Measure Theory and Fine Properties of Functions

Abhi Wadhwa
University of Southern California

Notes for an independent study based on Evans & Gariepy, Fall 2024.

Contents

1 August 26, 2024 **5**
1.1 Outer measures . . . 5
1.2 Carathéodory's criterion . . . 5
1.3 σ-algebras and Borel sets . . . 6
1.4 Radon measures . . . 6
1.5 Measurable functions . . . 7

2 August 28, 2024 **7**
2.1 Egoroff's theorem . . . 7
2.2 Lusin's theorem . . . 8
2.3 The Lebesgue integral . . . 8
2.4 Monotone convergence theorem . . . 8
2.5 Fatou's lemma . . . 9
2.6 Dominated convergence . . . 9

3 September 4, 2024 **9**
3.1 Product measures and Fubini's theorem . . . 9
3.1.1 Lebesgue measure . . . 10
3.2 Vitali's covering theorem . . . 10
3.3 Besicovitch's covering theorem . . . 11

4 September 9, 2024 **11**
4.1 Differentiation of Radon measures . . . 11
4.2 Lebesgue points . . . 12
4.3 Approximate continuity . . . 12

5 September 11, 2024 **13**
5.1 Riesz representation theorem . . . 13
5.2 Weak* convergence . . . 14

6 September 16, 2024 **15**
6.1 Hausdorff pre-measure and Hausdorff measure . . . 15
6.2 Hausdorff dimension . . . 15

7 September 18, 2024 **16**
7.1 Steiner symmetrization . . . 16
7.2 The isodiametric inequality . . . 17
7.3 $\mathcal{H}^n = \mathcal{L}^n$ . . . 18
7.4 Densities . . . 18
7.5 Connection to functions . . . 19

8 September 30, 2024 **19**
8.1 Lipschitz Functions . . . 19
8.2 Rademacher's Theorem . . . 20
8.3 Linear Maps and Jacobians . . . 21
8.4 The Area Formula . . . 22
8.5 Surface area as an application . . . 23

9 October 2, 2024 23
9.1 The Coarea Formula . . . 23
9.2 Why this is a nonlinear Fubini theorem . . . 23
9.3 The case $m = 1$ . . . 24
9.4 Proof sketch of the coarea formula . . . 24

10 October 7, 2024 25
10.1 Weak Derivatives . . . 25
10.2 Sobolev Spaces . . . 25
10.3 Approximation by Smooth Functions . . . 26

11 October 9, 2024 26
11.1 Traces . . . 26
11.2 Extensions . . . 27
11.3 Gagliardo–Nirenberg–Sobolev ($p < n$) . . . 27
11.4 Morrey's Inequality ($p > n$) . . . 28
11.5 The borderline $p = n$ . . . 28
11.6 Embedding summary . . . 28

12 October 14, 2024 28
12.1 Rellich–Kondrachov Compactness . . . 28
12.2 Capacity . . . 29
12.3 Quasicontinuity and Precise Representatives . . . 30
12.4 Differentiability on Lines . . . 30

13 October 21, 2024 30
13.1 Definitions and first properties . . . 31
13.2 The structure theorem . . . 31
13.3 Approximation and compactness . . . 32
13.4 Traces and extensions . . . 32

14 October 23, 2024 33
14.1 Coarea formula for BV . . . 33
14.2 Isoperimetric inequalities . . . 33

15 October 28, 2024 34
15.1 Sets of finite perimeter and the reduced boundary . . . 34
15.2 The blow-up theorem . . . 34
15.3 Measure-theoretic boundary and Gauss–Green . . . 35
15.4 Pointwise properties of BV functions . . . 35
15.5 Essential variation on lines . . . 36
15.6 Criterion for finite perimeter . . . 36

16 November 4, 2024 36
16.1 L^p differentiability . . . 36
16.2 Approximate differentiability . . . 37

17 November 6, 2024 **37**
17.1 Differentiability a.e. for $W^{1,p}$, $p > n$ . . . 38
17.2 Convex functions . . . 38
17.3 Aleksandrov's theorem . . . 38
17.4 Whitney's extension theorem . . . 39
17.5 Approximation by C^1 functions . . . 39

1 August 26, 2024

Pop quiz: what is a measure? If you answered "a function on sets," fine, but that skips the interesting part. The whole game in this chapter is that you *cannot* define a measure on all subsets of $\mathbb{R}^n$ and keep the properties you want. Vitali killed that dream in 1905. So you start with something weaker, an outer measure, and then carve out the sets that behave.

I got a coffee from Dulce on Figueroa before class and nearly burned my tongue off. Good omen for the semester.

Today we set up outer measures and Carathéodory's criterion.

1.1 Outer measures

Outer measures are cheap. You can define them on *every* subset. The cost is you only get subadditivity, not full additivity.

Definition 1.1 (Outer measure)**.** An **outer measure** on a set X is a function $\mu^*\colon \mathcal{P}(X) \to [0, \infty]$ satisfying

1. $\mu^*(\varnothing) = 0$,
2. (Monotonicity) $A \subseteq B \implies \mu^*(A) \le \mu^*(B)$,
3. (Countable subadditivity) $\mu^*\big(\bigcup_{k=1}^{\infty} A_k\big) \le \sum_{k=1}^{\infty} \mu^*(A_k)$.

That's it. No σ-algebra needed yet. We are measuring everything, just badly. Subadditivity is what you pay for working with all of $\mathcal{P}(X)$.

1.2 Carathéodory's criterion

Now we want to find the "well-behaved" sets, the ones where the outer measure actually splits additively.

Definition 1.2 (μ^*-measurability)**.** A set $A \subseteq X$ is **μ^*-measurable** if for every $E \subseteq X$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A).$$

I stared at this definition for an embarrassingly long time the first week. Why test against *every* set E? Brian explained it over lunch in the RTH courtyard: you already get $\le$ from subadditivity for free. So the real content is $\ge$. You are asking that A acts as a "clean cut." Slice any test set E along A, and no mass leaks out. That is the whole condition.

Theorem 1.3 (Carathéodory)**.** *Let μ^* be an outer measure on X. Then the collection*

$$\mathcal{M} = \{A \subseteq X : A \textit{ is } \mu^*\textit{-measurable}\}$$

is a σ-algebra, and μ^* restricted to $\mathcal{M}$ is a (complete) measure.

Proof sketch. Complements are free — the criterion is symmetric in A and A^c. For finite unions: take $A_1, A_2 \in \mathcal{M}$ and test against E. Split first by A_1, then split $E \cap A_1^c$ by A_2. You get three pieces that reassemble by subadditivity.

Now the hard part. Countable unions. Let $A = \bigcup_k A_k$ with the A_k disjoint (WLOG — replace A_k with $A_k \setminus \bigcup_{j<k} A_j$, which stays in $\mathcal{M}$ by the finite case). Set $B_n = \bigcup_{k=1}^{n} A_k$. Then for any E,

$$\mu^*(E \cap B_n) = \sum_{k=1}^{n} \mu^*(E \cap A_k)$$

by induction. So

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \setminus B_n) \geq \sum_{k=1}^{n} \mu^*(E \cap A_k) + \mu^*(E \setminus A),$$

since $A \supseteq B_n$. Let $n \to \infty$ and do the usual epsilon management. Completeness is free: any μ^*-null set passes the criterion trivially. □

Remark 1.4. Completeness coming for free matters more than it first looks. Build Lebesgue measure via Carathéodory and it is automatically complete. No need for the separate completion step you would do starting from Borel sets. I didn't appreciate that until I tried the Borel-first construction for a homework set and had to bolt on completion by hand.

