Differentiability of Lipschitz Functions on $\mathbb{R}^n$

Reese Lance

November 2022

Title: Differentiability of Lipschitz Functions on $\mathbb{R}^n$

Abstract: Lipschitz continuity is an enhancement of ordinary continuity. We know that continuity is not able to guarantee any smoothness: for example, the Weierstrass function is continuous everywhere but differentiable nowhere. Lipschitz continuity is powerful enough to imply differentiability a.e. for functions defined on $\mathbb{R}^n$. We will present a sketch of the proof in 1-dimension, then show how this case, along with some classical advanced calculus/analysis, will show the analogous statement for arbitrary dimensions. The basic blueprint is as follows:

Lipschitz $\Rightarrow$ Absolutely continuous $\Rightarrow$ Bounded variation $\Rightarrow$ Difference of monotone increasing functions

$\Rightarrow$ Differentiability a.e. of Lipschitz fcns on $\mathbb{R}^1$ $\Rightarrow$ Differentiability a.e. of Lipschitz fcns on $\mathbb{R}^n$

Theorem (Rademacher, 1919): *A Lipschitz function defined on U an open subset of $\mathbb{R}^n$ is differentiable a.e. wrt the Lebesgue measure.*

The general result on open subsets of $\mathbb{R}^n$ is known as Rademacher's theorem, which was proven in 1919. This is a generalization of the $n = 1$ case which was proved by Lebesgue, and is now a standard sequence of results in real analysis textbooks. The now-standard proof of the generalization to open subsets of $\mathbb{R}^n$ was first introduced by Charles Morrey in 1960, and we are roughly following his proof strategy.

Definition: A Lipschitz function is a function $f : \mathbb{R} \to \mathbb{R}$ such that there exists a constant $K > 0$ such that for all $x_1, x_2 \in \mathbb{R}$,

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$$

Exercise: Lipschitz continuity implies continuity. However, it is a stronger notion, as the example $\sqrt{x}$ on $[0, 1]$ shows. This is uniformly continuous but not Lipschitz.

Definition: A function $f : \mathbb{R} \to \mathbb{R}$ is called absolutely continuous if for every $\epsilon > 0$, there exists $\delta > 0$ such that for every finite collection of pairwise disjoint intervals (x_i, y_i),

$$\sum_k (y_k - x_k) < \delta \Rightarrow \sum_k |f(y_k) - f(x_k)| < \epsilon$$

This is some sort of notion of shrinking intervals. In fact, if f is absolutely continuous, then choosing $|I| = 1$ shows that f is uniformly continuous.

Proposition: *Lipschitz implies absolutely continuous.*

Proof: Let K be the Lipschitz constant of f. For $\epsilon > 0$, let $\delta = \frac{\epsilon}{K}$. Then for any collection of pairwise disjoint intervals, (x_i, y_i) whose total length is less than δ, we have

$$\sum_k |f(y_k) - f(x_k)| \leq K \sum_k |y_k - x_k| < K\delta = \epsilon$$

as required.

□

The converse is also true if you assume the derivative is bounded by the Lipschitz constant. Thus taking any unbounded, integrable function, such as $f = \frac{1}{\sqrt{x}}$ on $[0, 1]$, then $F = \int_0^x f(x)dx$ is absolutely continuous but not Lipschitz, since f is unbounded.

Definition: $f : \mathbb{R} \to \mathbb{R}$ is of bounded variation if

$$\lim_{x\to\infty} T_F(x) \equiv \lim_{x\to\infty} \sup \left\{ \sum_{j=1}^{n} \left|F(x_{j-1}) - F(x_j)\right| \right\} < \infty$$

where (x_j) is a partition of the interval (x_0, x).

Proposition: *An absolutely continuous function $f : [a,b] \to \mathbb{R}$ is of bounded variation.*

Proof(sketch): Suppose we are given a partition into n intervals. Let $\delta > 0$ be the number realizing the absolute continuity of f after choosing $\epsilon = 1$. Absolute continuity tells us that given a finite collection of intervals whose sum of lengths is less than delta, then the sum $\sum_{i=1}^{k} |F(y_i) - F(x_i)| < 1$. We are going to "pass over" the interval $[a,b]$ N times, selecting intervals until the sum of their lengths surpasses δ, possibly by also adding in extra subdivision points if necessary. Then for each pass through of $[a,b]$, the quantity $\sum |F(x_j) - F(x_{j-1})|$ is less than 1. There are N of these such groups, so the total variation must be less than N. The idea is that: We don't have control over δ. So the smaller it gets, the larger N must become, i.e. the more passes through we must make because we can pick up less and less intervals before surpassing δ, as δ shrinks. This idea is made precise by letting $N = \lfloor \frac{b-a}{\delta} + 1 \rfloor$. Since we showed this for an arbitrary n, the supremum must be finite.

