Rademacher's theorem

Math212b

January 28, 2005

This theorem says that a Lipschitz function on $\mathbb{R}^n$ is differentiable almost everywhere.

Here is a more detailed description of this theorem. If X and Y are Banach spaces, with U an open subset of X, and $f : U \to Y$ a map, we say that f is differentiable at $x \in U$ if there is a bounded linear map $df_x : X \to Y$ such that

$$f(x+h) - f(x) = df_x(h) + 0(h).$$

This means that for any $\epsilon > 0$ there is a $\delta > 0$ such that

$$\|f(x+h) - f(x) - df_x(h)\| \leq \epsilon \|h\| \quad \forall h \text{ such that } \|h\| < \delta.$$

Rademacher's theorem asserts that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz, then it is differentiable almost everywhere. We will need a Sobolev space version of this theorem. So I plan to state and prove the Sobolev space version (which relies on some elementary mollification) in lecture. But I thought for general education purposes you should see the original version which depends on a entirely different style arguments.

Contents

1 The Vitali covering theorem.

Definition 1 *Let E be a subset of the metric space (X, d). Let Φ be a class of subsets of X such that for all $\epsilon > 0$ and $x \in E$ there is some $U \in \Phi$ with $x \in U$ and diam $(U) < \epsilon$. Then the family Φ is called a* **Vitali class** *for E.*

We will let $|A|$ denote the diameter of any set A. Also, for any $\alpha > 0$ we will let $\mathcal{H}^\alpha$ denote the corresponding Hausdorff measure.

Theorem 1 [The Vitali covering theorem.] *Let $E \subset X$ and let Φ be a Vitali class for E consisting of closed subsets of E. Then there is a (finite or infinite) sequence $\{U_i\}$ of disjoint elements of Φ such that for each $\alpha > 0$ either $\sum_i |U|_i^\alpha = +\infty$ or $\mathcal{H}^\alpha(E \setminus \bigcup_i U_i)$.*

Proof. We may assume that all elements of Φ have diameter at most 1 since we may throw away the others. We will define the sets U_i recursively, starting with an arbitrary U_1. Suppose we have chosen $U_1, U_2, \ldots, U_m$. Let

$$d_m := \sup\{|U| \, |U \in \Phi \text{ and } U \cap U_i = \emptyset \text{ for } i = 1, \ldots, m\}.$$

If $d_m = 0$ then stop. Otherwise choose U_{m+1} such that

$$|U_{m+1}| > \frac{1}{2} d_m$$

and

$$U_{m+1} \cap U_i = \emptyset \text{ for } i = 1, \ldots, m.$$

For each i let B_i be a ball with center in U_i and radius $3|U_i|$. Suppose that

$$\sum |U_i|^s < \infty$$

for some $s > 0$. We claim that this implies that for each k we have

$$E \setminus \bigcup_{i=1}^{k} U_i \subset \bigcup_{j=k+1}^{\infty} B_j. \tag{1}$$

Indeed, let $x \in E \setminus \bigcup_{i=1}^k U_i$. Since the U_i are closed we have

$$d(x, \bigcup_{i=1}^{k} U_i) > 0.$$

The Vitali property implies that there is some $U \in \Phi$ with $x \in U$ and $U \cap U_i = \emptyset$ for all $i \leq k$, Since $\sum |U_i|^s < \infty$ and $|U_m| > \frac{1}{2} d_m$, there is a first index m with $|U| > d_m$. This implies that $U \cap U_j \neq \emptyset$ for some $j < m$ for otherwise d_m would have to have been larger than its definition.

Let x_j be the center of B_j. Let y be any element of $U \cap U_j$ Then for any $z \in U$ we have

$$d(y, z) \leq |U| \leq d_m \leq 2|U_j|$$

so

$$U \subset B_j.$$

In particular, $x_j \in B_j$. But since $U \cap U_i = \emptyset$ for $i \leq k$ we must have $j > k$. this proves (1).