1.3 σ-algebras and Borel sets

Definition 1.5. A σ-algebra on X is a collection $\mathcal{A} \subseteq \mathcal{P}(X)$ containing $\varnothing$, closed under complements, and closed under countable unions.

Nothing to see here. Closure under countable intersections follows from De Morgan.

Definition 1.6 (Borel σ-algebra)**.** For a topological space X, the **Borel σ-algebra** $\mathcal{B}(X)$ is the σ-algebra generated by the open sets.

In $\mathbb{R}^n$ you can equivalently generate $\mathcal{B}(\mathbb{R}^n)$ by open balls, closed sets, compact sets, or half-spaces $\{x : x_i < a\}$. Evans doesn't dwell on this. Why would he? Every real analysis book covers it, and Folland (Section 1.2) does it better anyway.

1.4 Radon measures

I'm trying to internalize this hierarchy now. Otherwise I will be looking it up every five minutes for the rest of the semester, and that got old fast during Schlag's 425a notes over the summer.

Definition 1.7. A **Borel measure** on $\mathbb{R}^n$ is a measure defined on $\mathcal{B}(\mathbb{R}^n)$.

Definition 1.8 (Borel regular)**.** A Borel measure μ is **Borel regular** if for every $A \subseteq \mathbb{R}^n$, there exists a Borel set $B \supseteq A$ with $\mu(B) = \mu^*(A)$.

Definition 1.9 (Radon measure)**.** A Borel regular measure μ on $\mathbb{R}^n$ is a **Radon measure** if $\mu(K) < \infty$ for every compact $K \subseteq \mathbb{R}^n$.

So: Radon = Borel regular + locally finite. Evans uses "Radon" as his default notion of "good measure on $\mathbb{R}^n$." Lebesgue measure $\mathcal{L}^n$ is the main example. Hausdorff measures on rectifiable sets are another, but we won't see those until Chapter 2.

Proposition 1.10 (Inner/outer regularity)**.** *If μ is a Radon measure on $\mathbb{R}^n$, then for every measurable set A:*

1. *(Outer regularity)* $\mu(A) = \inf\{\mu(U) : U \supseteq A,\ U \text{ open}\}$.
2. *(Inner regularity)* $\mu(A) = \sup\{\mu(K) : K \subseteq A,\ K \text{ compact}\}$.

Proof: Evans pp. 5–7. Outer regularity uses Borel regularity. Inner regularity uses σ-finiteness (exhaust by balls $B(0, k)$).

I spent two hours on the inner regularity proof before spotting the trick: first prove it for finite-measure sets, then extend by writing $A = \bigcup_k (A \cap B(0, k))$ and using continuity from below. Two hours for that. I could have been doing literally anything else.

1.5 Measurable functions

Definition 1.11. Let $(X, \mathcal{A})$ be a measurable space. A function $f\colon X \to [-\infty, \infty]$ is **$\mathcal{A}$-measurable** if $\{f > t\} \in \mathcal{A}$ for all $t \in \mathbb{R}$.

You could equivalently use $\{f \geq t\}$, $\{f < t\}$, or $\{f \leq t\}$. All the same: complements and countable intersections/unions convert between them. Evans uses the $\{f > t\}$ convention.

Proposition 1.12. *Measurable functions are closed under:*

1. *Pointwise limits: if $f_k \to f$ pointwise, each f_k measurable, then f is measurable.*
2. *Sums, products,* max, min, $|f|$.
3. *Countable* sup *and* inf*: $\sup_k f_k$ and $\inf_k f_k$ are measurable.*

Proof of (iii). $\{\sup_k f_k > t\} = \bigcup_k \{f_k > t\}$. Countable union of measurable sets. Done. The inf case is similar. □

Remark 1.13. Uncountable suprema can fail to be measurable. Pointwise limits work (sequential = countable) but arbitrary sups do not. Measurability is a countable-operations game.

TODO: Read Evans pp. 7–14 more carefully for the approximation by simple functions stuff.

2 August 28, 2024

Quick question before we start: if a sequence of measurable functions converges pointwise, is the convergence "almost uniform"? Turns out yes, as long as you are on a finite measure space. That is Egoroff.

Today we do Egoroff and Lusin, then start on integration.

2.1 Egoroff's theorem

The statement says convergence of measurable functions is "almost uniform." The proof looked like pure epsilon management at first glance. Then I drew the sets $E_{m,n}$ on the whiteboard in KAP 414 and the geometry clicked.

Theorem 2.1 (Egoroff)**.** *Let μ be a finite measure on X, and let f_k, f be measurable functions with $f_k \to f$ a.e. Then for every $\varepsilon > 0$, there exists a measurable set A with $\mu(X \setminus A) < \varepsilon$ such that $f_k \to f$* ***uniformly*** *on A.*

Proof. For each $m, n \in \mathbb{N}$, define

$$E_{m,n} = \bigcup_{k=n}^{\infty} \{x : |f_k(x) - f(x)| \geq \tfrac{1}{m}\}.$$

Then $E_{m,n} \downarrow$ as $n \to \infty$, and the a.e. convergence gives $\mu\big(\bigcap_n E_{m,n}\big) = 0$. So by continuity from above (here we use $\mu(X) < \infty$!), for each m pick n_m with $\mu(E_{m,n_m}) < \varepsilon/2^m$. Set $A = X \setminus \bigcup_m E_{m,n_m}$. Then $\mu(X \setminus A) < \varepsilon$, and on A we have $|f_k - f| < 1/m$ for all $k \geq n_m$. That's uniform convergence. □

Finiteness of μ is doing real work. Counterexample: $f_k = \mathbf{1}_{[k,k+1]}$ on $\mathbb{R}$ with Lebesgue measure. Converges to 0 pointwise everywhere. But not uniformly on any set of finite complement.

TODO: check if σ-finiteness suffices, or whether we truly need $\mu(X) < \infty$. I think you need finiteness on the set where convergence happens. Brian thinks σ-finiteness works if you are careful, but I have not verified it.

2.2 Lusin's theorem

Theorem 2.2 (Lusin). *Let μ be a Radon measure on $\mathbb{R}^n$ and $f\colon \mathbb{R}^n \to \mathbb{R}$ measurable. Then for every $\varepsilon > 0$ and every measurable set A with $\mu(A) < \infty$, there exists a compact set $K \subseteq A$ with $\mu(A \setminus K) < \varepsilon$ such that $f|_K$ is continuous.*

Proof: Evans pp. 15–16. Use inner regularity to approximate level sets $\{a_i \le f < a_{i+1}\}$ by compact sets. On the union of those compact sets, f is a staircase that can be refined to continuous. Tietze extension shows up at one point.

Lusin is philosophically nice: measurable functions are "almost continuous." But I have not needed it as often as Egoroff in practice. It comes back for the Riesz representation theorem later.

2.3 The Lebesgue integral

Evans builds the integral the standard way. No surprises if you have seen Royden or Folland.

Definition 2.3. For $f\colon X \to [0,\infty]$ measurable,

$$\int_X f \,\mathrm{d}\mu = \sup\left\{ \int_X s \,\mathrm{d}\mu : 0 \le s \le f,\ s \text{ simple} \right\},$$

where for a simple function $s = \sum_{i=1}^N a_i \mathbf{1}_{A_i}$ (with the A_i measurable, disjoint), $\int_X s \,\mathrm{d}\mu = \sum_{i=1}^N a_i\, \mu(A_i)$.

For general measurable f, write $f = f^+ - f^-$ and define $\int f \,\mathrm{d}\mu = \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu$, provided at least one side is finite. If both are finite, f is **integrable** ($f \in L^1$).

2.4 Monotone convergence theorem

This is the foundation. DCT and Fatou both reduce to this.

