R}$ but $\sup\{h^*(b) : b \in \mathbf{R}\} = \infty$.

12 Show that $|\{b \in \mathbf{R} : h^*(b) = \infty\}| = 0$ for every $h \in \mathcal{L}^1(\mathbf{R})$.

13 Show that there exists $h \in \mathcal{L}^1(\mathbf{R})$ such that $h^*(b) = \infty$ for every $b \in \mathbf{Q}$.

14 Suppose $h \in \mathcal{L}^1(\mathbf{R})$. Prove that

$$|\{b \in \mathbf{R} : h^*(b) \geq c\}| \leq \frac{3}{c}\|h\|_1$$

for every $c > 0$.

[*This result slightly strengthens the Hardy–Littlewood maximal inequality (4.8) because the set on the left side above includes those $b \in \mathbf{R}$ such that $h^*(b) = c$. A much deeper strengthening comes from replacing the constant 3 in the Hardy–Littlewood maximal inequality with a smaller constant. In 2003, Antonios Melas answered what had been an open question about the best constant. He proved that the smallest constant that can replace 3 in the Hardy–Littlewood maximal inequality is $(11 + \sqrt{61})/12 \approx 1.56752$; see* Annals of Mathematics *157 (2003), 647–688.*]

4B Derivatives of Integrals

Lebesgue Differentiation Theorem

The next result states that the average amount by which a function in $\mathcal{L}^1(\mathbf{R})$ differs from its values is small almost everywhere on small intervals. The 2 in the denominator of the fraction in the result below could be deleted, but its presence makes the length of the interval of integration nicely match the denominator $2t$.

The next result is called the Lebesgue Differentiation Theorem, even though no derivative is in sight. However, we will soon see how another version of this result deals with derivatives. The hard work takes place in the proof of this first version.

4.10 *Lebesgue Differentiation Theorem, first version*

Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Then

$$\lim_{t \downarrow 0} \frac{1}{2t} \int_{b-t}^{b+t} |f - f(b)| = 0$$

for almost every $b \in \mathbf{R}$.

Before getting to the formal proof of this first version of the Lebesgue Differentiation Theorem, we pause to provide some motivation for the proof. If $b \in \mathbf{R}$ and $t > 0$, then 3.25 gives the easy estimate

$$\frac{1}{2t} \int_{b-t}^{b+t} |f - f(b)| \leq \sup\{|f(x) - f(b)| : |x - b| \leq t\}.$$

If f is continuous at b, then the right side of this inequality has limit 0 as $t \downarrow 0$, proving 4.10 in the special case in which f is continuous on $\mathbf{R}$.

To prove the Lebesgue Differentiation Theorem, we will approximate an arbitrary function in $\mathcal{L}^1(\mathbf{R})$ by a continuous function (using 3.48). The previous paragraph shows that the continuous function has the desired behavior. We will use the Hardy–Littlewood maximal inequality (4.8) to show that the approximation produces approximately the desired behavior. Now we are ready for the formal details of the proof.

Proof of 4.10 Let $\delta > 0$. By 3.48, for each $k \in \mathbf{Z}^+$ there exists a continuous function $h_k \colon \mathbf{R} \to \mathbf{R}$ such that

4.11 $$\|f - h_k\|_1 < \frac{\delta}{k2^k}.$$

Let

$$B_k = \{b \in \mathbf{R} : |f(b) - h_k(b)| \leq \tfrac{1}{k} \text{ and } (f - h_k)^*(b) \leq \tfrac{1}{k}\}.$$

Then

4.12 $$\mathbf{R} \setminus B_k = \{b \in \mathbf{R} : |f(b) - h_k(b)| > \tfrac{1}{k}\} \cup \{b \in \mathbf{R} : (f - h_k)^*(b) > \tfrac{1}{k}\}.$$

Markov's inequality (4.1) as applied to the function $f - h_k$ and 4.11 imply that

4.13 $$|\{b \in \mathbf{R} : |f(b) - h_k(b)| > \tfrac{1}{k}\}| < \frac{\delta}{2^k}.$$

The Hardy–Littlewood maximal inequality (4.8) as applied to the function $f - h_k$ and 4.11 imply that

4.14 $$|\{b \in \mathbf{R} : (f - h_k)^*(b) > \tfrac{1}{k}\}| < \frac{3\delta}{2^k}.$$

