Vitali covers

1 Definition. A *Vitali cover* of a set $E \subseteq \mathbb{R}$ is a set $\mathcal{V}$ of closed intervals with positive length so that, for every $\delta > 0$ and every $x \in E$, there is some $I \in \mathcal{V}$ with $\lambda(I) < \delta$ and $x \in I$.

2 Lemma (Vitali covering) *Given a set $E \subseteq \mathbb{R}$ with $\lambda^*(E) < \infty$, a Vitali cover $\mathcal{V}$ of E, and some $\varepsilon > 0$, there are disjoint $I_1, \dots, I_n \in \mathcal{V}$ with $\lambda^*\big(E \setminus (I_1 \sqcup \cdots \sqcup I_n)\big) < \varepsilon$.*

Proof:

Start by picking an open set $O \supseteq E$ with $\lambda(O) < \infty$, and let $\mathcal{W} = \{I \in \mathcal{V} \colon I \subseteq O\}$. It is easy to check that $\mathcal{W}$ is a Vitali cover of E.

We now forget about E for a while, and pick pairwise disjoint intervals I_1, I_2, ... in $\mathcal{W}$. We shall do so one by one, starting with I_1, and we shall do it *greedily*, trying to make each interval as large as possible. Now there may not be any largest size available, so we settle for getting within a constant factor:

$$\lambda(I_{k+1}) > \tfrac{1}{2} \sup\big\{\lambda(J) \colon J \in \mathcal{W} \text{ and } J \cap I_j = \emptyset \text{ for } j = 1, \dots, k\big\}. \tag{1}$$

In ordinary prose: *At each step, pick the next interval disjoint from the other intervals picked this far, so that any other interval we could have picked is less than twice as long.* It is conceivable that this process cannot go on forever. But if so, we have already covered all of E with a finite, pairwise disjoint set of intervals from $\mathcal{W}$, so we're done.

If $x \in E \setminus (I_1 \sqcup \cdots \sqcup I_k)$, then since $I_1 \sqcup \cdots \sqcup I_k$ is closed, there is some δ-neighbourhood of x that does not meet $I_1 \sqcup \cdots \sqcup I_k$, and so we can continue with the selection for one more step.

If the process continues forever, then we find

$$\sum_{k=1}^{\infty} \lambda(I_k) = \lambda\Big(\bigsqcup_{k=1}^{\infty} I_k\Big) \le \lambda(O) < \infty,$$

so the sum converges. Pick n large enough so that

$$\sum_{k=n+1}^{\infty} \lambda(I_k) < \varepsilon,$$

and let J_k be the interval with the same center as I_k, but five times the length.

I claim

$$E \setminus \bigsqcup_{k=1}^{n} I_k \subseteq \bigcup_{k=n+1}^{\infty} J_k,$$

which in turn implies

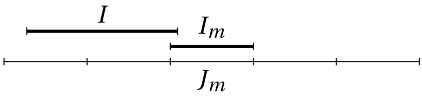
$$\lambda^*\Big(E \setminus \bigsqcup_{k=1}^{n} I_k\Big) \le \sum_{k=n+1}^{\infty} \lambda(J_k) = 5 \sum_{k=n+1}^{\infty} \lambda(I_k) < 5\varepsilon,$$

and that is enough to finish the proof.

To prove the claim, let $x \in E \setminus (I_1 \sqcup \cdots \sqcup I_n)$. Since $\mathcal{W}$ is a Vitali cover for E, we can find some $I \in \mathcal{W}$ with $x \in I$ and $I \cap I_k = \emptyset$ for $k = 1, \ldots, n$.

Let m be the smallest natural number with $I \cap I_m \neq \emptyset$.

Such an m must exist, for otherwise, I would always be one of the intervals we *could* have picked, and therefore $\lambda(I_{k+1}) > \frac{1}{2}\lambda(I)$ for all k. But this is impossible, since $\sum \lambda(I_k) < \infty$. There is more – when $k = m-1$, I is still one of the J's, so we must have $\lambda(I_m) > \frac{1}{2}\lambda(I)$.