It follows from (1) that

$$\mathcal{H}^s(E \setminus \bigcup_{i=1}^{k} U_i) \leq \mathcal{H}^s\left(t \bigcup_{j=k+1}^{\infty} B_j\right) \leq \sum_{j=k+1}^{\infty} |B_j|^s \leq 3^s \sum_{j=k+1}^{\infty} |U_j|^s$$

which approaches 0 as $k \to \infty$. $\square$

Corollary 2 *Let Φ be a family of closed balls (for some norm on $\mathbb{R}^m$ and let $A \subset \mathbb{R}^m$ be a set such that each point in A is contained in a member of Φ of arbitrarily small radius. Then there is a sequence of disjoint elements of Φ whose union covers all of A except for a set of Lebesgue measure zero.*

Proof. We may assume that all the balls have radius at most one. Let $B(x, r)$ denote the ball of radius r centered at the point x, so, for example, $B(0, 1)$ denotes the ball of radius one centered at the origin. Let

$$A_1 := A \cap B(0, 1).$$

Take $E = A_1$ in the Vitali covering theorem and choose a corresponding sequence of balls $U_i^1 \in \Phi$ with the additional property that $A_1 \cap U_i \neq \emptyset$. Since $\bigcup U_i^1$ is a bounded subset of $\mathbb{R}^m$ and the union is disjoint, we know that $\sum |U_i^1|^m < \infty$, so only the second alternative in the theorem can hold. (Recall that up to an overall constant, $\mathcal{H}^m$ coincides with Lebesgue measure.) Choose an index n_1 such that $\bigcup_{i=1}^{n_1}$ covers all of A_1 up to a set of measure $\frac{1}{2}$,.

We now recursively define A_j, n_j and pairwise disjoint $U_i^i \in \Phi$ such that

$$lambda^m \left(A_j \setminus \bigcup_{i=1}^{n_j} U_i^j\right) < \frac{1}{2^j}$$

for $j < k$ where λ^m denotes Lebesgue measure. Set

$$A_k := (A \cap B(0, k)) \setminus \bigcup j < k, i \leq n_j U_i^j.$$

Since the balls U_i^j are closed, the family Φ_k of members of Φ which are disjoint from from the U_i^j with $j < k$ form a Vitali class for A_k. So we can find n_k and U_i^k with

$$\lambda^m \left(A_k setminus \bigcup_{i=1}^{n_k} U_i^j\right) < \frac{1}{2^k}.$$

The set

$$\bigcap_k A_k$$

has Lebesgue measure zero, and every point of A is either contained in this null set or is contained in one of the disjoint collection of balls U_i^j. $\square$

2 The Lebesgue differentiation theorem.

Theorem 3 **[The Lebsgue differentiation theorem.]** *Let f be a Lebesgue integrable function on $\mathbb{R}^{>}$. Then for almost all $x \in \mathbb{R}^m$ the following holds: For every sequence $\{K_n\}$ of balls which contain x and whose diameter converges to 0 we have*

$$\lim_{n\to 0} \frac{\int_{K_n} f(y)dy}{\lambda^m(K_n)} = f(x). \tag{2}$$

Proof. We may assume that f is real valued and non-negative. Define the function $\tilde{f}$ by

$$\tilde{f}(x) := \lim_{\delta\to 0} \sup \left\{ \frac{\int_K f(y)dy}{\lambda^m(K)} | \ K \text{ is a ball containing } x \text{ with radius } \ r < \delta \right\}.$$

We first claim that the function $\tilde{f}$ is measurable. Indeed, if we take the balls to be open and define

$$s_\delta(x) = \sup \left\{ \frac{\int_K f(y)dy}{\lambda^m(K)} | \ K \text{ is a ball containing } x \text{ with radius } \ r < \delta \right\},$$

then the sets where $s_\delta(x) > c$ are open. In other words s_δ is lower semi-continuous and letting the δ converge to zero through a countable set we see that $\tilde{f}$ is the limit of a sequence of lower semi-continuous functions.

Let $s > 0$ and consider the set

$$A_s := \{x | f(x) < s < \tilde{f}(x)\|.$$

We wish to show that A_r has measure zero. Suppose not. Then

$$\int_{A_r} f(x)dx < \int_{A_r} rdx = r\lambda^m(A_r).$$

So we can find an open set $U \supset A_r$ with

$$\int_U f(x)dx < r\lambda^m(A_r).$$

Let Φ be the collection of all closed balls K contained in U such that

$$\frac{\int_K f(y)dy}{\lambda^m(K)} > r.$$

By the definition of A_r, each point of A_r is an element of an arbitrarily small member of Φ. In other words, Φ is a Vitali class for A_r. By Corollary 2, there is a disjoint sequence $\{K_i\}$ of elements of Φ which cover A_r up to a set of measure zero. Therefore

$$r\lambda^m(A_r) > \int_U f(y)dy \geq \int_{\bigcup K_i} f(y)dy = \sum_i \int_{K_i} f(y)dy > r \sum \lambda^m(K_i) \geq r\lambda^m(A_r),$$

a contradiction. So A_r has measure zero. Let r range over the rationals and taking $A = \bigcup A_r$ shows that (2) with lim replaced by lim sup. But the same argument yields lim inf and hence proves the theorem. $\square$

Corollary 4 [Fundamental theorem of the calculus according to Lebesgue.] *Let f be a Lebesgue integrable function on the real line. Let $F(x) = \int_{-\infty}^{x} f(t)dt$. Then F is differentiable almost everywhere with $F'(x) = f(x)$ a.e..*

Proof. If x is a point where the conclusion of the theorem holds we may take the "balls" containing x to be intervals with x at the right or left end points, in which case the conclusion says that the right and left handed derivatives exist and equal $f(x)$. □.