Now 4.12, 4.13, and 4.14 imply that

$$|\mathbf{R} \setminus B_k| < \frac{\delta}{2^{k-2}}.$$

Let

$$B = \bigcap_{k=1}^{\infty} B_k.$$

Then

4.15 $$|\mathbf{R} \setminus B| = \Big|\bigcup_{k=1}^{\infty} (\mathbf{R} \setminus B_k)\Big| \le \sum_{k=1}^{\infty} |\mathbf{R} \setminus B_k| < \sum_{k=1}^{\infty} \frac{\delta}{2^{k-2}} = 4\delta.$$

Suppose $b \in B$ and $t > 0$. Then for each $k \in \mathbf{Z}^+$ we have

$$\begin{aligned}\frac{1}{2t}\int_{b-t}^{b+t} |f - f(b)| &\le \frac{1}{2t}\int_{b-t}^{b+t} \big(|f - h_k| + |h_k - h_k(b)| + |h_k(b) - f(b)|\big)\\ &\le (f - h_k)^*(b) + \Big(\frac{1}{2t}\int_{b-t}^{b+t} |h_k - h_k(b)|\Big) + |h_k(b) - f(b)|\\ &\le \frac{2}{k} + \frac{1}{2t}\int_{b-t}^{b+t} |h_k - h_k(b)|.\end{aligned}$$

Because h_k is continuous, the last term is less than $\frac{1}{k}$ for all $t > 0$ sufficiently close to 0 (how close is sufficiently close depends upon k). In other words, for each $k \in \mathbf{Z}^+$, we have

$$\frac{1}{2t}\int_{b-t}^{b+t} |f - f(b)| < \frac{3}{k}$$

for all $t > 0$ sufficiently close to 0.

Hence we conclude that

$$\lim_{t \downarrow 0} \frac{1}{2t}\int_{b-t}^{b+t} |f - f(b)| = 0$$

for all $b \in B$.

Let A denote the set of numbers $a \in \mathbf{R}$ such that

$$\lim_{t \downarrow 0} \frac{1}{2t}\int_{a-t}^{a+t} |f - f(a)|$$

either does not exist or is nonzero. We have shown that $A \subset (\mathbf{R} \setminus B)$. Thus

$$|A| \le |\mathbf{R} \setminus B| < 4\delta,$$

where the last inequality comes from 4.15. Because δ is an arbitrary positive number, the last inequality implies that $|A| = 0$, completing the proof. ■

Derivatives

You should remember the following definition from your calculus course.

4.16 Definition ***derivative***

Suppose $g\colon I \to \mathbf{R}$ is a function defined on an open interval I of $\mathbf{R}$ and $b \in I$. The *derivative* of g at b, denoted $g'(b)$, is defined by

$$g'(b) = \lim_{t \to 0} \frac{g(b+t) - g(b)}{t}$$

if the limit above exists, in which case g is called *differentiable* at b.

We now turn to the Fundamental Theorem of Calculus and a powerful extension that avoids continuity. These results show that differentiation and integration can be thought of as inverse operations.

You saw the next result in your calculus class, except now the function f is only required to be Lebesgue measurable (and its absolute value must have a finite Lebesgue integral). Of course, we also need to require f to be continuous at the crucial point b in the next result, because changing the value of f at a single number would not change the function g.

4.17 ***Fundamental Theorem of Calculus***

Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Define $g\colon \mathbf{R} \to \mathbf{R}$ by

$$g(x) = \int_{-\infty}^{x} f.$$

Suppose $b \in \mathbf{R}$ and f is continuous at b. Then g is differentiable at b and

$$g'(b) = f(b).$$

Proof If $t \neq 0$, then

$$\begin{aligned}
\left| \frac{g(b+t) - g(b)}{t} - f(b) \right| &= \left| \frac{\int_{-\infty}^{b+t} f - \int_{-\infty}^{b} f}{t} - f(b) \right| \\
&= \left| \frac{\int_{b}^{b+t} f}{t} - f(b) \right| \\
&= \left| \frac{\int_{b}^{b+t} \big(f - f(b)\big)}{t} \right| \qquad (4.18) \\
&\leq \sup_{\{x \in \mathbf{R}\,:\, |x-b| < |t|\}} |f(x) - f(b)|.
\end{aligned}$$

If $\varepsilon > 0$, then by the continuity of f at b, the last quantity is less than ε for t sufficiently close to 0. Thus g is differentiable at b and $g'(b) = f(b)$. ∎

A function in $\mathcal{L}^1(\mathbf{R})$ need not be continuous anywhere. Thus the Fundamental Theorem of Calculus (4.17) might provide no information about differentiating the integral of such a function. However, our next result states that all is well almost everywhere, even in the absence of any continuity of the function being integrated.