Theorem 2.4 (Monotone Convergence). *Let $0 \le f_1 \le f_2 \le \cdots$ be measurable with $f_k \nearrow f$ pointwise. Then*

$$\int_X f \,\mathrm{d}\mu = \lim_{k\to\infty} \int_X f_k \,\mathrm{d}\mu.$$

Proof. The $\le$ direction: $f_k \le f$ implies $\int f_k \le \int f$, so $\lim \int f_k \le \int f$.

For $\ge$: pick any simple s with $0 \le s \le f$ and fix $0 < t < 1$. Define $A_k = \{x : f_k(x) \ge t \cdot s(x)\}$. Then $A_k \nearrow X$ (since $f_k \nearrow f \ge s > ts$ wherever $s > 0$). So

$$\int_X f_k \,\mathrm{d}\mu \ge \int_{A_k} f_k \,\mathrm{d}\mu \ge t \int_{A_k} s \,\mathrm{d}\mu.$$

Since s is simple, $\int_{A_k} s \,\mathrm{d}\mu \to \int_X s \,\mathrm{d}\mu$ by continuity from below. So $\lim \int f_k \ge t \int s$. Let $t \nearrow 1$, then sup over simple $s \le f$. $\square$

I initially tried to use DCT to prove this. Circular, obviously. MCT has to come first. I felt dumb for about ten minutes, then realized half the class had made the same mistake based on what Brian told me after.

2.5 Fatou's lemma

Lemma 2.5 (Fatou). *If $f_k \geq 0$ are measurable, then*

$$\int_X \liminf_{k\to\infty} f_k \, d\mu \leq \liminf_{k\to\infty} \int_X f_k \, d\mu.$$

Proof. Set $g_k = \inf_{j\geq k} f_j$. Then $g_k \nearrow \liminf f_k$ and $g_k \leq f_k$, so $\int g_k \leq \int f_k$. Apply MCT to the left side. □

Short. The inequality can be strict: $f_k = k \cdot \mathbf{1}_{(0,1/k)}$ on $\mathbb{R}$. Each integral is 1, but the pointwise liminf is 0 a.e.

2.6 Dominated convergence

Theorem 2.6 (Dominated Convergence). *Let $f_k \to f$ a.e., with $|f_k| \leq g$ a.e. for some integrable g. Then f is integrable and*

$$\lim_{k\to\infty} \int_X f_k \, d\mu = \int_X f \, d\mu.$$

Proof. The trick is to apply Fatou twice. $g + f_k \geq 0$ and $g - f_k \geq 0$, so:

$$\int (g+f) \leq \liminf \int (g+f_k), \qquad \int (g-f) \leq \liminf \int (g-f_k).$$

Since $\int g < \infty$, cancel:

$$\int f \leq \liminf \int f_k, \qquad -\int f \leq -\limsup \int f_k.$$

So $\limsup \int f_k \leq \int f \leq \liminf \int f_k$. We're done. □

The "apply Fatou twice" move is one of those things that is obvious after you have seen it and genuinely confusing before. The domination $|f_k| \leq g$ does all the work: it makes the Fatou inequalities point the right way and keeps everything integrable.

Remark 2.7. You can't drop the domination. Same counterexample: $f_k = k \cdot \mathbf{1}_{(0,1/k)}$ converges to 0 pointwise but $\int f_k = 1$. No integrable dominator exists.

3 September 4, 2024

I walked to the classroom on Monday. Labor Day. Nobody there. Sat on the steps outside KAP for five minutes feeling ridiculous before checking my phone.

Today: product measures, Fubini–Tonelli, and then the covering theorems that set up the second half of the chapter.

3.1 Product measures and Fubini's theorem

Definition 3.1 (Product σ-algebra). If $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces, the **product σ-algebra** $\mathcal{A} \otimes \mathcal{B}$ on $X \times Y$ is generated by measurable rectangles $A \times B$ ($A \in \mathcal{A}$, $B \in \mathcal{B}$).

Theorem 3.2 (Product measure). *If μ and ν are σ-finite measures, there exists a unique measure $\mu \otimes \nu$ on $\mathcal{A} \otimes \mathcal{B}$ with $(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B)$.*

σ-finiteness is needed for uniqueness. Without it you can cook up weird examples. (Halmos has one on counting measure on an uncountable set, if I remember right.)

Theorem 3.3 (Fubini–Tonelli)**.** *Let μ, ν be σ-finite measures.*

1. ***(Tonelli)*** *If $f \geq 0$ is $\mathcal{A} \otimes \mathcal{B}$-measurable, then*

$$\int_{X \times Y} f \, \mathrm{d}(\mu \otimes \nu) = \int_X \left(\int_Y f(x, y) \, \mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) = \int_Y \left(\int_X f(x, y) \, \mathrm{d}\mu(x) \right) \mathrm{d}\nu(y).$$

2. ***(Fubini)*** *If $f \in L^1(\mu \otimes \nu)$, the same holds, and the inner integrals are finite for a.e. value of the outer variable.*

The proof is a slog but the strategy is clean: verify for indicators of measurable rectangles (trivial), extend to simple functions by linearity, extend to nonneg measurable by MCT, extend to integrable by splitting $f = f^+ - f^-$.

Tonelli is for $f \geq 0$. No integrability needed. Fubini needs $f \in L^1$. Common trick: use Tonelli on $|f|$ to check integrability, then Fubini on f.

3.1.1 Lebesgue measure

Definition 3.4. n-dimensional **Lebesgue measure** $\mathcal{L}^n$ is the product $\underbrace{\mathcal{L}^1 \otimes \cdots \otimes \mathcal{L}^1}_{n}$, where $\mathcal{L}^1$ is the completion of the Borel measure on $\mathbb{R}$ assigning to each interval $[a, b]$ its length $b - a$.

Evans constructs $\mathcal{L}^n$ via the outer measure

$$\mathcal{L}^n_*(A) = \inf \left\{ \sum_{k=1}^{\infty} |Q_k| : A \subseteq \bigcup_{k=1}^{\infty} Q_k, \ Q_k \text{ cubes} \right\},$$

where $|Q|$ is the volume of the cube. Carathéodory then gives the σ-algebra and measure. The two constructions agree on Borel sets.

Fact 3.5. *$\mathcal{L}^n$ is:*

1. *A Radon measure.*
2. *Translation invariant: $\mathcal{L}^n(A + x) = \mathcal{L}^n(A)$.*
3. *Scales correctly: $\mathcal{L}^n(rA) = r^n \mathcal{L}^n(A)$ for $r > 0$.*
4. *Uniquely characterized (up to scalar) by translation invariance on Borel sets.*

3.2 Vitali's covering theorem

Covering theorems. Not glamorous, but you literally cannot differentiate measures without them. I found this section painful on first reading. Like chewing on cardboard, except the cardboard turns out to be load-bearing.

Theorem 3.6 (Vitali)**.** *Let $\mathcal{F}$ be a collection of closed balls in $\mathbb{R}^n$ with $\sup\{\operatorname{diam}(B) : B \in \mathcal{F}\} < \infty$. Then there exists a countable subcollection of **disjoint** balls $\{B_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ such that*

$$\bigcup_{B \in \mathcal{F}} B \subseteq \bigcup_{i=1}^{\infty} \hat{B}_i,$$

where $\hat{B}_i$ is the ball with the same center as B_i but 5 times the radius.

The “$5r$-trick.” The proof is a greedy algorithm: pick the biggest ball (or close to it), throw out everything it intersects, repeat. Why 5 and not 3 or 7? Because if two balls overlap and one has radius at most the other’s, the smaller one sits inside the $5\times$ enlargement. The constant 5 is sharp for this particular greedy strategy. Nothing deeper.