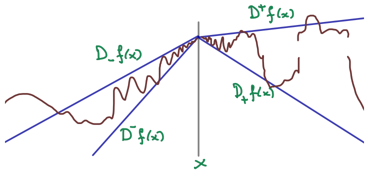
This drawing illustrates the fact that, when $I_m \cap I \neq \emptyset$ and $\lambda(I_m) > \frac{1}{2}\lambda(I)$, then $J_m \supseteq I$. In particular, $x \in J_m$, and the proof is finally complete. ∎

Dini derivatives

These are defined by

$$D^+ f(x) = \overline{\lim_{y \searrow x}} \frac{f(y) - f(x)}{y - x}, \quad D_+ f(x) = \underline{\lim_{y \searrow x}} \frac{f(y) - f(x)}{y - x},$$
$$D^- f(x) = \overline{\lim_{y \nearrow x}} \frac{f(y) - f(x)}{y - x}, \quad D_- f(x) = \underline{\lim_{y \nearrow x}} \frac{f(y) - f(x)}{y - x},$$

and called the *upper right*, *lower right*, *upper left*, andn *lower left* Dini derivatives, respectively.



The Dini derivatives have these simple properties:

$$-\infty \le D_\pm f \le D^\pm f \le \infty,$$
$$D^\pm(-f) = -D_\pm f,$$
$$D^\pm g = D_\mp f \qquad \text{where } g(x) = f(-x).$$

Note that any of the Dini derivatives can take on either of the values $\pm\infty$.

Our interest in the Dini derivatives stems from the fact that they always exists, and moreover f has a left derivative x if and only if $D_- f(x) = D^- f(x)$, a right derivative if and only if $D_+ f(x) = D^+ f(x)$, and a (two-sided) derivative if and only if all four Dini derivatives are the same.

Limits to growth for a monotone function

From now on, $a < b$ are real numbers, and $f\colon (a,b) \to \mathbb{R}$ is an increasing function.

3 Lemma *$D^+ f(x) < \infty$ and $D^- f(x) < \infty$ for λ-a.e. $x \in (a,b)$.*

Proof: Let $E = \{x \in (a,b)\colon D^+ f(x) = \infty\}$. The idea of the proof is that f must have unbounded growth on this set, if it has positive measure. To get this to work, we will first assume that $|f| \le M$ on (a,b).

Pick any (large) number m. It follows from the definition that the set of intervals $[x,y]$ where

$$x \in E, \quad y \in (x,b), \quad \text{and } \frac{f(y)-f(x)}{y-x} > m$$

is a Vitali cover for E. Pick any $\varepsilon > 0$, and let $[x_k,y_k]$ be pairwise disjoint intervals of this type for $k = 1,\dots,n$ with $\lambda^*(E \setminus ([x_1,y_1] \sqcup \cdots \sqcup [x_n,y_n])) < \varepsilon$. It follows that $(y_1 - x_1) + \cdots + (y_n - x_n) > \lambda^*(E) - \varepsilon$, and therefore

$$\bigl(f(y_1) - f(x_1)\bigr) + \cdots + \bigl(f(y_n) - f(x_n)\bigr) > m(\lambda^*(E) - \varepsilon).$$

But because f is increasing, the intervals $(f(x_k), f(y_k))$ are pairwise non-overlapping in $[-M,M]$, so we must finally have $2M > m(\lambda^*(E) - \varepsilon)$. If $\lambda^*(E) > 0$, we can choose $\varepsilon < \lambda^*(E)$ and m large enough for this to be a contradiction.

If f is unbounded, we apply the above to the restriction of f to slightly smaller intervals $(a + n^{-1}, b - n^{-1})$ instead, and use the fact that a countable union of sets of measure zero still has measure zero.

We have proved that $D^+ f < \infty$ a.e. To show the same for $D^- f$, apply the first result to the function $x \mapsto -f(-x)$ on $(-b,-a)$. ∎

Limits to wiggliness for a monotone function

4 Lemma *$D^+f(x) = D_+f(x)$ and $D^-f(x) = D_-f(x)$ for λ-a.e. $x \in (a,b)$.*

Proof: As in the previous lemma, the second a.e. equality follows from the first by applying it to $x \mapsto -f(-x)$. So we only prove the first one. Note that $D^+f(x) \ge D_+f(x)$ always, so we need to show that the set of x where $D^+f(x) > D_+f(x)$ has measure zero. But for any such x, we can always find two rational numbers r and s with $D^+f(x) > s > r > D_+f(x)$. Since the number of rational pairs (r,s) with $r < s$ is countable, we only need to show that the set

$$E = \{x \in (a,b) \colon D^+f(x) > s > r > D_+f(x)\} \qquad (r,s \in \mathbb{Q} \text{ fixed})$$

has measure zero.