4.19 *Lebesgue Differentiation Theorem, second version*

Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Define $g\colon \mathbf{R} \to \mathbf{R}$ by

$$g(x) = \int_{-\infty}^{x} f.$$

Then $g'(b) = f(b)$ for almost every $b \in \mathbf{R}$.

Proof Suppose $t \neq 0$. Then from 4.18 we have

$$\begin{aligned}\left|\frac{g(b+t)-g(b)}{t} - f(b)\right| &= \left|\frac{\int_b^{b+t}\big(f - f(b)\big)}{t}\right| \\ &\leq \frac{1}{t}\int_b^{b+t}|f - f(b)| \\ &\leq \frac{1}{t}\int_{b-t}^{b+t}|f - f(b)|\end{aligned}$$

for all $b \in \mathbf{R}$. By the first version of the Lebesgue Differentiation Theorem (4.10), the last quantity has limit 0 as $t \to 0$ for almost every $b \in \mathbf{R}$. Thus $g'(b) = f(b)$ for almost every $b \in \mathbf{R}$. ■

Now we can answer the question raised on the opening page of this chapter.

4.20 *no set constitutes exactly half of each interval*

There does not exist a Lebesgue measurable set $E \subset [0,1]$ such that

$$|E \cap [0,b]| = \frac{b}{2}$$

for all $b \in [0,1]$.

Proof Suppose there does exist a Lebesgue measurable set $E \subset [0,1]$ with the property above. Define $g\colon \mathbf{R} \to \mathbf{R}$ by

$$g(b) = \int_{-\infty}^{b} \chi_E.$$

Thus $g(b) = \frac{b}{2}$ for all $b \in [0,1]$. Hence $g'(b) = \frac{1}{2}$ for all $b \in (0,1)$.

The Lebesgue Differentiation Theorem (4.19) implies that $g'(b) = \chi_E(b)$ for almost every $b \in \mathbf{R}$. However, χ_E never takes on the value $\frac{1}{2}$, which contradicts the conclusion of the previous paragraph. This contradiction completes the proof. ■

The next result says that a function in $\mathcal{L}^1(\mathbf{R})$ is equal almost everywhere to the limit of its average over small intervals. These two-sided results generalize more naturally to higher dimensions (take the average over balls centered at b) than the one-sided results.

4.21 $\mathcal{L}^1(\mathbf{R})$ *function equals its local average almost everywhere*

Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Then

$$f(b) = \lim_{t \downarrow 0} \frac{1}{2t} \int_{b-t}^{b+t} f$$

for almost every $b \in \mathbf{R}$.

Proof Suppose $t > 0$. Then

$$\begin{aligned} \left| \left(\frac{1}{2t} \int_{b-t}^{b+t} f \right) - f(b) \right| &= \left| \frac{1}{2t} \int_{b-t}^{b+t} \bigl(f - f(b)\bigr) \right| \\ &\leq \frac{1}{2t} \int_{b-t}^{b+t} |f - f(b)|. \end{aligned}$$

The desired result now follows from the first version of the Lebesgue Differentiation Theorem (4.10). ∎

Again, the conclusion of the result above holds at every number b at which f is continuous. The remarkable part of the result above is that even if f is discontinuous everywhere, the conclusion holds for almost every real number b.

Density

The next definition captures the notion of the proportion of a set in small intervals centered at a number b.

4.22 Definition *density*

Suppose $E \subset \mathbf{R}$. The *density* of E at a number $b \in \mathbf{R}$ is

$$\lim_{t \downarrow 0} \frac{|E \cap (b-t, b+t)|}{2t}$$

if this limit exists (otherwise the density of E at b is undefined).

4.23 Example *density of an interval*

$$\text{The density of } [0,1] \text{ at } b = \begin{cases} 1 & \text{if } b \in (0,1), \\ \frac{1}{2} & \text{if } b = 0 \text{ or } b = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The next beautiful result shows the power of the techniques developed in this chapter.