Corollary 5 [Integration by parts according to Lebesgue.] *Let f and g be Lebesgue integrable and let*

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad G(x) = \int_{-\infty}^{x} g(t)dt.$$

Then

$$\int_a^b f(x)G(x)dx = FG|_a^b - \int_a^b g(x)F(x)dx.$$

Proof using Fubini.

$$\begin{aligned}
\int_a^b f(x)G(x)dx &= \int_a^b [F(a) + \int_a^x f(y)dy]g(x)dx \\
&= \int_a^b [F(a) + \int_a^b f(y)\mathbf{1}_{[a,x]}(y)dy]g(x)dx \\
&= F(a)[G(b) - G(a)] + \int_a^b \int_a^b f(y)\mathbf{1}_{[a,x]}(y)g(x)dydx \\
&= F(a)[G(b) - G(a)] + \int_a^b \int_a^b f(y)\mathbf{1}_{y<x}(y)g(x)dxdy \quad \text{by Fubini} \\
&= F(a)[G(b) - G(a)] + \int_a^b \int_a^b f(y)\mathbf{1}_{[a,x]}(y)g(x)dydx \quad \text{by Fubini} \\
&= F(a)[G(b) - G(a)] + \int_a^b \int_a^b f(y)dy \int_y^b g(g)dxdy \quad \text{by Fubini} \\
&= F(a)[G(b) - G(a)] + \int_a^b \int_a^b [G(b) - G(y)]f(y)dy[\\
&= F(a)[G(b) - G(a)] + G(b)[F(b) - F(a)] - \int_a^b f(y)G(y)dy. \quad \square
\end{aligned}$$

3 Rademacher's theorem in one dimension.

Theorem 6 *Let $F : [a, b] \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constant L.Then*

$$F(x) = \int_a^x f(x)dx$$

where f is a Lebesgue-integrable function with $|f| < L$ a.e. So every Lipschitz function on $\mathbb{R}$ is differentiable almost everywhere.

Proof via the Riesz representation theorem. We may assume that $[a,b]=[0,1]$. For any step function

$$s=\sum_i \alpha_i \mathbf{1}[a_i,a_{i+1}]$$

define

$$I(s):=\sum_i \alpha_i[F(a_{i+1})-F(a_i).$$

This is clearly linear in s and, by the Lipschitz condition,

$$|I(s)|\leq L\sum_i |\alpha_i|a_{i+1}-a_i|=L\|s\|_1=L(\mathbf{1}_{[0,1)},|s|)_2\leq L\|s\|_2$$

where we are using the L_1 and L_2 norms for $[0,1]$ and the last inequality is the Caucy-Schwarz inequality. So I is continuous in the L_2 norm and we know that the step functions are dense. So I extends to a unique continuous linear function on $L_2([0,1])$. By the Riesz representation theorem there exists a (unique) function f such that

$$I(\psi)=(f,\psi)\quad \forall\ \psi\in L_2([0,1]).$$

If we choose $\psi=\mathbf{1}_{[x,y]}$ for some $0\leq x<y\leq 1$ this says that

$$F(y)-F(x)=\int_x^y f(t)dt.$$

We now apply Corollary 4. □

4 The Rademacher theorem.

Theorem 7 [Rademacher.] *Let U be an open subset of $\mathbb{R}^m$ and let $f:U\to\mathbb{R}^n$ be a Lipschitz map. Then f is differentiable almost everywhere.*

Proof. f is differentiable if and only if each of the coordinates of f thought of as a map from $\mathbb{R}^m$ to $\mathbb{R}$ is differentiable, so it is enough to prove the theorem when $n=1$.