The proof idea is like that for the previous lemma, only done twice: Use one inequality to show that the average growth rate of f on a set of intervals nearly covering E is less than r, then use the other to find subintervals, still nearly covering E, where the average growth rate is greater than s. Then combine the two to get a contradiction, if $\lambda^*(E) > 0$.

To get started, then, we first pick some $\varepsilon > 0$, and an open set $O \supset E$ with $\lambda(O) < \lambda^*(E) + \varepsilon$. Consider the Vitali cover for E consisting of those intervals $[x,y]$ where

$$x \in E, \quad x < y, \quad y \in O, \quad \text{and } \frac{f(y)-f(x)}{y-x} < r.$$

Then let $\varepsilon > 0$, and use the Vitali covering lemma to pick pairwise disjoint intervals $[x_k, y_k]$ of this type, with with $\lambda^*(E \setminus ([x_1,y_1] \sqcup \cdots \sqcup [x_n,y_n])) < \varepsilon$.

It follows that

$$\sum_{k=1}^{n} \bigl(f(y_k) - f(x_k)\bigr) < r \sum_{k=1}^{n} (y_k - x_k).$$

Now let $U = (x_1, y_n) \sqcup \cdots \sqcup (x_n, y_n)$, and create yet another Vitali cover, this time of $E \cap U$, consisting of those intervals (u,v) where

$$u \in E \cap U, \quad u < v, \quad (u,v) \subseteq U, \quad \text{and } \frac{f(v)-f(u)}{v-u} > s.$$

Use the Vitali lemma to almost cover $E \cap U$ with pairwise disjoint intervals $[u_j, v_j]$ of this type, with $\lambda^*(E \cap U \setminus ([u_1,v_1] \sqcup \cdots \sqcup [u_m,v_m])) < \varepsilon$. Using reasoning that should be familiar by now, we get

$$\sum_{j=1}^{n} \bigl(f(v_j) - f(u_j)\bigr) > s \sum_{j=1}^{n} (v_j - u_j).$$

But since f is increasing, the intervals $[f(u_j), f(v_j)]$ are non-overlapping. They are also each contained in some $[f(x_k), f(y_k)]$, so

$$\sum_{j=1}^{n} \bigl(f(v_j) - f(u_j)\bigr) \le \sum_{k=1}^{n} \bigl(f(y_k) - f(x_k)\bigr).$$

From the inequalities shown we get

$$s\sum_{j=1}^{m}(v_j-u_j)<r\sum_{k=1}^{n}(y_k-x_k).$$

To estimate the right hand side, note $\bigsqcup_{k=1}^{n}[x_k,y_k]\subseteq O$, so

$$r\sum_{k=1}^{n}(y_k-x_k)\le r\lambda(O)<r\big(\lambda^*(E)+\varepsilon\big).$$

And for the left hand side,

$$s\sum_{k=1}^{m}(v_j-u_j)=s\lambda\Big(\bigsqcup_{j=1}^{m}[u_j,v_j]\Big)>s\big(\lambda^*(E\cap U)-\varepsilon\big)>s\big(\lambda^*(E)-2\varepsilon\big),$$

so now we have

$$s\big(\lambda^*(E)-2\varepsilon\big)<r\big(\lambda^*(E)+\varepsilon\big).$$

Recall that $r<s$. So if $\lambda^*(E)>0$, we can pick $\varepsilon>0$ small enough that the above inequality becomes a contradiction. ∎

Limited number of corners

From now on, $a < b$ are real numbers, and $f\colon (a,b) \to \mathbb{R}$ is an arbitrary function.

5 Definition. Call $u \in (a,b)$ a *strict local maximum point* for f if there is some $\delta > 0$ so that $f(x) < f(u)$ whenever $x \in (a,b)$ is such that $|x-u| < \delta$.

