Real Analysis Exchange
Vol. 29(2), 2003/2004, pp. 957–973

Brian S. Thomson, Mathematics Department, Simon Fraser University, B. C., Canada V5A 1S6. email: thomson@cs.sfu.ca

VITALI COVERINGS AND LEBESGUE'S DIFFERENTIATION THEOREM

Abstract

The standard techniques used to prove the Lebesgue differentiation theorem (that monotonic functions are a.e. differentiable) are presented in an unusual way that reveals more about their nature and allows greater generality.

1 Introduction

We have just seen in these pages two elementary proofs of the Lebesgue differentiation theorem. Claude-Alain Faure [2] has presented a clean and elegant exposition using a rising-sun argument. The presentation is particularly elementary in the sense that minimal apparatus from measure theory is needed: only the most rudimentary properties of Lebesgue outer measure are used. Since the rising sun lemma has several other applications and a very natural geometric appeal this would be well worth presenting at an elementary level.

John Hagood's [4] presentation of the same theorem depends on a clever new covering lemma that is a variant on the Heine-Borel theorem. This proof is equally elementary, needing only a few facts about measurable sets and Lebesgue measure. Again by appealing only to a familiar compactness argument the proof is accessible and attractive at a beginning level.

In particular, in both presentations the Vitali covering theorem is not invoked, although the usual growth lemmas (cf. [1, Lemma 7.1 and Lemma 7.4]) are proved and the usual device used to complete the proof. The question arises as to how to present the "nonelementary" version: if all the apparatus of measure theory including the Vitali theorem can be used, how should one present a proof of Lebesgue's theorem?

Key Words: Lebesgue Differentiation theorem, Vitali covering theorem, variational measures
Mathematical Reviews subject classification: 26A45, 26A24
Received by the editors February 19, 2004
Communicated by: Clifford E. Weil

The goal of this paper is to present this program from a different perspective than the reader is likely to have encountered. The Vitali theorem itself is (I hope) unrecognizable as it appears here and the Lebesgue differentiation theorem is stated for a larger class of functions than monotonic functions. The Vitali covering theorem for Lebesgue measure on the real line is taken for granted and is the only deep result needed. From it we can deduce much about the differentiation and variational structure of real functions.

One advantage we find in such a presentation is that it allows Lebesgue's program to be carried forward to a more general question. We know that a function of bounded variation is a.e. differentiable. What variational properties should a function have on a set E to enable us to deduce its a.e. differentiability on E? Since our growth lemmas apply to a general class of functions we can answer this question in an economical way.

2 Full and Fine Covers, Partitions and Subpartitions

Vitali covers and a dual notion of covering that is stronger than Vitali coverings are presented by way of *covering relations*, i.e., collections whose elements are pairs $([x,y],z)$ with $x<y$ and $z\in[x,y]$.

We say that a covering relation β is a *full cover* of a set E if for every $z\in E$ there is a $\delta>0$ so that every pair $([x,y],z)$ with $0<y-x<\delta$ and $z\in[x,y]$ must belong to β. We say that a covering relation β is a *fine cover* of a set E if for every $z\in E$ and every $\delta>0$ there must exist at least one pair $([x,y],z)$ with $0<y-x<\delta$ and $z\in[x,y]$ that belongs to β.

The full and fine covers play a dual role and can be expressed in a way that reveals a genuine dual structure. The following pair of theorems (whose proofs are left to the reader) show this.

Theorem 2.1. *A necessary and sufficient condition for a covering relation β to be a full cover of a set E is that for every fine cover β_1 of E and every $z\in E$ there is at least one pair $([x,y],z)$ in $\beta\cap\beta_1$.*

Theorem 2.2. *A necessary and sufficient condition for a covering relation β to be a fine cover of a set E is that for every full cover β_1 of E and every $z\in E$ there is at least one pair $([x,y],z)$ in $\beta\cap\beta_1$.*

A finite covering relation π

$$\pi=\{([x_i,y_i],z_i):i=1,2,3,\dots,n\}$$

is a *partition* of $[a,b]$ if

a. any two distinct $([x_1, y_1], z_1)$, $([x_2, y_2], z_2)$ elements of π require nonoverlapping $[x_1, y_1]$ and $[x_2, y_2]$, and

b. $\bigcup_{([x,y],z)\in\pi}[x, y] = [a, b]$.

Partitions themselves play no role in the present theory, but arbitrary subsets shall. Any subset π of a partition is a *subpartition.* For a subpartition

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \dots, n\}$$

we write $\ell(\pi) = \sum_{i=1}^{n}(y_i - x_i)$ and refer to this as the *total length* of the subpartition (i.e., it is the total length of the intervals that appear in the subpartition).

3 Definition of the Measures $\mathcal{L}$ and $\mathcal{L}_*$

We present here two equivalent formulations of the Lebesgue outer measure on the real line, formulations which arise from the Vitali theorem. The two measures that shall play a role in the statement of the Vitali covering theorem are defined by using estimates of the total length $\ell(\pi)$ of subpartitions contained in full and fine covers of the set to be measured. If β is an arbitrary covering relation write $\ell(\beta) = \sup_{\pi\subset\beta} \ell(\pi)$ where the supremum is taken over all subpartitions π that are contained in β.