There is also a refined version for Radon measures:

Theorem 3.7 (Vitali covering theorem for Radon measures)**.** *Let μ be a Radon measure, $A \subseteq \mathbb{R}^n$, and $\mathcal{F}$ a collection of closed balls such that for each $x \in A$, there exist balls in $\mathcal{F}$ centered at x with arbitrarily small radius (a “fine cover”). Then there exist disjoint balls $\{B_i\} \subseteq \mathcal{F}$ with*

$$\mu\left(A \setminus \bigcup_{i=1}^{\infty} B_i\right) = 0.$$

So a fine Vitali cover lets you cover μ-almost all of A with disjoint balls.

3.3 Besicovitch’s covering theorem

Theorem 3.8 (Besicovitch)**.** *There exists a constant N_n (depending only on dimension n) such that: if $A \subseteq \mathbb{R}^n$ is bounded and $\mathcal{F}$ is a collection of closed balls with a ball centered at each point of A, then there exist at most N_n countable subcollections $\mathcal{G}_1, \dots, \mathcal{G}_{N_n}$, each consisting of disjoint balls, such that*

$$A \subseteq \bigcup_{j=1}^{N_n} \bigcup_{B \in \mathcal{G}_j} B.$$

Besicovitch is harder than Vitali. Evans devotes pp. 30–35 to the proof. The advantage over Vitali: the balls do not get enlarged. The trade-off: bounded overlap (N_n families) instead of disjointness in one family. The exact value of N_n does not matter for anything we do. For $n = 1$ it is 2, for $n = 2$ it is around 19, and after that nobody memorizes the numbers.

The geometric core is a counting argument: at any point, how many balls of comparable radius can contain that point? The bound is dimensional. Annoying but mechanical.

I followed the proof once in lecture and could not reproduce it cold. I will not pretend otherwise.

TODO: redo the Besicovitch proof more carefully.

4 September 9, 2024

Evans derives Radon–Nikodym as a *consequence* of measure differentiation. That is the reverse of the functional-analysis approach you see in Rudin’s *Real and Complex Analysis*, Chapter 6. I think Evans’s way is more geometric and more natural for what follows (BV functions, sets of finite perimeter), but it makes this section of the chapter harder on a first reading.

4.1 Differentiation of Radon measures

Where the covering theorems pay off. The goal: given two Radon measures μ and ν, differentiate ν with respect to μ by looking at ratios on shrinking balls.

Definition 4.1. For Radon measures μ, ν on $\mathbb{R}^n$, define

$$D_\mu \nu(x) = \lim_{r \to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))},$$

when the limit exists (and $\mu(B(x,r)) > 0$ for small r).

Theorem 4.2 (Lebesgue–Besicovitch Differentiation)**.** *Let μ be a Radon measure on $\mathbb{R}^n$ and ν a (signed) Radon measure. Then:*

1. *$D_\mu \nu(x)$ exists and is finite μ-a.e.*
2. *The Lebesgue decomposition $\nu = \nu_{ac} + \nu_s$ (where $\nu_{ac} \ll \mu$ and $\nu_s \perp \mu$) satisfies $D_\mu \nu(x) = \frac{d\nu_{ac}}{d\mu}(x)$ μ-a.e.*
3. *$D_\mu \nu_s(x) = 0$ for μ-a.e. x.*

Part (iii) says the singular part is "invisible" to the derivative. I went through the proof twice and still find it hard to hold in my head all at once. The idea makes sense locally. The orchestration across all the sub-cases is what gets me.

Proof sketch. Evans constructs the Lebesgue decomposition *using* the derivative. Define $D^+ = \limsup_{r\to 0}$ and $D_- = \liminf_{r\to 0}$ of the ratio $\nu(B(x,r))/\mu(B(x,r))$. The main technical step: show that

$$\mu\{x : D^+\nu(x) > t > s > D_-\nu(x)\} = 0$$

for all $t > s$. This uses Besicovitch to build efficient covers at two different scales. Once you know $D^+ = D_-$ a.e., the limit exists. The identification with the Radon–Nikodym derivative is then a uniqueness argument. See Evans pp. 39–43. □

Corollary 4.3 (Lebesgue decomposition)**.** *Every signed Radon measure ν has a unique decomposition $\nu = \nu_{ac} + \nu_s$ with $\nu_{ac} \ll \mu$ and $\nu_s \perp \mu$.*

Corollary 4.4 (Radon–Nikodym)**.** *If $\nu \ll \mu$ (both Radon on $\mathbb{R}^n$), then there exists $f \in L^1_{\text{loc}}(\mu)$ with $\nu = f\mu$, i.e., $\nu(A) = \int_A f \, d\mu$ for all Borel A. And $f = D_\mu \nu$ a.e.*

4.2 Lebesgue points

Definition 4.5 (Lebesgue point)**.** Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. A point x is a **Lebesgue point** of f if

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, dy = 0.$$

This is stronger than saying the averages converge to $f(x)$. It says the L^1 oscillation around x vanishes. That distinction matters, and I confused them on a problem set.

Theorem 4.6. *If $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, then $\mathcal{L}^n$-a.e. x is a Lebesgue point of f.*

Proof idea. Apply Lebesgue–Besicovitch (Theorem 4.2) to $\nu(A) = \int_A |f(y) - c| \, dy$ for each rational c. This gives

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} |f(y) - c| \, dy = |f(x) - c| \quad \text{a.e.}$$

for each rational c. Take a countable intersection (one for each rational), then choose $c = c_k \to f(x)$ along rationals. You get the full Lebesgue point condition. □

4.3 Approximate continuity

Definition 4.7. A measurable function f is **approximately continuous** at x if there exists a measurable set A with density 1 at x (meaning $\lim_{r\to 0} \mathcal{L}^n(A \cap B(x,r))/\mathcal{L}^n(B(x,r)) = 1$) such that $f|_A$ is continuous at x.

Theorem 4.8. *If $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, then f is approximately continuous at $\mathcal{L}^n$-a.e. x.*

This follows from the Lebesgue point theorem: if x is a Lebesgue point, take $A = \{y : |f(y) - f(x)| < \varepsilon\}$ and check it has density 1. Details on Evans p. 46.

Lebesgue points, approximate limits, approximate continuity. These all come back in Chapters 5 and 6 when Evans discusses BV functions and the fine structure of derivatives. Filing them away for now and hoping I remember them when it matters.

5 September 11, 2024

Last lecture on Chapter 1. We finish with the Riesz representation theorem and weak* compactness for measures. This section connects measure theory to functional analysis. After the covering-theorem grind from last week, it is a relief to think about linear functionals again. I have a sandwich from Lemonade on Trousdale and I'm trying to pay attention.

5.1 Riesz representation theorem

This is why we built all the Radon measure machinery.

Theorem 5.1 (Riesz Representation). *Let $L\colon C_c(\mathbb{R}^n) \to \mathbb{R}$ be a positive linear functional (i.e., $L(f) \geq 0$ whenever $f \geq 0$). Then there exists a unique Radon measure μ on $\mathbb{R}^n$ such that*

$$L(f) = \int_{\mathbb{R}^n} f \, \mathrm{d}\mu \quad \textit{for all } f \in C_c(\mathbb{R}^n).$$

Here $C_c(\mathbb{R}^n)$ is continuous functions with compact support.

Proof outline. **Step 1.** Define μ on open sets by

$$\mu(U) = \sup\{L(f) : f \in C_c(\mathbb{R}^n),\ 0 \leq f \leq 1,\ \mathrm{spt}(f) \subseteq U\}.$$

Step 2. Extend to an outer measure via $\mu^*(A) = \inf\{\mu(U) : U \supseteq A,\ U \text{ open}\}$.

Step 3. Carathéodory gives a measure. Show it's Radon.