6 Lemma *The set of strict local maximum points for f is countable.*

Proof: Any two u values satisfying the definition above for some $\delta > 0$ must be at least a distance δ apart, so there is only room for a finite number of them in (a,b). But any strict local maximum must satisfy the definition for some $\delta \in \{1/n\colon n \in \mathbb{N}\}$, so their total number is countable. ∎

7 Definition. A *corner point* for f is a $u \in (a,b)$ so that either $D_- f(u) > D^+ f(u)$ or $D^- f(u) < D_+ f(u)$.

To understand the meaning of this definition, consult the illustration of Dini derivatives above, which shows a corner of the first kind. Turn the picture upside down to see a corner of the second kind.

8 Lemma *Any real valued function f has only a countable number of corner points.*

Proof: We only show that the set of u satisfying the first inequality is countable. The same result for the second inequality will then follow by replacing f by $-f$.

If the first inequality holds, there is some $q \in \mathbb{Q}$ with $D_- f(u) > q > D^+ f(u)$. If this holds, then x is a strict local maximum for the function $g(x) = f(x) - qx$, since then $D_- g(u) > 0 > D^+ g(u)$.

By Lemma 6 only a countable number of such u exist for any q, and since $\mathbb{Q}$ is countable, we are done. ∎

Almost everywhere differentiability

9 Theorem *Any monotone function is differentiable almost everywhere.*

Proof: We only need to consider an increasing function $f\colon (a,b) \to \mathbb{R}$. By Lemma 3 and 4, f has finite one-sided derivatives almost everywhere. If f has one-sided derivatives at some point, and they are different, then f has a corner point there. But Lemma 8 implies that the corner points have measure zero. ∎

10 Example. Let $f\colon [0,1] \to [0,1]$ be the Cantor function, and $C \subset [0,1]$ the (standard) Cantor set. Then f is locally constant on the open set $(0,1) \setminus C$, so $f' = 0$ there. But $\lambda(C) = 0$, so $f' = 0$ almost everywhere. In particular,

$$\int_0^1 f'\,d\lambda = 0, \quad \text{and yet } f(1) - f(0) = 1.$$

This shows that we cannot expect the fundamental theorem of calculus to hold for arbitrary monotone functions – more is needed. What is happening here is that all of the growth happens within a set of measure zero.

An integral (in)equality

11 Proposition *Let $f\colon [a,b]\to\mathbb{R}$ be an increasing function. Then f' is integrable, and*

$$\int_a^b f'(x)\,dx \le f(b)-f(a).$$

If there is some constant L so that f satisfies

$$f(y)\le f(x)+L(y-x) \qquad \text{whenever } a\le x<y\le b,$$

then equality holds instead:

$$\int_a^b f'(x)\,dx = f(b)-f(a).$$

By *increasing*, I mean what some prefer to call *non-decreasing*.

Proof: For simplicity, extend f by setting $f(x)=f(b)$ when $x>b$. Define $f_n\colon [a,b]\to\mathbb{R}$ by

$$f_n(x)=n\cdot\bigl(f(x+n^{-1})-f(x)\bigr).$$

Then $f_n\ge 0$, and $f_n\to f'$ a.e. Therefore Fatou's lemma gives us

$$\begin{aligned}
\int_a^b f'(x)\,dx &\le \varliminf_{n\to\infty}\int_a^b f_n(x)\,dx\\
&= \varliminf_{n\to\infty} n\int_a^b \bigl(f(x+n^{-1})-f(x)\bigr)\,dx\\
&= \varliminf_{n\to\infty} n\Bigl(\int_{a+n^{-1}}^{b+n^{-1}} f(x)\,dx-\int_a^b f(x)\,dx\Bigr)\\
&= \varliminf_{n\to\infty}\Bigl(n\int_b^{b+n^{-1}} f(x)\,dx-n\int_a^{a+n^{-1}} f(x)\,dx\Bigr)\\
&\le f(b)-f(a).
\end{aligned}$$

If the stated linear growth condition holds, then $f_n\le L$, so instead of using Fatou's lemma, we can use the bounded convergence theorem (BCT), and the first inequality in the calculation above becomes equality. Also, $\varliminf$ can be replace by an ordinary limit. The final inequality also becomes an equality, since f is now continuous. ∎