Then for any set E of real numbers we write

$$\mathcal{L}(E) = \inf_{\beta\ full} \ell(\beta) \text{ and } \mathcal{L}_*(E) = \inf_{\beta\ fine} \ell(\beta)$$

where the infima are taken over all covering relations β that are full covers of E (for the first measure) and fine covers of E (for the second measure).

Lemma 3.1. *$\mathcal{L}$ and $\mathcal{L}_*$ are metric outer measures on the real line.*

The proof is not difficult and is, in any case, well-known, although the ideas are often expressed in different language. The most elegant and compact treatment of metric outer measures is Federer [3] (where they are called simply "metric measures"); a more leisurely treatment is given in [1, Chap. 3].

4 Vitali Covering Theorem

Let λ denote the Lebesgue outer measure on the real line. Our version of the covering theorem is the assertion that the measures λ, $\mathcal{L}$ and $\mathcal{L}_*$ are identical.

Theorem 4.1 (Vitali Covering Theorem)**.** $\lambda = \mathcal{L} = \mathcal{L}_*$.

Rather than prove the theorem we shall show that it is equivalent to the usual form of that theorem, namely the following statement:

[VCT] *Let E be a bounded set, $\varepsilon > 0$ and β any fine cover of E. Then there exists a subpartition $\pi \subset \beta$,*

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \dots, n\},$$

so that

$$\lambda\left(E \setminus \bigcup_{i=1}^{n} [x_i, y_i]\right) < \varepsilon.$$

Let E be an arbitrary bounded set. We show, using [VCT] and elementary properties of the Lebesgue measure, that $\lambda(E) = \mathcal{L}(E) = \mathcal{L}_*(E)$; from this it follows that the three measures agree on all sets.

Since every full cover of E is also a fine cover of E it follows immediately from the definitions that $\mathcal{L}_*(E) \leq \mathcal{L}(E)$. Let $\varepsilon > 0$ and, using properties of λ, select an open set G containing E so that $\lambda(G) < \lambda(E) + \varepsilon$. Let β denote the covering relation consisting of all elements $([x, y], z)$ with $z \in E$ for which $[x, y] \subset G$. Observe that β is a full cover of E. For any subpartition

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \dots, n\}$$

contained in β we have each $[x_i, y_i] \subset G$ and hence

$$\ell(\pi) = \sum_{i=1}^{n} [y_i - x_i] \leq \lambda(G) < \lambda(E) + \varepsilon.$$

From this and the way in which $\mathcal{L}(E)$ is defined it follows that $\mathcal{L}(E) \leq \lambda(E) + \varepsilon$. As ε is an arbitrary positive number we obtain one more inequality: $\mathcal{L}(E) \leq \lambda(E)$.

The final inequality which will provide the identity among the three measures is obtained from [VCT]: suppose that we are given an arbitrary fine cover β of E. Then by [VCT] we may select a subpartition $\pi \subset \beta$,

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \dots, n\},$$

so that

$$\lambda\left(E \setminus \bigcup_{i=1}^{n} [x_i, y_i]\right) < \varepsilon.$$

Thus

$$\lambda(E) \leq \lambda\left(E \setminus \bigcup_{i=1}^{n} [x_i, y_i]\right) + \sum_{i=1}^{n} [y_i - x_i] \leq \ell(\pi) + \varepsilon.$$

Since this is true for any fine cover β of E and any ε, we must have $\lambda(E) \leq \mathcal{L}_*(E)$.

Putting our three inequalities together gives us

$$\lambda(E) \leq \mathcal{L}_*(E) \leq \mathcal{L}(E) \leq \lambda(E)$$

and the theorem follows, at least for bounded sets. Usual measure theoretic arguments show that the identity is true for all sets.

Conversely let us now show that the identity of the three measures $\mathcal{L}$, $\mathcal{L}_*$ and λ can be used to establish the the assertion [VCT].

For let E be a bounded set and let $\varepsilon > 0$. Choose G open so that $E \subset G$ and $\lambda(G) < \lambda(E) + \varepsilon/2$. Let β be an arbitrary fine cover of E. Let β_1 denote the collection of all elements $([x, y], z)$ of β for which $[x, y] \subset G$. Observe that β_1 is also a fine cover of E. By the definition of $\mathcal{L}_*(E)$ there must be a subpartition $\pi \subset \beta_1$,

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \ldots, n\}$$

so that $\mathcal{L}_*(E) < \sum_{i=1}^{n}(y_i - x_i) + \varepsilon/2$. Using familiar properties of Lebesgue measure and the fact that each interval $[x_i, y_i]$ is a subset of G, we obtain

$$\lambda\left(G \setminus \bigcup_{i=1}^{n}[x_i, y_i]\right) = \lambda(G) - \sum_{i=1}^{n}[y_i - x_i].$$

Finally then, using the identity $\mathcal{L}_* = \lambda$ and the computations above, we have the inequality in [VCT] that we require:

$$\lambda\left(E \setminus \bigcup_{i=1}^{n}[x_i, y_i]\right) \leq \lambda\left(G \setminus \bigcup_{i=1}^{n}[x_i, y_i]\right) \leq (\mathcal{L}_*(E) + \varepsilon/2) - \sum_{i=1}^{n}[y_i - x_i] < \varepsilon.$$

5 The Full and Fine Total Variation Measures of a Function

We study the differentiation properties of a function F defined on the real line, by defining measures analogous to the two measures $\mathcal{L}$ and $\mathcal{L}_*$ that carry the variational information about F.