Step 4. Show $L(f) = \int f \, \mathrm{d}\mu$ by approximating f from below with simple functions built from the sets $\{f > t\}$.

Step 5. Uniqueness: if two Radon measures agree on C_c, they agree on open sets by inner regularity, hence on all Borel sets. □

I won't claim I have internalized every detail here. Step 1 is the key idea. The linear functional "tells you" what the measure of each open set should be, and you bootstrap from there.

Wait. Let me back up. The reason Step 1 works at all is that open sets are determined from below by continuous functions with compact support. That bridging observation is the entire engine connecting functional analysis and measure theory. Without it, you would have a linear functional and no way to turn it into a set function.

TODO: work through the proof more carefully, maybe try it for $\mathbb{R}^1$ first.

Remark 5.2. There is a signed version: every bounded linear functional on $C_0(\mathbb{R}^n)$ (continuous functions vanishing at infinity, with the sup norm) is represented by a signed Radon measure. Evans does not state this but it is standard (Rudin's *Real and Complex Analysis*, Chapter 6).

5.2 Weak* convergence

Definition 5.3. A sequence of Radon measures μ_k converges **weakly*** to μ (written $\mu_k \overset{*}{\rightharpoonup} \mu$) if

$$\int_{\mathbb{R}^n} f \, d\mu_k \to \int_{\mathbb{R}^n} f \, d\mu \quad \text{for all } f \in C_c(\mathbb{R}^n).$$

Evans calls this "weak convergence" but it is really weak*. Measures live in $(C_c)^*$, not in a space that is itself a dual. This terminology confusion is everywhere in the PDE and GMT literature. It used to trip me up. Now I just mentally substitute.

Theorem 5.4 (Weak* compactness)**.** *Let $\{\mu_k\}$ be Radon measures on $\mathbb{R}^n$ with $\sup_k \mu_k(K) < \infty$ for every compact $K \subseteq \mathbb{R}^n$. Then there exists a subsequence $\mu_{k_j} \overset{*}{\rightharpoonup} \mu$ for some Radon measure μ.*

Proof sketch. $C_c(\mathbb{R}^n)$ is separable. Let $\{f_m\}$ be a countable dense subset. For each m, the sequence $\{\int f_m \, d\mu_k\}$ is bounded by the uniform bound on compact sets plus compact support of f_m. Diagonalize: extract a subsequence so that $\int f_m \, d\mu_{k_j}$ converges for every m.

By density, $\int f \, d\mu_{k_j}$ converges for every $f \in C_c$. This limit defines a positive linear functional, so by Riesz (Theorem 5.1) it's integration against a Radon measure. □

Riesz is doing real work here. It converts the convergent functional back into a measure. Without it you have a limit that is a number for each test function but no guarantee that the limit is a measure.

Proposition 5.5 (Lower semicontinuity)**.** *If $\mu_k \overset{*}{\rightharpoonup} \mu$ and $U \subseteq \mathbb{R}^n$ is open, then $\mu(U) \leq \liminf_{k\to\infty} \mu_k(U)$.*

Proof. Approximate $\mathbf{1}_U$ from below by $f \in C_c$ with $0 \leq f \leq 1$, $\operatorname{spt}(f) \subseteq U$. Then $\int f \, d\mu = \lim \int f \, d\mu_k \leq \liminf \mu_k(U)$. Take sup over such f. □

One direction of the Portmanteau theorem. The other directions ($\limsup \mu_k(C) \leq \mu(C)$ for closed C, convergence of $\mu_k(A)$ when $\mu(\partial A) = 0$) follow by similar approximation.

Example 5.6. $\mu_k = \delta_{1/k}$ on $\mathbb{R}$. Then $\mu_k \overset{*}{\rightharpoonup} \delta_0$ since $f(1/k) \to f(0)$ for all $f \in C_c(\mathbb{R})$. But $\mu_k(\{0\}) = 0$ for all k while $\delta_0(\{0\}) = 1$. Weak convergence does not give convergence of measures of specific sets unless the boundary has measure zero.

Remark 5.7. In probability this compactness result is Prokhorov's theorem. The condition $\sup_k \mu_k(K) < \infty$ is related to tightness. For probability measures on $\mathbb{R}^n$, tightness and relative sequential compactness in the weak topology are equivalent.

That wraps Chapter 1. Lots of machinery. The covering theorems felt the hardest; the Riesz representation and weak compactness feel the most useful going forward, because Evans can freely differentiate measures, pass to weak limits, and decompose into absolutely continuous and singular parts without reproving anything.

TODO: go back and do more examples for the differentiation section. I understand Lebesgue–Besicovitch's statement but the proof still feels fuzzy. Maybe try it for $\mathbb{R}^1$ where Besicovitch is basically just Vitali.

6 September 16, 2024

Brian asked me over lunch what Hausdorff measure even *is*, and my attempt to explain it made me realize I did not understand the definition as well as I thought. Specifically, I could not say why the limit $\delta \to 0$ matters. Humbling.

Here is the problem. Lebesgue measure only "sees" n-dimensional things in $\mathbb{R}^n$. A curve in $\mathbb{R}^3$? Lebesgue measure says it has zero volume and walks away. Same for the surface of a sphere or a fractal. You need something that can handle non-integer dimensions. That is Hausdorff measure.

6.1 Hausdorff pre-measure and Hausdorff measure

We build $\mathcal{H}^s$ in two steps: approximate at scale δ, then take a limit.

Definition 6.1 (Hausdorff pre-measure)**.** Let $A \subseteq \mathbb{R}^n$, $0 \le s < \infty$, $0 < \delta \le \infty$. Define

$$\mathcal{H}^s_\delta(A) = \inf\left\{ \sum_{j=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_j}{2}\right)^s : A \subseteq \bigcup_{j=1}^{\infty} C_j,\ \operatorname{diam} C_j \le \delta \right\},$$

where $\alpha(s) = \pi^{s/2}/\Gamma(s/2+1)$ is the normalization constant chosen so that $\mathcal{H}^n$ agrees with $\mathcal{L}^n$.

I keep forgetting the normalization. Writing it out one more time: $\alpha(n) = \pi^{n/2}/\Gamma(n/2+1)$. For integer n this is the volume of the unit ball in $\mathbb{R}^n$. So $\alpha(1) = 2$, $\alpha(2) = \pi$, $\alpha(3) = 4\pi/3$. I have a flashcard for this now.

Definition 6.2 (s-dimensional Hausdorff measure)**.**

$$\mathcal{H}^s(A) = \lim_{\delta \to 0^+} \mathcal{H}^s_\delta(A) = \sup_{\delta > 0} \mathcal{H}^s_\delta(A).$$

The limit exists (possibly $= +\infty$) because $\mathcal{H}^s_\delta(A)$ is non-decreasing as $\delta \to 0$: shrinking δ restricts the admissible covers.

As δ shrinks, covers must use smaller sets, which can only make the infimum larger. Why force small covers at all? Because at large scales, covering sets can "cheat" by straddling geometric features that only become visible at finer resolution. But I am not sure I could make that precise without more work.

Remark 6.3. Some sanity checks:

- $\mathcal{H}^0$ is counting measure. Each nonempty covering set contributes $\alpha(0)(\operatorname{diam} C_j/2)^0 = 1$.
- $\mathcal{H}^1$ on $\mathbb{R}^1$ agrees with $\mathcal{L}^1$. Not hard but takes a small argument — intervals are the most efficient covers on the real line.
- $\mathcal{H}^n$ on $\mathbb{R}^n$ agrees with $\mathcal{L}^n$. That's the content of the isodiametric inequality below.
- $\mathcal{H}^s \equiv 0$ on $\mathbb{R}^n$ for $s > n$. Cover by cubes of side δ; each has diameter $\delta\sqrt{n}$. The sum goes to 0 as $\delta \to 0$ because the exponent $s > n$ beats the count.