The first fundamental theorem of calculus

12 Theorem *If $f \in \mathcal{L}^1([a,b])$ and $F(x) = \int_a^x f(t)\,dt$ for $x \in [a,b]$, then F is differentiable a.e., and $F' = f$ a.e. in $[a,b]$.*

The proof of this theorem builds on the following

13 Lemma *If $f \in \mathcal{L}^1([a,b])$ and $\int_a^x f(t)\,dt = 0$ for all $x \in [a,b]$, then $f = 0$ a.e. in $[a,b]$.*

Proof: Assume that $f > 0$ in a subset of (a,b) of positive measure. Then this subset has a compact subset K of positive measure, and then $\int_K f(t)\,dt > 0$. Since $\int_a^b f(t)\,dt = 0$, we must have $\int_{(a,b)\setminus K} f(t)\,dt < 0$. But $(a,b)\setminus K$ is an open set, and hence is a countable disjoint union of open intervals. For at least one of those intervals, say (c,d), we must have $\int_{(c,d)} f(t)\,dt < 0$. But $\int_{(c,d)} f(t)\,dt = \big(\int_a^d - \int_a^c\big) f(t)\,dt = 0$, by assumption – and this contradiction shows that $f \le 0$ a.e.

Applying this to $-f$, we see that $f \ge 0$ a.e., so $f = 0$ a.e.

The above proof is for real-valued f. The extension to complex-valued f by considering the real and complex parts separately is straight-forward. ∎

Proof of the theorem: Assume first that $f \ge 0$. Then F is increasing, so it is differentiable a.e., and as Proposition 11 shows that

$$\int_a^x F'(t)\,dt \le F(x) \qquad \text{for all } x \in [a,b].$$

Furthermore, *if f is bounded*, then F satisfies the extra condition of Proposition 11, so the above inequality is actually an equality. But that is equivalent to

$$\int_a^x \big(F'(t) - f(t)\big)\,dt = 0 \qquad \text{for all } x \in [a,b],$$

and then the above lemma shows that $F' = f$ a.e. This completes the proof in that special case.

If $f \ge 0$ but f is unbounded, write

$$F_n(x) = \int_a^x \big(f(t) \wedge n\big)\,dt,$$

and note that then $F(x) - F_n(x)$ is the integral of a nonnegative function, so that $F - F_n$ is an increasing function. In particular, $F' \ge F_n'$ at any point where both derivatives exists (i.e., a.e.). Therefore, and using the result of the previous paragraph,

$$\int_a^x F'(t)\,dt \ge \int_a^x F_n'(t)\,dt = \int_a^x \big(f(t) \wedge n\big)\,dt.$$

Finally, since $f \wedge n \nearrow f$ when $n \to \infty$, we can use the monotone convergence theorem to conclude that

$$\int_a^x F'(t)\,dt \geq \int_a^x f(t)\,dt \qquad \text{for all } x \in [a,b].$$

Just like in the bounded case, this once more completes the proof, now for the case when $f \geq 0$.

If f is a real-valued integrable function, the theorem readily follows by using the result for the positive and negative parts of f.

Similarly, if f is complex valued, using the result for real and imaginary parts separately readily yields the desired result. ∎

Absolutely continuous functions

Recall that the Cantor function ψ is continuous and differentiable a.e. on $[0,1]$, whith $\psi' = 0$ a.e. yet $\psi(1) \neq \psi(0)$, hence it cannot satisfy $\psi(1) = \psi(0) + \int_0^1 \psi'(x)\,dx$, as the second fundamental theorem of calculus would have us believe.

It turns out that the remedy is to introduce a stronger form of continuity.

14 Definition. A function f on an interval $[a,b]$ is called *absolutely continuous* if for every $\varepsilon > 0$ there is a $\delta > 0$ so that, whenever $[a_k, b_k]$ for $k = 1,\ldots,n$ are nonoverlapping intervals with $\sum_{k=1}^n (b_k - a_k) < \delta$, we have $\sum_{k=1}^n \bigl(f(b_k) - f(a_k)\bigr) < \varepsilon$.