For a subpartition

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \ldots, n\}$$

we write

$$V(F, \pi) = \sum_{i=1}^{n} |F(y_i) - F(x_i)|$$

and refer to this as the *total variation of* F on the subpartition. For an arbitrary covering relation β we extend this notation by writing also

$$V(F,\beta) = \sup_{\pi \subset \beta} V(F,\pi)$$

where the supremum is taken over all subpartitions π that are contained in β.

The two measures that are used in the statement of the Vitali property for a function F are defined by using estimates of the total variation $V(F,\pi)$ of subpartitions contained in full and fine covers of the set to be measured. For any set E of real numbers we write

$$\mathcal{L}^F(E) = \inf_{\beta \; full} V(F,\beta) \text{ and } \mathcal{L}_*^F(E) = \inf_{\beta \; fine} V(F,\beta)$$

where, exactly as before, the infima are taken over all covering relations β that are full covers of E (for the first measure) and fine covers of E (for the second measure).

Lemma 5.1. *$\mathcal{L}^F$ and $\mathcal{L}_*^F$ are metric outer measures on the real line.*

6 Derived Numbers

The analysis of the differentiation structure of a function F can often be carried out quite transparently by using the notion of a derived number. A number r (including $\pm\infty$) is a *derived number* of F at a point z provided there is some sequence $x_n \to z$ ($x_n \neq z$) for which

$$\frac{F(x_n) - F(z)}{x_n - z} \to r.$$

The nature of full and fine covers forces on us also a weaker form of derived number to analyze: A number r (including $\pm\infty$) is a $*$-*derived number* of F at a point z provided there are sequences $x_n \to z$, $y_n \to z$, $x_n \leq z \leq y_n$ ($x_n \neq y_n$) for which

$$\frac{F(x_n) - F(y_n)}{x_n - y_n} \to r.$$

Related to this too are the values:

$$\overline{d}(F,z) = \inf_{\delta>0} \sup \left\{ \left| \frac{F(x) - F(y)}{x - y} \right| : x \leq z \leq y,\ 0 < y - x < \delta \right\}$$

and

$$\underline{d}(F,z) = \sup_{\delta>0} \inf \left\{ \left| \frac{F(x) - F(y)}{x - y} \right| : x \leq z \leq y,\ 0 < y - x < \delta \right\}.$$

Some immediate or elementary observations are needed and will also help to clarify the concepts.

a. If $F(x) = |x|$ then, at the point $z = 0$, the only derived numbers are 1 and -1 but every number in $[-1, 1]$ is a $*$-derived number of F at 0, $\underline{d}(F, 0) = 0$ and $\overline{d}(F, 0) = 1$.

b. Every derived number is also a $*$-derived number.

c. $F'(z) = r$ (including $r = \pm\infty$) if and only if r is the only $*$-derived number of F at z.

d. $\underline{d}(F, x)$ and $\overline{d}(F, x)$ are Borel functions.

e. If r is a $*$-derived number at z of F then $\underline{d}(F, z) \leq |r| \leq \overline{d}(F, z)$.

f. $F'(z) = r$ implies that $\underline{d}(F, z) = \overline{d}(F, z) = |r|$.

g. Conversely to (f), if $\underline{d}(F, z) = \overline{d}(F, z) = |r| \neq \infty$ then F is differentiable at z with $|F'(z)| = r$.

h. The identity $\underline{d}(F, z) = \overline{d}(F, z) = \infty$ does not imply in general that $|F'(z)| = \infty$.

i. If F is also continuous or monotonic then $\underline{d}(F, z) = \infty$ *does* imply that $|F'(z)| = \infty$.

Assertion (g) is easy if you also assume that F is everywhere continuous or is monotonic. Without one of those assumptions (g) requires a geometric argument to show that there cannot exist a function with both $+r$ and $-r$ as $*$-derived numbers and no other values.

7 Growth Lemmas

The pair of growth lemmas [1, Lemma 7.1 and Lemma 7.4] mentioned in the introduction are restricted to monotonic functions and have proofs which depend on the Vitali covering theorem. The growth lemmas we now present apply to any function and, making no appeal to the Vitali covering theorem, are entirely elementary. By stating them for the measures $\mathcal{L}$ and $\mathcal{L}_*$ rather than for λ we are stressing that their proof is not using the Vitali theorem, although we shall certainly take advantage of the identity $\mathcal{L} = \mathcal{L}_* = \lambda$ when we need to.