$\mathcal{H}^s$ is a Borel regular outer measure on $\mathbb{R}^n$. Countable subadditivity is straightforward. Borel regularity takes more work (Evans, p. 62).

6.2 Hausdorff dimension

Lemma 6.4 (Dimension threshold)**.** *Let $A \subseteq \mathbb{R}^n$.*

1. *If $\mathcal{H}^s(A) < \infty$, then $\mathcal{H}^t(A) = 0$ for all $t > s$.*

2. *If $\mathcal{H}^t(A) > 0$, then $\mathcal{H}^s(A) = +\infty$ for all $0 \le s < t$.*

Proof sketch. For (i): let $\{C_j\}$ be a δ-cover of A. Then

$$\sum_j \alpha(t)\left(\frac{\operatorname{diam} C_j}{2}\right)^t \le \left(\frac{\delta}{2}\right)^{t-s} \frac{\alpha(t)}{\alpha(s)} \sum_j \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s.$$

Take the infimum: $\mathcal{H}^t_\delta(A) \le C \cdot \delta^{t-s} \cdot \mathcal{H}^s_\delta(A)$. Send $\delta \to 0$. The δ^{t-s} factor kills everything since $t - s > 0$ and $\mathcal{H}^s(A) < \infty$. Part (ii) is the contrapositive. □

So there is a critical value of s where $\mathcal{H}^s$ jumps from $+\infty$ to 0. A phase transition, if you want to be dramatic about it.

Definition 6.5 (Hausdorff dimension)**.**

$$\dim_{\mathcal{H}}(A) = \inf\big\{s \ge 0 : \mathcal{H}^s(A) = 0\big\} = \sup\big\{s \ge 0 : \mathcal{H}^s(A) = +\infty\big\}.$$

At the critical dimension itself, $\mathcal{H}^s(A)$ can be 0, ∞, or anything in between. The middle-thirds Cantor set has $\dim_{\mathcal{H}} = \log 2/\log 3$ and its Hausdorff measure at that dimension is 1. That is special. Most sets are not so well-behaved at the critical exponent.

Examples:

- Smooth k-dimensional submanifolds of $\mathbb{R}^n$: $\dim_{\mathcal{H}} = k$.
- Middle-thirds Cantor set: $\dim_{\mathcal{H}} = \log 2/\log 3 \approx 0.631$.
- Countable sets: $\dim_{\mathcal{H}} = 0$.
- $\mathbb{R}^n$: $\dim_{\mathcal{H}} = n$. Obviously.

I tried computing $\mathcal{H}^{\log 2/\log 3}$ of the Cantor set directly but got tangled in the covering argument. The upper bound is easy (use the 2^k remaining intervals at level k). The lower bound needs more care: you have to show those "natural" covers are actually optimal, and that is where I got stuck. I think the density argument in Rogers's book (Chapter 1, around Theorem 30) handles it, but I have not worked through it line by line.

TODO: work through the Cantor set dimension calculation carefully. Check Rogers's book for the density argument.

7 September 18, 2024

The isodiametric inequality via Steiner symmetrization is one of the cleanest proofs I have seen all year. I was sitting in the back row of KAP 414 with my laptop half-closed and I put it away to watch Evans write it out on the board.

7.1 Steiner symmetrization

Definition 7.1 (Steiner symmetrization)**.** Let $A \subseteq \mathbb{R}^n$ be Borel and $a \in S^{n-1}$ a unit vector. The *Steiner symmetrization* of A with respect to the hyperplane $\{x : x \cdot a = 0\}$ is obtained by replacing, for each line parallel to a, the intersection of that line with A by a symmetric interval of the same length centered on the hyperplane.

Precisely: decompose $x = y + ta$ where $y \perp a$, $t \in \mathbb{R}$. For each y, let $A_y = \{t : y + ta \in A\}$. Then

$$S_a(A) = \big\{\, y + ta : y \perp a,\ |t| \le \tfrac{1}{2}\mathcal{L}^1(A_y) \,\big\}.$$

Picture it: slide each vertical "slice" of A so it is centered on the hyperplane. Each slice keeps its length, so volume is preserved. But the set gets rounder. The diameter can only shrink or stay the same.

Lemma 7.2. *Let $A \subseteq \mathbb{R}^n$ be Borel.*

1. $\operatorname{diam} S_a(A) \le \operatorname{diam} A$.
2. $\mathcal{L}^n(S_a(A)) = \mathcal{L}^n(A)$.

Proof sketch. Part (ii) is immediate from Fubini: each slice A_y gets replaced by an interval of the same $\mathcal{L}^1$-measure. Integrate over y.

Part (i) is the interesting one. Take $y_1 + t_1 a$ and $y_2 + t_2 a$ in $S_a(A)$. We need their distance $\le \operatorname{diam} A$. By the symmetrization construction, there exist s_1, s_2 with $y_1 + s_1 a \in A$ and $y_2 + s_2 a \in A$, and you can choose them so that $|s_1 - s_2| \ge |t_1 - t_2|$ (because the original slices are at least as long as the symmetrized intervals, so you can pick points that are at least as far apart in the a-direction). Then:

$$|(y_1 + t_1 a) - (y_2 + t_2 a)|^2 = |y_1 - y_2|^2 + (t_1 - t_2)^2 \le |y_1 - y_2|^2 + (s_1 - s_2)^2 \le (\operatorname{diam} A)^2.$$

Centering the intervals can only bring points in the a-direction closer together. □

7.2 The isodiametric inequality

Theorem 7.3 (Isodiametric inequality)**.** *For any Borel set $A \subseteq \mathbb{R}^n$,*

$$\mathcal{L}^n(A) \le \alpha(n) \left(\frac{\operatorname{diam} A}{2} \right)^n .$$

Among all sets of a given diameter, the ball has the largest volume.

Proof. We can assume $\operatorname{diam} A < \infty$ (otherwise the right side is $+\infty$).

Step 1: Successive symmetrization. Let $e_1, \ldots, e_n$ be the standard basis. Set $A_0 = A$ and $A_k = S_{e_k}(A_{k-1})$ for $k = 1, \ldots, n$. By Lemma 7.2:

$$\mathcal{L}^n(A_n) = \mathcal{L}^n(A), \qquad \operatorname{diam} A_n \le \operatorname{diam} A.$$

And A_n is symmetric about the origin — symmetric with respect to each coordinate hyperplane.

Step 2. Since A_n is origin-symmetric, for every $x \in A_n$ we have $-x \in A_n$. So

$$2|x| = |x - (-x)| \le \operatorname{diam} A_n \le \operatorname{diam} A.$$

This gives $A_n \subseteq B(0, \operatorname{diam} A/2)$.

Step 3.

$$\mathcal{L}^n(A) = \mathcal{L}^n(A_n) \le \mathcal{L}^n\big(B(0, \operatorname{diam} A/2)\big) = \alpha(n) \left(\frac{\operatorname{diam} A}{2} \right)^n .$$ □

Symmetrize along each axis. Volume stays the same, diameter can only shrink, and the result is origin-symmetric, so it fits inside a ball of radius diam /2. Done. Three steps, each one line of real content. I love this proof.

Remark 7.4. The theorem extends to Lebesgue measurable sets and even arbitrary sets (via outer regularity: every set sits inside a Borel set of the same outer measure).

7.3 $\mathcal{H}^n = \mathcal{L}^n$

This is what all the machinery was building toward.

Theorem 7.5. *On $\mathbb{R}^n$, $\mathcal{H}^n = \mathcal{L}^n$.*

Proof. Two inequalities.