15 Lemma *If $f \in \mathcal{L}([a,b])$ and $F(x) = \int_a^x f(t)\,dt$, then F is absolutely continuous.*

Proof: The lemma follows from the fact that, for any $\varepsilon > 0$, there exists $\delta > 0$ so that for any measurable $E \subseteq [a,b]$ with $\lambda(E) < \delta$, we have $\int_E |f(x)|\,dx < \varepsilon$.

This follows immediately by taking $E = \bigcup_{k=1}^n [a_k, b_k]$ in the definition of absolute continuity.

(the fact referred to above was an exercise. I'll add a proof here later.) ∎

We shall prove the converse of the above theorem. An important step along the way is the following:

16 Lemma *If f is absolutely continuous and differentiable a.e. on $[a,b]$ with $f' = 0$ a.e., then f is constant.*

Later, we shall see that absolutely continuous functions necessarily are differentiable a.e. (they have bounded variation), but for now, we just assume it.

Proof: Let D be the set of points in (a,b) in which f is differentiable and the derivative is zero. Let $\varepsilon > 0$, and let $\delta > 0$ be as in the definition of absolute continuity. The intervals $[c,d] \subseteq [a,b]$ for which $|f(d) - f(c)| < \varepsilon|d-c|$ form a Vitali cover of D, so by the Vitali covering lemma and the fact that $\lambda[a,b] \setminus D = 0$, there are a finite, disjoint sequence of such intervals $[c_k, d_k] \subseteq [a,b]$ with $k = 1,\ldots,n$ such that $\lambda([a,b] \setminus \bigsqcup_{k=1}^n (c_k, d_k)) < \delta$. Order these intervals so that $c_1 < d_1 < c_2 < d_2 < \cdots < c_n < d_n$, and write $[a_k, b_k]$ for the gaps: So let $[a_0, b_0] = [a, c_1]$, $[a_1, b_1] = [d_1, c_2]$, and so on up to $[a_n, b_n] = [d_n, b]$. Then $\sum_{k=0}^n (b_k - a_k) < \delta$, so $\sum_{k=0}^n \bigl(f(b_k) - f(a_k)\bigr) < \varepsilon$. Then

$$
\begin{aligned}
f(b) - f(a) &= f(b) - f(d_n) + f(d_n) - f(c_n) + f(c_n) - \cdots - f(d_1) + f(c_1) - f(a) \\
&= \sum_{k=0}^n \bigl(f(b_k) - f(a_k)\bigr) + \sum_{k=1}^n \bigl(f(d_k) - f(c_k)\bigr),
\end{aligned}
$$

and so

$$\begin{aligned}\left|f(b)-f(a)\right| &\le \sum_{k=0}^{n}\left|f(b_k)-f(a_k)\right| + \sum_{k=1}^{n}\left|f(d_k)-f(c_k)\right| \\ &< \varepsilon + \sum_{k=1}^{n} \varepsilon|d_k-c_k| \\ &\le (1+b-a)\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, $f(b) = f(a)$. The same argument with b replaced by any $x \in [a,b]$ shows that $f(x) = f(a)$, and the proof is therefore complete. ∎

The second fundamental theorem of calculus

17 Theorem *If f is absolutely continuous on $[a,b]$, then f is differentiable a.e. on $[a,b]$, $f' \in \mathcal{L}^1([a,b])$, and*

$$f(x) = f(a) + \int_a^x f'(t)\,dt$$

for all $x \in [a,b]$.

Proof: It is shown elsewhere that f is differentiable a.e. and that $f' \in \mathcal{L}^1([a,b])$ (see the textbook – or perhaps I will add a proof here later).

Let

$$g(x) = f(x) - \int_a^x f'(t)\,dt.$$

It follows from the first fundamental theorem of calculus that $g' = 0$ a.e., and from Lemma 15 that g is absolutely continuous. Lemma 16 then shows that g is constant. The value $x = 0$ shows that then $g(x) = f(a)$ for all x, and the proof is complete. ∎

A note on the missing pieces in the above proof: One proceeds by showing that any absolutely continuous function has *bounded variation* (BV). The proof is quite simple, once you know the definition. It turns out that the BV functions are differences between increasing functions, and we already know that increasing functions are differentiable a.e., and their derivatives are integrable – see Propositin 11 above. Thus the only missing piece is a small bit of theory for BV functions.