Lemma 7.1. *If $\underline{d}(F, z) < r$ for every $z \in E$ then $\mathcal{L}_*^F(E) \leq r\mathcal{L}(E)$.*

Proof. Assume that $\mathcal{L}(E) < \infty$ and choose any t so that $\mathcal{L}(E) < t$. Choose a full cover β of E so that $\ell(\beta) < t$. Let β_1 denote the collection of all pairs $([x,y],z)$ for which $z \in E$ and $|F(y) - F(z)| < r|y-z|$. It is easy to check that β_1 is a fine cover of E and that $\beta \cap \beta_1$ is too. Note that if

$$\pi = \{([x_i, y_i], z_i) : i = 1, 2, 3, \ldots, n\}$$

is a subpartition contained in $\beta \cap \beta_1$ then

$$V(F,\pi) = \sum_{i=1}^{n} |F(y_i) - F(x_i)| \le \sum_{i=1}^{n} r(y_i - x_i) \le r\ell(\beta) < rt.$$

From this it follows that $\mathcal{L}^F_*(E) \le V(F, \beta \cap \beta_1) \le rt$, and the conclusion of the lemma then follows. □

There are three other growth lemmas corresponding to the remaining configurations in the associated inequality. The proofs are nearly identical except for swapping full for fine or reversing an inequality. We give the details as these lemmas, even while elementary, are the key tools in the theory.

Lemma 7.2. *If $\overline{d}(F,z) > r > 0$ for every $z \in E$ then $r\mathcal{L}_*(E) \le \mathcal{L}^F(E)$.*

Proof. Assume that $\mathcal{L}^F(E) < \infty$ and choose any t so that $\mathcal{L}^F(E) < t$. Choose a full cover β of E so that $V(F,\beta) < t$. Let β_1 denote the collection of all pairs $([x,y],z)$ for which $z \in E$ and $|F(y) - F(z)| > r|y-z|$. It is easy to check that β_1 is a fine cover of E and that $\beta \cap \beta_1$ is too. Arguing as in Lemma 7.1 we note that $\ell(\pi) < r^{-1}t$ for all subpartitions $\pi \subset \beta \cap \beta_1$. From this it follows that $\mathcal{L}_*(E) \le \ell(\beta \cap \beta_1) \le r^{-1}t$, and the conclusion of the lemma then follows. □

Lemma 7.3. *If $\overline{d}(F,z) < r$ for every $z \in E$ then $\mathcal{L}^F(E) \le r\mathcal{L}(E)$.*

Proof. Assume that $\mathcal{L}(E) < \infty$ and choose any t so that $\mathcal{L}(E) < t$. Choose a full cover β of E so that $\ell(\beta) < t$. Let β_1 denote the collection of all pairs $([x,y],z)$ for which $z \in E$ and $|F(y) - F(z)| < r|y-z|$. It is easy to check that β_1 is a full cover of E and that $\beta \cap \beta_1$ is too. As before, we obtain

$$\mathcal{L}^F(E) \le V(F, \beta \cap \beta_1) \le rt,$$

and the conclusion of the lemma then follows. □

Lemma 7.4. *If $\underline{d}(F,z) > r > 0$ for every $z \in E$ then $r\mathcal{L}(E) \le \mathcal{L}^F(E)$ and $r\mathcal{L}_*(E) \le \mathcal{L}^F_*(E)$.*

Proof. Assume that $\mathcal{L}_*^F(E) < \infty$ and choose any t so that $\mathcal{L}_*^F(E) < t$. Choose a fine cover β of E so that $V(F,\beta) < t$. Let β_1 denote the collection of all pairs $([x,y],z)$ for which $z \in E$ and $|F(y) - F(z)| > r|y-z|$. It is easy to check that β_1 is a full cover of E and that $\beta \cap \beta_1$ is then a fine cover of E. Repeating our technique once again, we show that

$$\mathcal{L}_*(E) \leq \ell(\beta \cap \beta_1) \leq r^{-1}t,$$

and the conclusion $r\mathcal{L}_*(E) \leq \mathcal{L}_*^F(E)$ of the lemma then follows. The assertion that $r\mathcal{L}(E) \leq \mathcal{L}^F(E)$ is proved the same way substituting "full" for "fine" throughout. □

Lemma 7.5. *If $\underline{d}(F,z) = \infty$ for every $z \in E$ then $\mathcal{L}_*(E) = 0$.*

Proof. Note that if $\mathcal{L}_*^F(E) < \infty$ then $\mathcal{L}_*(E) = 0$ is an immediate consequence of Lemma 7.4. The same is true if $\mathcal{L}_*^F$ is σ-finite. (For a function F that is continuous or which has the Darboux property one can prove that $\mathcal{L}_*^F$ must be σ-finite.)