$\mathcal{H}^n \geq \mathcal{L}^n$: Let $\{C_j\}$ be any covering of A with $\operatorname{diam} C_j \leq \delta$. By the isodiametric inequality,

$$\mathcal{L}^n(A) \leq \sum_j \mathcal{L}^n(C_j) \leq \sum_j \alpha(n)\left(\frac{\operatorname{diam} C_j}{2}\right)^n.$$

Take the infimum over covers: $\mathcal{L}^n(A) \leq \mathcal{H}^n_\delta(A)$ for all δ, so $\mathcal{L}^n(A) \leq \mathcal{H}^n(A)$.

$\mathcal{H}^n \leq \mathcal{L}^n$: Cover A by cubes of side $\delta/\sqrt{n}$ (so diameter $\leq \delta$). The sum $\sum_j \alpha(n)(\operatorname{diam} C_j/2)^n$ is at most $\mathcal{L}^n(A) + \varepsilon$ if you choose the cover carefully using the definition of Lebesgue outer measure. Send $\varepsilon \to 0$: $\mathcal{H}^n_\delta(A) \leq \mathcal{L}^n(A)$, so $\mathcal{H}^n(A) \leq \mathcal{L}^n(A)$.

The isodiametric inequality does the heavy lifting in the first direction. The second direction is checking that the normalization constant works out. □

Brian and I spent a whole afternoon on why $\mathcal{H}^n = \mathcal{L}^n$ is not obvious. You might think: of course covering by balls and covering by arbitrary sets give the same answer. But they do not, at least not until you prove a geometric inequality bounding the volume of arbitrary sets by their diameter. That is the isodiametric inequality. Without it you are stuck. We sat in the courtyard outside Kaprielian Hall until it got dark, and I still wasn't sure I believed the second direction until I wrote it up later that night.

7.4 Densities

Definition 7.6 (Upper and lower s-dimensional densities)**.** Let $\mu = \mathcal{H}^s \llcorner A$. For $x \in \mathbb{R}^n$, define

$$\Theta^{*s}(\mu, x) = \limsup_{r\to 0^+} \frac{\mu(B(x,r))}{\alpha(s)\, r^s},$$
$$\Theta^{s}_{*}(\mu, x) = \liminf_{r\to 0^+} \frac{\mu(B(x,r))}{\alpha(s)\, r^s}.$$

When the limit exists, $\Theta^s(\mu, x) = \Theta^{*s}(\mu, x) = \Theta^s_*(\mu, x)$ is the *s-dimensional density* of μ at x.

Density measures how uniformly s-dimensional a set looks at small scales near x. If A is a smooth k-manifold, then $\Theta^k(\mathcal{H}^k \llcorner A, x) = 1$ at every $x \in A$. Locally the set looks like a flat k-plane, filling the ball at the same rate as $\mathbb{R}^k$.

Theorem 7.7 (Density bounds)**.** *Let $A \subseteq \mathbb{R}^n$ with $\mathcal{H}^s(A) < \infty$. Then:*

1. $\Theta^{*s}(\mathcal{H}^s \llcorner A,\, x) = 0$ *for $\mathcal{H}^s$-a.e. $x \notin A$.*
2. $2^{-s} \leq \Theta^{*s}(\mathcal{H}^s \llcorner A,\, x) \leq 1$ *for $\mathcal{H}^s$-a.e. $x \in A$.*

The upper bound $\Theta^{*s} \leq 1$ is the easier direction. A covering argument using the definition of $\mathcal{H}^s$. The lower bound $2^{-s} \leq \Theta^{*s}$ is harder and relies on a Vitali-type covering lemma for Hausdorff measure.

Remark 7.8. For integer $s = n$ everything is nicer. The n-dimensional density of $\mathcal{L}^n$-measurable sets equals 1 a.e. on the set and 0 a.e. off it. That is the Lebesgue density theorem. For non-integer s the density may not exist, and the gap between 2^{-s} and 1 is genuine. Sets with non-integer Hausdorff dimension do not have clean local structure.

I thought this gap was an artifact of the proof until I saw how it shows up in rectifiability in Chapter 5. It is real.

TODO: understand the proof of the lower bound 2^{-s} better. Evans uses a covering argument (pp. 72–73) that I need to go through line by line.

7.5 Connection to functions

The last bit connects Hausdorff measure back to the "fine properties" theme of the book. Main tool: Fubini applied to subgraphs.

Definition 7.9 (Subgraph)**.** For $f\colon \mathbb{R}^n \to [0,\infty]$ measurable, the *subgraph* is

$$\mathrm{SG}(f) = \{\, (x,t) \in \mathbb{R}^n \times \mathbb{R} : 0 \le t \le f(x) \,\} \subseteq \mathbb{R}^{n+1}.$$

Lemma 7.10. *If $f\colon \mathbb{R}^n \to [0,\infty]$ is $\mathcal{L}^n$-measurable, then $\mathrm{SG}(f)$ is $\mathcal{L}^{n+1}$-measurable and*

$$\mathcal{L}^{n+1}\big(\mathrm{SG}(f)\big) = \int_{\mathbb{R}^n} f \,\mathrm{d}\mathcal{L}^n.$$

Proof sketch. By Fubini: $\mathcal{L}^{n+1}(\mathrm{SG}(f)) = \int_{\mathbb{R}^n} \mathcal{L}^1(\{t : 0 \le t \le f(x)\})\,\mathrm{d}\mathcal{L}^n(x) = \int_{\mathbb{R}^n} f(x)\,\mathrm{d}\mathcal{L}^n(x)$. The measurability of $\mathrm{SG}(f)$ as a subset of $\mathbb{R}^{n+1}$ follows from approximation by simple functions. □

One of those lemmas that feels obvious but needs Fubini to make honest. The payoff comes later: you can study integrability, level sets, and regularity of a function by looking at its subgraph as a subset of $\mathbb{R}^{n+1}$ and applying Hausdorff measure tools. The Cavalieri principle ($\int f = \int_0^\infty \mathcal{L}^n(\{f > t\})\,\mathrm{d}t$) is a close relative.

That is Chapter 2. Hausdorff measure, Hausdorff dimension, the isodiametric inequality via Steiner symmetrization, the proof that $\mathcal{H}^n = \mathcal{L}^n$, densities, and the subgraph connection. The isodiametric proof is the gem. Everything else is careful bookkeeping with covers and limits.

TODO: read Evans Section 2.4 more carefully for measurability details. Revisit the Vitali covering argument in the density lower bound.

8 September 30, 2024

I slept through most of the weekend trying to catch up on everything. Hausdorff measure is still sitting in my head half-processed, but we are moving on. Area and coarea formulas. These are the real payoff of Chapters 1–2: the machinery we built starts producing actual geometric results.

8.1 Lipschitz Functions

Definition 8.1 (Lipschitz function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz** if there exists a constant $L \ge 0$ such that

$$|f(x) - f(y)| \le L|x - y| \qquad \text{for all } x, y \in \mathbb{R}^n.$$

The smallest such L is denoted $\mathrm{Lip}(f)$.

Why Lipschitz? Smooth maps are too restrictive for geometric measure theory, and continuous maps are too wild. Space-filling curves are continuous, for one thing. Lipschitz maps sit in between: differentiable a.e. by Rademacher, they send null sets to null sets, and they do not blow up Hausdorff measure by more than a controlled factor. The right class for most of what follows.

Theorem 8.2 (McShane Extension)**.** *Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}$ be Lipschitz with constant L. Then there exists $\bar{f} : \mathbb{R}^n \to \mathbb{R}$ with $\bar{f} = f$ on A and* $\operatorname{Lip}(\bar{f}) = \operatorname{Lip}(f)$*.*

The extension is explicit:

$$\bar{f}(x) = \inf_{a \in A} \big\{ f(a) + L|x - a| \big\}.$$

Check it satisfies the Lipschitz bound. Check it agrees with f on A. Done. That is the whole proof. It fits on an index card. Proofs like this are why I got into math instead of, I don't know, accounting.