4.24 *Lebesgue Density Theorem*

Suppose $E \subset \mathbf{R}$ is a Lebesgue measurable set. Then the density of E is 1 at almost every element of E and is 0 at almost every element of $\mathbf{R} \setminus E$.

Proof First suppose $|E| < \infty$. Thus $\chi_E \in \mathcal{L}^1(\mathbf{R})$. Because

$$\frac{|E \cap (b-t, b+t)|}{2t} = \frac{1}{2t} \int_{b-t}^{b+t} \chi_E$$

for every $t > 0$ and every $b \in \mathbf{R}$, the desired result follows immediately from 4.21.

Now consider the case where $|E| = \infty$ [which means that $\chi_E \notin \mathcal{L}^1(\mathbf{R})$ and hence 4.21 as stated cannot be used]. For $k \in \mathbf{Z}^+$, let $E_k = E \cap (-k, k)$. If $|b| < k$, then the density of E at b equals the density of E_k at b. By the previous paragraph as applied to E_k, there are sets $F_k \subset E_k$ and $G_k \subset \mathbf{R} \setminus E_k$ such that $|F_k| = |G_k| = 0$ and the density of E_k equals 1 at every element of $E_k \setminus F_k$ and the density of E_k equals 0 at every element of $(\mathbf{R} \setminus E_k) \setminus G_k$.

Let $F = \bigcup_{k=1}^{\infty} F_k$ and $G = \bigcup_{k=1}^{\infty} G_k$. Then $|F| = |G| = 0$ and the density of E is 1 at every element of $E \setminus F$ and is 0 at every element of $(\mathbf{R} \setminus E) \setminus G$. ■

The bad Borel set provided by the next result leads to a bad Borel measurable function. Specifically, let E be the bad Borel set in 4.25. Then χ_E is a Borel measurable function that is discontinuous everywhere. Furthermore, the function χ_E cannot be modified on a set of measure 0 to be continuous anywhere (in contrast to the function $\chi_{\mathbf{Q}}$).

The Lebesgue Density Theorem makes the example provided by the next result somewhat surprising. Be sure to spend some time pondering why the next result does not contradict the Lebesgue Density Theorem. Also, compare the next result to 4.20.

Even though the function χ_E discussed in the paragraph above is continuous nowhere and every modification of this function on a set of measure 0 is also continuous nowhere, the function g defined by

$$g(b) = \int_0^b \chi_E$$

is differentiable almost everywhere (by 4.19).

The proof of 4.25 given below is based on an idea of Walter Rudin.

4.25 *bad Borel set*

There exists a Borel set $E \subset \mathbf{R}$ such that

$$0 < |E \cap I| < |I|$$

for every nonempty bounded open interval I.

Proof We use the following fact in our construction:

4.26 Suppose G is a nonempty open subset of $\mathbf{R}$. Then there exists a closed set $F \subset G \setminus \mathbf{Q}$ such that $|F| > 0$.

To prove 4.26, let J be a closed interval contained in G such that $0 < |J|$. Let $r_1, r_2, \dots$ be a list of all the rational numbers. Let

$$F = J \setminus \bigcup_{k=1}^{\infty} \Bigl(r_k - \frac{|J|}{2^{k+2}}, r_k + \frac{|J|}{2^{k+2}} \Bigr).$$

Then F is a closed subset of $\mathbf{R}$ and $F \subset J \setminus \mathbf{Q} \subset G \setminus \mathbf{Q}$. Also, $|J \setminus F| \leq \frac{1}{2}|J|$ because $J \setminus F \subset \bigcup_{k=1}^{\infty} \Bigl(r_k - \frac{|J|}{2^{k+2}}, r_k + \frac{|J|}{2^{k+2}} \Bigr)$. Thus

$$|F| = |J| - |J \setminus F| \geq \tfrac{1}{2}|J| > 0,$$

completing the proof of 4.26.