□

Note we had to restrict to bounded interval here.
This implication is also not a biconditional: For example, the Cantor function is of bounded variation but not absolutely continuous.

Proposition: *If $F : \mathbb{R} \to \mathbb{R}$ is of bounded variation, then the functions $T_F + F$ and $T_F - F$ are increasing.*

Proof: Suppose $x < y$ and choose some $\epsilon > 0$. Then there exists a partition $x_0 < \cdots < x_n = x$ s.t.

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \epsilon$$

because $T_F(x)$ is the sup over all such partitions. Then we can define a new partition

$$x_0 < \cdots < x_n = x < y$$

which we may use to approximate the total variation $T_F(y)$. This is just given by the original variation $T_F(x)$, adding on the single extra term $|F(y) - F(x)|$:

$$T_F(y) \geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)|$$

$$T_F(y) \pm F(y) \geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)| \pm (F(y) - F(x) + F(x))$$

$$T_F(y) \pm F(y) \geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)| \pm F(y) - F(x) \pm F(x)$$

$$T_F(y) \pm F(y) \geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)| \pm F(y) - F(x) \pm F(x)$$

This reduces to (There are 4 cases to check depending on the sign of the absolute value and which of the $\pm$ signs you choose. There is no unifying argument here, you just have to check each of the 4 cases.)

$$T_F(y) \pm F(y) \geq T_F(x) \pm F(x)$$

as desired.

□

Thus any function of bounded variation can be written as the difference of increasing functions:

$$F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$$

Proposition: *An increasing function $F : \mathbb{R} \to \mathbb{R}$ is differentiable a.e.*

Proof:
First we note that F being increasing can only be discontinuous at countably many points, i.e. F is continuous a.e. This is a standard classical analysis problem: we even had it on a previous comp exam. Thus we may define a new function, $G(x) = F(x+)$, which is continuous and agrees with F a.e. In particular, G is increasing and right continuous, thus we may associate to it a premeasure:

$$\mu_0\left(\bigcup^n (a, b]\right) := \sum^n G(b) - G(a)$$

Here the increasing condition guarantees that μ_0 is always non-negative, and right continuity implies one inequality in the statement of countable sub-additivity, though this is somewhat subtle.

In fact we showed on an exam in this class that sets of these type generate the Borel algebra, so it suffices to define μ_0 on these. Then we apply Caratheodory theorem to obtain a Borel measure, μ_G. I don't want to go too in detail here because I believe this is the content of another student's presentation.

So we have a measure μ_G, and we may consider its LRN decomposition into an absolutely continuous piece and a singular piece wrt the Lebesgue measure, m:

$$d\mu_G = d\lambda + f dm$$

for some fcn f, which is known as the Radon-Nikodym derivative.

Theorem (Folland 3.22): *If ν is a regular measure on $\mathbb{R}^n$ with LRN decomposition $d\nu = d\lambda + f dm$, then for m-almost every $x \in \mathbb{R}^n$,*

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

for every family $\{E_r\}$ that shrinks nicely to x.

This theorem is a direct corollary of the Lebesgue Differentiation Theorem.

So in our case (indeed μ_G regular, the proof of which we omit), we first observe that the families $E_r^1 = \{(x-r, x]\}$ and $E_r^2 = \{(x, x+r]\}$ shrink nicely to x for every x. Applying the above theorem, then, we have for a.e. x, considering E_r^2 for example,

$$\lim_{r\to 0} \frac{\nu(E_r^2)}{m(E_r^2)}$$
$$= \lim_{r\to 0} \frac{\mu_G((x, x+r])}{m((x, x+r])}$$
$$= \lim_{r\to 0} \frac{G(x+r) - G(x)}{r} = f(x)$$

Thus, with the addition of the limit provided by considering E_r^1, we have shown that G is differentiable a.e. As G agrees with F away from a countable subset of $\mathbb{R}$, this implies F is also differentiable.