This is enough for the purposes of this paper and as far as we can go with the simple Vitali arguments of this section. Nonetheless we point out that the lemma is true in general. Appealing to the methods in Saks [5, p. 270] one can prove that for any function F the set of points z where $\underline{d}(F,z) = \infty$ has measure zero. The essential feature of the argument is that the graph of the function has at every such point $(z, f(z))$ a vertical tangent and the projection of that set onto the x-axis has then measure zero. □

8 Functions with the Vitali Property

Now, in contrast to the Vitali covering theorem, the identity of $\mathcal{L}^F$ and $\mathcal{L}_*^F$ becomes a definition describing the class of functions that are of interest in the theory. Note that, in general, $\mathcal{L}_*^F \leq \mathcal{L}^F$ because every full cover is also a fine cover. When there is equality there are considerable implications about the differentiation properties of F and that is what motivates the definition.

Definition 8.1. *A function F is said to have the* Vitali property *on a Borel set E if the outer measures $\mathcal{L}^F$ and $\mathcal{L}_*^F$ agree on every Borel subset of E.*

Our first main theorem shows how intimately related is the Vitali property of a function to its differentiation properties.

Theorem 8.2. *Let F have a finite derivative at every point of a Borel set E. Then F has the Vitali property on E and, moreover,*

$$\mathcal{L}^F(E) = \mathcal{L}_*^F(E) = \int_E |F'(x)|\,dx.$$

PROOF. We can suppose that E is bounded making some of the computations more transparent. Let $r > 1$. Since F has a derivative at every point z in E we can use the fact that $\underline{d}(F, z) = \overline{d}(F, z) = |F'(z)|$. Use the notation

$$E_k = \{x \in E : r^{k-1} < |F'(x)| \leq r^k\}$$

and

$$Z = \{x \in E : |F'(x)| = 0\}.$$

Then E is expressed as a countable disjointed union of a sequence of Borel subsets, namely the sets Z, $\{E_n\}$ $(n = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots)$.

Note first that $\mathcal{L}^F(Z) = 0$. This follows easily from the third growth lemma (Lemma 7.3) since, for any $\varepsilon > 0$, we obtain $\mathcal{L}^F(Z) \leq \varepsilon\mathcal{L}(Z) \leq \varepsilon\mathcal{L}(E)$ and we have assumed that E is bounded so that its Lebesgue measure is finite.

We know then, since these are metric outer measures and all these sets are Borel, that

$$\mathcal{L}_*^F(E) = \sum_{k=-\infty}^{\infty} \mathcal{L}_*^F(E_k) \text{ and } \mathcal{L}^F(E) = \sum_{k=-\infty}^{\infty} \mathcal{L}^F(E_k).$$

Thus we have, using once again Lemma 7.3 and the identity of $\mathcal{L}$ and λ that

$$\mathcal{L}^F(E) = \sum_{k=-\infty}^{\infty} \mathcal{L}^F(E_k) \leq \sum_{k=-\infty}^{\infty} r^k \mathcal{L}(E_k)$$
$$\leq r\left(\sum_{k=-\infty}^{\infty} \int_{E_k} |F'(x)|\,dx\right) = r\int_E |F'(x)|\,dx.$$

Similarly using Lemma 7.4 and the identity of $\mathcal{L}_*$ and λ we obtain that

$$\mathcal{L}_*^F(E) \geq \sum_{k=-\infty}^{\infty} \mathcal{L}_*^F(E_k) \geq \sum_{k=-\infty}^{\infty} r^{k-1}\mathcal{L}_*(E_k) \geq r^{-1}\int_E |F'(x)|\,dx.$$

Since these two inequalities are true for all $r > 1$ and since in general $\mathcal{L}_*^F(E) \leq \mathcal{L}^F(E)$ the identity of the theorem must follow. As this identity holds as well for all Borel subsets of E we have also established that F has the Vitali property on E. □

9 Lebesgue's Differentiation Theorem

We can now give our version of Lebesgue's theorem, stated not merely for monotonic functions but for any function possessing the Vitali property. The proof is quite simply the usual one (cf. [1, p. 288] or [2, Theorem 9]) using the growth lemmas Lemma 7.1 and 7.2, which now apply to any function. The Vitali-type identities $\mathcal{L}^F = \mathcal{L}^F_*$ and $\mathcal{L} = \mathcal{L}_* = \lambda$ supply the final step.

Theorem 9.1. *Let F have the Vitali property on a Borel set E. Then F has a finite derivative at $\mathcal{L}$-almost every point of E and at $\mathcal{L}^F$-almost every point z either F has a finite derivative $F'(z)$ or else $\underline{d}(F,z) = \infty$.*

Proof. It is enough to prove the theorem under the assumption that E is a bounded Borel set. We examine the Borel subset

$$A = \{z \in E : \underline{d}(F,z) < \overline{d}(F,z)\}.$$

As usual in arguments of this type introduce rational numbers $0 < r < s$ and the further Borel subsets

$$A_{rs} = \{z \in A : \underline{d}(F,z) < r < s < \overline{d}(F,z)\}.$$

Since A is the countable union of this collection of sets (taken over all rationals r and s with $r < s$) the measure arguments are simple.