To construct the set E with the desired properties, let $I_1, I_2, \dots$ be a sequence consisting of all nonempty bounded open intervals of $\mathbf{R}$ with rational endpoints. Let $F_0 = \hat{F}_0 = \varnothing$, and inductively construct sequences $F_1, F_2, \dots$ and $\hat{F}_1, \hat{F}_2, \dots$ of closed subsets of $\mathbf{R}$ as follows: Suppose $n \in \mathbf{Z}^+$ and $F_0, \dots, F_{n-1}$ and $\hat{F}_0, \dots, \hat{F}_{n-1}$ have been chosen as closed sets that contain no rational numbers. Thus

$$I_n \setminus (\hat{F}_0 \cup \dots \cup \hat{F}_{n-1})$$

is a nonempty open set (nonempty because it contains all rational numbers in I_n). Applying 4.26 to the open set above, we see that there is a closed set F_n contained in the set above such that F_n contains no rational numbers and $|F_n| > 0$. Applying 4.26 again, but this time to the open set

$$I_n \setminus (F_0 \cup \dots \cup F_n),$$

which is nonempty because it contains all rational numbers in I_n, we see that there is a closed set $\hat{F}_n$ contained in the set above such that $\hat{F}_n$ contains no rational numbers and $|\hat{F}_n| > 0$.

Now let

$$E = \bigcup_{k=1}^{\infty} F_k.$$

Our construction implies that $F_k \cap \hat{F}_n = \varnothing$ for all $k, n \in \mathbf{Z}^+$. Thus $E \cap \hat{F}_n = \varnothing$ for all $n \in \mathbf{Z}^+$. Hence $\hat{F}_n \subset I_n \setminus E$ for all $n \in \mathbf{Z}^+$.

Suppose I is a nonempty bounded open interval. Then $I_n \subset I$ for some $n \in \mathbf{Z}^+$. Thus

$$0 < |F_n| \leq |E \cap I_n| \leq |E \cap I|.$$

Also,

$$|E \cap I| = |I| - |I \setminus E| \leq |I| - |I_n \setminus E| \leq |I| - |\hat{F}_n| < |I|,$$

completing the proof. ■

EXERCISES 4B

For $f \in \mathcal{L}^1(\mathbf{R})$ and I an interval of $\mathbf{R}$ with $0 < |I| < \infty$, let f_I denote the average of f on I. In other words, $f_I = \frac{1}{|I|}\int_I f$.

1 Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Prove that

$$\lim_{t \downarrow 0} \frac{1}{2t} \int_{b-t}^{b+t} |f - f_{[b-t,\,b+t]}| = 0$$

for almost every $b \in \mathbf{R}$.

2 Suppose $f \in \mathcal{L}^1(\mathbf{R})$. Prove that

$$\lim_{t \downarrow 0} \sup\Bigl\{\frac{1}{|I|}\int_I |f - f_I| : I \text{ is an interval of length } t \text{ containing } b\Bigr\} = 0$$

for almost every $b \in \mathbf{R}$.

3 Suppose $f\colon \mathbf{R} \to \mathbf{R}$ is a Lebesgue measurable function such that $f^2 \in \mathcal{L}^1(\mathbf{R})$. Prove that

$$\lim_{t \downarrow 0} \frac{1}{2t} \int_{b-t}^{b+t} |f - f(b)|^2 = 0$$

for almost every $b \in \mathbf{R}$.

4 Prove that the Lebesgue Differentiation Theorem (4.19) still holds if the hypothesis that $\int_{-\infty}^{\infty} |f| < \infty$ is weakened to the requirement that $\int_{-\infty}^{x} |f| < \infty$ for all $x \in \mathbf{R}$.

5 Suppose $f\colon \mathbf{R} \to \mathbf{R}$ is a Lebesgue measurable function. Prove that

$$|f(b)| \le f^*(b)$$

for almost every $b \in \mathbf{R}$.

6 Prove that if $h \in L^1(\mathbf{R})$ and $\int_{-\infty}^{s} h = 0$ for all $s \in \mathbf{R}$, then $h = 0$.

7 Give an example of a Borel subset of $\mathbf{R}$ whose density at 0 is not defined.

8 Give an example of a Borel subset of $\mathbf{R}$ whose density at 0 is $\frac{1}{3}$.

9 Prove that if $t \in [0,1]$, then there exists a Borel set $E \subset \mathbf{R}$ such that the density of E at 0 is t.

10 Suppose E is a Lebesgue measurable subset of $\mathbf{R}$ such that the density of E equals 1 at every element of E and equals 0 at every element of $\mathbf{R} \setminus E$. Prove that $E = \emptyset$ or $E = \mathbf{R}$.