□

Corollary: *A Lipschitz continuous function $f : \mathbb{R} \to \mathbb{R}$ is differentiable a.e.*

Proof: This is just summing up what we have done: Take your f. Restricted to the interval $[n, n+1]$, it is absolutely continuous and thus of bounded variation. This implies the functions $T_f + f$ and $T_f - f$ are increasing functions, and by the above, are thus differentiable a.e. on $[n, n+1]$. But

$$f = \frac{1}{2}(T_f + f) - \frac{1}{2}(T_f - f)$$

so f is also differentiable a.e. on $[n, n+1]$, i.e. it is differentiable apart from a set of measure 0. Taking unions as n goes to ∞ and $-\infty$, and noting that the Lebesgue measure obeys countable sub-additivity, it follows that $f : \mathbb{R} \to \mathbb{R}$ is differentiable a.e.

□

Proposition: *A Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable a.e.*

Proof (sketch): First, we show that all the directional derivatives

$$D_v f(x) \equiv \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}$$

exist. This follows from Fubini's theorem: The n-dimensional measure of the set where these directional derivatives don't exist is given by the product of the one-dimensional measures, all of which are 0 by the $n = 1$ case.

Further, we have that

$$D_v f(x) = grad\ f \cdot v$$

a generalization of the classical result in calculus. Choose a countable, dense subset of the unit sphere in $\mathbb{R}^n$, and let A denote the set of points of $\mathbb{R}^n$ where $D_v f = grad\ f \cdot v$ for all v. Because there are countably many v's, $m(A^C) = 0$. So we need to show

$$f(x+w) - f(w) = df(x) \cdot w + o(\|w\|)$$

To do this, observe, if u is the unit vector corresponding to w, $u = \frac{w}{\|w\|}$,

$$\frac{f(x+w) - f(x)}{\|w\|} = \frac{f(x+\|w\|v) - f(x)}{\|w\|} + \frac{f(x+\|w\|u) - f(x+\|w\|v)}{\|w\|}$$

Sketch of how to bound: The first term on RHS approximates the directional derivative a.e. as we claimed. The second term is bounded by $K\|u-v\|$, by the Lipschitz assumption. We may choose v such that this is bounded by $\frac{1}{2}\epsilon$, and $df(x) \cdot w$ is bounded by $M\|u-v\|$, since it is a linear function on a f.d. VS, so we can also choose v such that this term is bounded by $\frac{1}{2}\epsilon$. This shows that the error term dies quickly enough.

□

Corollary: *For U an open subset of $\mathbb{R}^n$, a Lipschitz continuous function $f : U \to \mathbb{R}^m$ is differentiable a.e.*

Proof: Note that a function $\mathbb{R}^n \to \mathbb{R}^m$ is differentiable iff its component functions $f = (f_1, \ldots, f_m)$ are differentiable, so this follows immediately from applying the above to each component.

□

Some results that Rademacher's theorem is used to prove:

Theorem (Uniqueness of Closest Point): *Let K be compact. Then the set of points $x \in \mathbb{R}^n \setminus K$ such that the closest point in K to x is not unique is of measure 0.*

To prove this, one shows that the distance function is Lipschitz, thus differentiable a.e. by Rademacher. At such an x where the distance function is differentiable, then the closest point in K to x ends up being unique.

Theorem (Characterization of Differentiability): *$f : U \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable a.e. iff*

$$\limsup_{y \to x} \frac{|f(y) - f(x)|}{|y - x|} < \infty, \quad \text{ae } x$$

This will originally shown via an application of Lebesgue density arguments applied to some suitable sets, but it was later proven even without reference to the density theorem.

The $n = 1$ case of Rademacher's theorem is contained within Folland's "Real Analysis", though you have to jump around the book a bit to produce it. Mostly it is contained within chapter 3.
The n dimensional case is covered in Federer's "Geometric Measure Theory" (Theorem 3.1.6, but really uses all of section 3.1 leading up to Theorem 3.1.6).
A primary source is Rademacher "Über partielle und totale differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale" (1919) doi:10.1007/BF01498415, but we did not follow this proof technique and I don't have an english version on hand.