By the two growth lemmas and the identity $\mathcal{L} = \mathcal{L}_* = \lambda$ we easily obtain

$$\mathcal{L}^F_*(E_{rs}) \leq r\mathcal{L}(E_{rs}) \leq s\mathcal{L}_*(E_{rs}) \leq \mathcal{L}^F(E_{rs}).$$

Our assumption that F has the Vitali property on E gives the identity $\mathcal{L}^F = \mathcal{L}^F_*$ on Borel subsets of E. None of these numbers are infinite, $r < s$ and so the inequality makes sense only in the case that $\mathcal{L}^F(A_{rs}) = \mathcal{L}(A_{rs}) = 0$. Consequently $\mathcal{L}^F(A) = \mathcal{L}(A) = 0$.

At every point z in $E \setminus A$ we know that either $\underline{d}(F,z) = \overline{d}(F,z) = r$ (finite) or else $\underline{d}(F,z) = \overline{d}(F,z) = +\infty$. In the former case (as observed in Section 6) $F'(z)$ exists and is $\pm r$. Thus we have proved that at each point $z \in E \setminus A$ either $F'(z)$ exists (finitely) or else $\underline{d}(F,z) = +\infty$. Lemma 7.5 then completes the proof since the set of points at which $\underline{d}(F,z) = +\infty$ has measure zero. □

Remark. In Saks [5, p. 125 and p. 230], proving similar theorems, one finds the conclusion that the function F has a derivative (possibly infinite) everywhere except at a set N for which $\lambda(N) = \lambda(F(N)) = 0$. In the version here the exceptional set has $\lambda(N) = \mathcal{L}^F(N) = 0$. In general $\lambda(F(N)) \leq \mathcal{L}^F(N))$ so it would seem that we have proved a sharper result. However if F is assumed to be continuous and VBG$_*$ on a set N then the two assertions $\lambda(F(N)) = 0$ and $\mathcal{L}^F(N) = 0$ are known to be equivalent. Since the condition VBG$_*$ (as we shall see below) is exactly equivalent to the Vitali property we might better present the exceptional set as having the property

$$\lambda(N) = \lambda(F(N)) = \mathcal{L}^F(N)) = 0$$

with the understanding that the latter two parts of the identity state the same thing.

10 Differentiation of Monotonic Functions

The original differentiation theorem of Lebesgue applies to monotonic functions. To obtain it here we need to show that such functions enjoy the Vitali property.

Theorem 10.1. *Let F be a continuous monotonic function. Then F has the Vitali property on every Borel set.*

Proof. We begin with the proof for the simpler case that F is strictly increasing. Let λ_F be the usual Lebesgue-Stieltjes outer measure associated with F, so that in particular for an open interval (a,b) we must have $\lambda_F((a,b)) = F(b) - F(a)$. Note that $\mathcal{L}^F((a,b)) \leq F(b) - F(a)$.

Let β be a fine cover of an interval (a,b). Then define β_1 as the collection of all pairs $([F(x), F(y)], F(z))$ corresponding to pairs $([x,y],z) \in \beta$. Since F is continuous and strictly increasing it is clear that β_1 is a fine cover of the interval $(F(a), F(b))$. Thus, by Theorem 4.1, for any $\varepsilon > 0$ there is a subpartition $\pi \subset \beta_1$

$$\pi = \{([F(x_i), F(y_i)], F(z_i)) : i = 1, 2, 3, \ldots, n\},$$

for which

$$\ell(\pi) = \sum_{i=1}^{n} (F(y_i) - F(x_i)) > F(b) - F(a) - \varepsilon.$$

But that supplies a subpartition $\pi' \subset \beta$,

$$\pi' = \{([x_i, y_i], z_i) : i = 1, 2, 3, \ldots, n\},$$

for which

$$V(F, \pi') = \sum_{i=1}^{n} (F(y_i) - F(x_i)) > F(b) - F(a) - \varepsilon.$$

We deduce that

$$F(b) - F(a) - \varepsilon < \mathcal{L}^F_*((a,b) \leq \mathcal{L}^F((a,b) \leq F(b) - F(a).$$

This proves the identity of the measures $\mathcal{L}^F$, $\mathcal{L}^F_*$, and λ_F on all open intervals. From that follows the identity on all open sets, all closed sets and then all Borel sets by properties of these measures (e.g., [1, pp. 133-135]).

If F is merely nondecreasing then a similar argument works but one has to take account of the countable set C of points x at which $F^{-1}(x)$ is an interval and in constructing β_1 dispense with pairs $([F(x), F(y)], F(z))$ for which $F(x) = F(y)$. Then β_1 becomes a fine cover of the set $(F(a), F(b)) \setminus F(C)$ and the proof continues unchanged. □

A consequence of this theorem is the following complete account of the differentiation structure of continuous monotonic functions, following immediately from the preceding material. The measure $\mathcal{L}^F$ is usually called the Lebesgue-Stieltjes measure associated with F and the display in the last line of the corollary is attributable to De La Vallée Poussin.