For each $v\in\mathbb{R}^m$ with $\|v\|=1$ define the directional derivative in the direction v at $x\in U$ by

$$D_vf(x)=\lim_{t\searrow 0}\frac{f(x+tv)--f(x)}{t}\quad \text{whenever the limit exists.}$$

Since f is continuous, we may restrict our attention to t rational. Therefore the set

$$A_v:=\{x|D_v(x)\text{ exists}\}$$

is measurable and the function $x\mapsto D_vf(x)$ restricted to this subset is measurable, and we may extend it to all of U by setting it equal to zero on the complement of A_v. Let us consider the orthogonal decomposition of $\mathbb{R}^m$ into

$$\mathbb{R}\cdot v\oplus v^\perp.$$

For each $y \in v^\perp$, the intersection of $A_v^c \cap U$ with the line parallel to $\mathbb{R} \cdot v$ and passing through y has one dimensi0nal measure 0 by Theorem 6. Hence by Fubini, the set $A^c \cap U$ has measure zero. In other words, each directional derivative exists almost everywhere. In particular, the partial derivatives with respect to the coordinates exist almost everywhere so the covector (row vector)

$$df(x) := \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

exists almost everywhere.

Lemma 1 *For each fixed v we have*

$$D_v f(x) = df(x) \cdot v \tag{3}$$

almost everywhere.

Proof of the lemma. It is enough to show that both sides give the same answer when integrate against any smooth function of compact support, g, i.e. to show that

$$\int_{\mathbb{R}^m} (D_v f(x)) g(x) dx = \int_{\mathbb{R}^m} (df(x) \cdot v) g(x) dx$$

for all smooth g of compact support contained in U. Decomposing $\mathbb{R}^m$ into $\mathbb{R} \cdot v \oplus v^\perp$ as before, it is enough by Fubini to show that we have

$$\int_{v^\perp} \int_{\mathbb{R}} D_v f(y + tv) g(y + tv) dt dy = \int_{v^\perp} \int_{\mathbb{R}} (df \cdot v)(y + tv) dt dy.$$

Lebesgue's integration by parts formula, Corollary 5, says that

$$\int_{\mathbb{R}} D_v f(y + tv) g(y + tv) dt = - \int_{\mathbb{R}} f(y + tv) D_v g dt.$$

But since g is smooth, we have

$$D_v g = dg \cdot v$$

so

$$\int_{\mathbb{R}^m} (D_v f(x)) g(x) dx = - \sum_{i=1}^m v_i \int_{\mathbb{R}^m} f \cdot \frac{\partial g}{\partial x_i} dx.$$

Applying Corollary 5 again to each of the coordinate directions and using Fubini, this last expression can be written as

$$\int_{\mathbb{R}^m} (df(x) \cdot v) g(x) dx$$

completing the proof of the Lemma.

Back to the proof of the theorem. Choose a countable dense set of vectors v_k on the unit sphere. Let

$$A_k = \{x \in U | D_{v_k} f \text{ exists, } df(x) \text{ exists, and} D_{v_k} f(x) = df(x) \cdot v\}.$$

Then $A_k^c \cap U$ has measure zero hence so does $A^c \cap U$ where

$$A := \int_k A_k.$$

We claim that f is differentiable at all $x \in A$¿ More precisely, we claim that for all $x \in A$ we have

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

Indeed, for $w \neq 0$ let

$$u := \frac{1}{\|w\|}$$

so

$$w = \|w\|u.$$

and

$$f(x+w) - f(x) = f(x + \|w\|u) - f(x).$$

For any $v = v_i$ we have

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

By the Lipschitz property of f the second term on the right is bounded in absolute value by $L\|u - v\|$. We want to choose v such that this term is $< \frac{1}{2}\epsilon$. For this v, we may find a sufficiently small δ such the first expression differs from $df(x) \cdot v$ by less than ϵ for all $\|w\| < \delta$. Now $df(x)$ is a fixed linear function so $df(x)(u-v)$ is bounded in absolute value by $M\|u-v\|$ for some constant M. We may also make our original choice of v so that $M\|u-v\| < \frac{1}{2}\epsilon$. So we proceed as follows. Cover the unit sphere by finitely many balls of radius η where $(M+L)\eta < \epsilon$. In each of these balls choose one v from among the v_i. Then choose a δ small enough so that for all of these v the first term on the right above differs from the corresponding $df(x) \cdot v$ by less than ϵ for all $\|w\| < \delta$. So for any $\epsilon > 0$ we can choose $\delta > 0$ so that

$$\frac{f(x+w) - f(x)}{\|w\|} = df(x) \cdot u + \text{error}$$

where the error is bounded in absolute value by 2ϵ. This says that

$$f(x+w) - f(x) = df(x) \cdot w + \text{Error}$$

where the absolute value of the Error is $< 2\epsilon\|w\|$ which is what we want to prove. □