Corollary 10.2. *Let F be a continuous nondecreasing function. Then F is differentiable almost everywhere on $[a, b]$ and has a finite or infinite derivative $\mathcal{L}^F$-almost everywhere. Moreover, if D denotes the set of points where F has a finite derivative, D_∞ denotes the set of points where F has an infinite derivative and E is any Borel set then*

$$\mathcal{L}^F_*(E) = \mathcal{L}^F(E) = \int_{E \cap D} F'(x)\,dx + \mathcal{L}^F(E \cap D_\infty).$$

What do we do with discontinuities in this setting? Perhaps the simplest way to dispense with them is to observe what happens for a monotonic saltus function.

Theorem 10.3. *Let F be a nondecreasing saltus function with jumps occurring only in the countable set C. Then $\mathcal{L}^F_*(\mathbb{R} \setminus C) = \mathcal{L}^F(\mathbb{R} \setminus C) = 0$ and $F'(x) = 0$ almost everywhere.*

PROOF. It is easy to establish that $\mathcal{L}^F(\mathbb{R} \setminus C) = 0$. Lemma 7.2 then quickly shows that the set of points $x \in \mathbb{R} \setminus C$ at which $\overline{d}(F, x) > 0$ has $\mathcal{L}_*$-measure zero. □

11 Characterization of the Vitali Property

The usual presentation of the Lebesgue differentiation theorem seems to end with the observation that functions of bounded variation are a.e. differentiable. Certainly the methods do not suggest that more can be done. (In fact, as readers of Saks [5] well know, the methods can be lifted somewhat arduously to the class of VBG_* functions.)

Here we are under an obligation to ask for more: what class of functions has the Vitali property on a given set E? The answer is quite clean and natural. (Again readers of Saks will expect that the class of VBG$_*$ functions offers an equivalent expression of this property.) Details, along with a large set of equivalent conditions, can be found in [6].

Theorem 11.1. *Let F be continuous and let E be a Borel set. A necessary and sufficient condition that F should have the Vitali property on E is that the measure $\mathcal{L}^F$ be σ-finite on E.*

It follows then from Theorem 9.1 that, for a continuous function F for which $\mathcal{L}^F$ is σ-finite on a Borel set E, F is differentiable almost everywhere in E and has a finite or infinite derivative $\mathcal{L}^F$-almost everywhere in E.

12 Criteria for the Vitali Property

Since the Vitali property for a function F is generally equivalent to continuity plus the VBG$_*$ property it might be useful to mention some classical criteria under which it can be established that a function is VBG$_*$ on a set. This will also allow us to bypass the classical proofs and reveal the structure of what is happening a bit more clearly. We are concerned with estimates on the derivates of a function which will ensure finiteness, σ-finiteness or absolute continuity of $\mathcal{L}^F$. These will replace the familiar versions in Saks [5, pp.234–235].

Theorem 12.1. *If $-r < \underline{D}F(x) \le \overline{D}F(x) < r$ at every $x \in E$, then $\mathcal{L}^F(E) \le r\lambda(E)$.*

PROOF. This follows directly from Lemma 7.3. □

Theorem 12.2. *If $-r < \underline{D}^+F(x) \le \overline{D}^+F(x) < r$ at every $x \in E$, then there is a countable set N so that $\mathcal{L}^F(E \setminus N) \le r\lambda(E)$.*

PROOF. By an early theorem of G. C. Young (eg., Saks [5, p. 261]) the set of points where $\underline{D}^-F(x) > \overline{D}^+F(x)$ or $\overline{D}^-F(x) < \underline{D}^+F(x)$ is countable. Let N denote this set. Then F satisfies the hypotheses of Theorem 12.1 on the set $E \setminus N$ and so the conclusion then follows. □

Theorem 12.3. *If $\overline{D}F(x) < \infty$ at every $x \in E$, then there is a sequence of closed set $\{C_i\}$ covering E so that each $\mathcal{L}^F(C_i) < \infty$.*

PROOF. For each integer n consider the set E_n of points z for which $\overline{D}F(z) < n$. Write $G(x) = nx - F(x)$. Let β be the collection of all pairs $([x, y], z)$ with $z \in [x, y]$ and $F(y) - F(x) \le n(y - x)$, i.e., so that $G(y) - G(x) \ge 0$. Certainly

β is a full cover of E_n. Choose $\eta(z)$ for each $z \in E_n$ so that $y - x < \eta(z)$ implies that $([x,y],z) \in \beta$. Define for integers m, k

$$E_{mnk} = \{x \in E_n : \eta(x) > 1/m\} \cap \left[\frac{k}{m}, \frac{k+1}{m}\right].$$

Note the following feature of this construction: if $[x,y]$ is a subinterval of $[k/m,(k+1)/m]$ and if $[x,y]$ contains at least one point z of E_{mnk} then, since $y - x \leq 1/m < \eta(z)$, necessarily $([x,y],z) \in \beta$ and so, in particular $G(y) - G(x) \geq 0$.

Define $a = \inf \overline{E_{mnk}}$ and $b = \sup \overline{E_{mnk}}$. Because of this feature we have noted we can deduce that

$$V(G, \beta[\overline{E_{mnk}}]) \leq G(b) - G(a). \tag{1}$$

Then, since $|F(y) - F(x)| \leq |G(y) - G(x)| + n|y - x|$ it follows from (1) that

$$V(F, \beta[\overline{E_{mnk}}]) \leq G(b) - G(a) + n/m.$$

Again, because of the feature noted above, it is easy to see that β is a full cover of $\overline{E_{mnk}}$. From this it follows now that

$$\mathcal{L}^F(\overline{E_{mnk}}) \leq V(F, \beta[\overline{E_{mnk}}]) \leq G(b) - G(a) + n/m < \infty.$$

Thus the theorem is completed by taking for the sequence of closed sets $\{C_i\}$ covering E a relabeling of the countable collection $\{\overline{E_{mnk}}\}$. □

13 One-Sided Vitali coverings

The one-sided derivates have an unusual geometry that is not available for bilateral derivates and not available in any other setting that I know, certainly not in higher dimensions. Both papers Faure [2] and Hagood [4] exploit this geometry, the former by using the rising-sun lemma and the latter by a simple compactness argument that arises from one-sided versions of Vitali covers.

It should be noticed that the version in Hagood [4] could be reframed as a special Vitali argument available for this particular geometry. This may be worth doing in an introductory course to introduce a simplified Vitali covering argument as a prelude to more advanced material. Here are two lemmas illustrating how this might work. The first uses Hagood's method, the second an old method of Lebesgue. This latter method (employing transfinite chains of intervals) was the basis for Lebesgue's early (1903) analysis of the differentiation of monotonic functions and the integration of their derivates.

Lemma 13.1. *Suppose that $\mathcal{C}$ is a collection of closed intervals with the property that to each point x in a compact set K there is at least one interval $[x, y]$ in $\mathcal{C}$. Then for any $t < 1$ there is finite collection of nonoverlapping intervals $\{[x_i, y_i]\}$ from $\mathcal{C}$ so that $\sum_i (y_i - x_i) > t\lambda(K)$.*

PROOF. Augment $\mathcal{C}$ by adding to it, for every $[x, y] \in \mathcal{C}$ all intervals $[s, y]$ for which $s \leq x$ and $t(y - s) < (y - x)$. By Hagood's lemma (Hagood [4]) there is a finite collection of nonoverlapping intervals $[s_i, y_i]$ from the augmented collection that covers K so that, in particular, there correspond nonoverlapping intervals $[x_i, y_i]$ from $\mathcal{C}$ for which

$$\sum_i (y_i - x_i) > \sum_i t(y_i - s_i) > t\lambda(K). \qquad \square$$

Lemma 13.2. *Suppose that $\mathcal{C}$ is a collection of closed intervals with the property that to each point x in a compact set K there is at least one interval $[x, y]$ in $\mathcal{C}$. Then there is a finite or infinite sequence of nonoverlapping intervals from $\mathcal{C}$ that covers K.*

PROOF. Let $a = \inf K$, $b = \sup K$. Inductively define a transfinite sequence [known as a Lebesgue chain] by starting with $x_0 = a$ and choosing x_1 so that $[x_0, x_1] \in \mathcal{C}$. For x_2 we have two cases: case (i) $x_1 \in K$ in which case choose $[x_1, x_2] \in \mathcal{C}$, or case (ii) $x_1 \notin K$ in which case choose $x_2 = \inf K \cap [x_1, b]$ so that, in particular $[x_1, x_2) \cap K = \emptyset$. For every ordinal α, $x_{\alpha+1}$ is determined as in cases (i) and (ii). At any limit ordinal α we set $x_\alpha = \sup\{x_\gamma : \gamma < \alpha\}$. This process stops in a countable number of steps when $x_\alpha > b$. The countable collection of nonoverlapping intervals $[x_\alpha, x_{\alpha+1}]$ covers $[a, b]$ and removing all of the case (ii) choices results in a subcollection of $\mathcal{C}$ that covers the compact set K as required. $\square$

References

[1] A. M. Bruckner, J. B. Bruckner and B. S. Thomson, *Real Analysis*, Prentice-Hall (1996).

[2] C.-A. Faure, *The Lebesgue Differentiation Theorem via the Rising Sun Lemma*, Real Anal. Exch., **29** (2004-05), this issue.

[3] H. Federer, *Geometric Measure Theory*, Springer-Verlag, (Berlin, 1969).

[4] J. Hagood, *The Lebesgue Differentiation Theorem via Nonoverlapping Interval Covers*, Real Anal. Exch., **29** (2004-05), this issue.

[5] S. Saks, *Theory of the Integral*, Dover, (1937).

[6] B. S. Thomson, *Derivates of Interval Functions,* Memoir Amer. Math. Soc., **452**, Providence, 1991.