For vector-valued $f : A \to \mathbb{R}^m$, apply McShane component by component. The Lipschitz constant picks up a factor of $\sqrt{m}$. Nobody cares about the exact constant in practice.

8.2 Rademacher's Theorem

This is one of those results that sounds impossible the first time.

Theorem 8.3 (Rademacher)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz. Then f is differentiable $\mathcal{L}^n$-a.e.*

Every locally Lipschitz function has a classical derivative at almost every point. No extra hypotheses. The proof is long. Each step is manageable on its own, but the orchestration is what I found hard. I tried to keep it all in my head while walking home from KAP and lost it by the time I got to Figueroa.

Proof sketch (Evans–Gariepy, pp. 81–84). It suffices to treat $f : \mathbb{R}^n \to \mathbb{R}$ (apply componentwise).

Step 1: Directional derivatives exist a.e. Fix a direction $v \in S^{n-1}$. For $\mathcal{L}^n$-a.e. x, the function $t \mapsto f(x + tv)$ is absolutely continuous on bounded intervals. This is basically Fubini: project $\mathbb{R}^n$ along v, and on almost every parallel line, f is absolutely continuous. So the derivative

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

exists for $\mathcal{L}^n$-a.e. x. Nothing deep here.

Step 2: A "gradient" exists a.e. Pick a countable dense set $\{v_k\} \subset S^{n-1}$. By Step 1 and a countable intersection, there's a full-measure set E where $D_{v_k} f(x)$ exists for all k. At a.e. point of E, the map $v \mapsto D_v f(x)$ turns out to be linear in v. Why? Because $D_{e_i} f$ exists a.e. for each coordinate direction e_i, and on lines we get

$$D_v f(x) = \sum_{i=1}^{n} v_i \, D_{e_i} f(x)$$

for the dense set of v's, and then for all v by continuity (the Lipschitz bound gives $|D_v f| \le \operatorname{Lip}(f)$, so we can pass to limits). Define $\nabla f(x) = (D_{e_1} f(x), \ldots, D_{e_n} f(x))$. Then $D_v f(x) = \nabla f(x) \cdot v$ for all v.

Step 3: Differentiability. We need

$$Q(x, v, t) := \frac{f(x + tv) - f(x) - \nabla f(x) \cdot (tv)}{t} \to 0$$

as $t \to 0$, *uniformly in* v. This is the hardest part. Define the "bad set"

$$A_\epsilon = \left\{x : \limsup_{t\to 0} \sup_{|v|=1} |Q(x, v, t)| > \epsilon\right\}.$$

We want $\mathcal{L}^n(A_\epsilon) = 0$. At points where ∇f exists and equals some linear map L, compare $f(x + tv) - f(x)$ against $L \cdot (tv)$ using the Lipschitz bound. A covering argument (Vitali, from a few lectures ago) shows the set where uniform convergence fails has measure zero. The uniform bound $|Q| \le 2\,\mathrm{Lip}(f)$ is what makes the covering argument work. □

Long proof. Worth it though.

Corollary 8.4. *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz and $f = 0$ on a set Z, then $Df = 0$ $\mathcal{L}^n$-a.e. on Z.*

Immediate: at a.e. point of Z where f is differentiable, the derivative has to vanish because f vanishes on Z and Z has full density at such points.

We also get a chain rule: if f and g are Lipschitz and the composition makes sense, $D(g \circ f) = Dg \circ Df$ a.e. Standard.

8.3 Linear Maps and Jacobians

Quick review of the linear algebra we need (Evans–Gariepy, pp. 87–90). Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be linear.

- The **adjoint** $L^* : \mathbb{R}^m \to \mathbb{R}^n$ satisfies $\langle Lx, y\rangle = \langle x, L^*y\rangle$. In matrices, $L^* = L^T$.
- $L^*L : \mathbb{R}^n \to \mathbb{R}^n$ is symmetric and positive semidefinite.
- **Polar decomposition**: $L = O \circ S$ where $S = (L^*L)^{1/2}$ is the unique positive semidefinite square root of L^*L, and $O : \mathbb{R}^n \to \mathbb{R}^m$ satisfies $|Ox| = |x|$ for $x \in \mathrm{range}(S)$. When $n \le m$ and L is injective, O extends to an orthogonal map on all of $\mathbb{R}^n$.

Definition 8.5 (Jacobian)**.** For a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$:

- If $n \le m$: $J_L = \sqrt{\det(L^*L)}$. This equals the product of singular values $\lambda_1 \cdots \lambda_n$ where λ_i^2 are eigenvalues of L^*L.
- If $n \ge m$: $J_L = \sqrt{\det(LL^*)}$.

For a Lipschitz map $f : \mathbb{R}^n \to \mathbb{R}^m$, the **Jacobian** at a point of differentiability is $J_f(x) = J_{Df(x)}$.

When $n = m$, both definitions agree and give $J_L = |\det L|$. Reassuring.

Lemma 8.6 (Change of variables for linear maps)**.** *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be linear with $n \le m$, and let $A \subset \mathbb{R}^n$ be Lebesgue measurable. Then*

$$\mathcal{H}^n(L(A)) = J_L \cdot \mathcal{L}^n(A).$$

Injective linear maps scale n-dimensional volume by exactly J_L. The area formula generalizes this to Lipschitz maps.

8.4 The Area Formula

This is the whole reason we defined Lipschitz maps and Jacobians. Everything before was setup.

Theorem 8.7 (Area Formula (Evans–Gariepy, Thm. 1, p. 96))**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$ *be Lipschitz, with* $n \le m$*. Then for any Lebesgue measurable* $g : \mathbb{R}^n \to [0, \infty]$,

$$\int_{\mathbb{R}^n} g(x)\, J_f(x)\, d\mathcal{L}^n(x) = \int_{\mathbb{R}^m} \left(\sum_{x \in f^{-1}(\{y\})} g(x) \right) d\mathcal{H}^n(y).$$

When $g = \chi_A$ for some measurable A, this becomes

$$\int_A J_f\, dx = \int_{\mathbb{R}^m} \#\big(A \cap f^{-1}(\{y\})\big)\, d\mathcal{H}^n(y),$$

which says the integral of the Jacobian over A equals the integral of the number of preimages, counted with Hausdorff measure. When f is injective on A, the right side collapses to $\mathcal{H}^n(f(A))$.

Proof sketch (pp. 96–101). Here is the architecture.

Step 1: Reduce to $g = \chi_A$**.** Standard approximation by simple functions.

Step 2: Handle $\{J_f = 0\}$**.** Where $J_f(x) = 0$, the derivative $Df(x)$ has rank $< n$, so f is crushing an n-dimensional region into something lower-dimensional. Need to show $\mathcal{H}^n(f(Z)) = 0$ for $Z = \{J_f = 0\}$. Cover Z with balls, estimate the Hausdorff measure of the image using the fact that f maps small balls into thin ellipsoids. Tedious but not conceptually hard.

Step 3: Linear case. Already done (Lemma 8.6).

Step 4: Lipschitz linearization. This is the core. By Rademacher, f is differentiable a.e. At points of differentiability, $f(y) \approx f(x) + Df(x)(y - x)$ for y near x. For any $\epsilon > 0$, decompose $\mathbb{R}^n = N \cup \bigcup_{k=1}^{\infty} E_k$ where $\mathcal{L}^n(N) = 0$ and on each E_k:

- Df is nearly constant: $\|Df(x) - Df(y)\| < \epsilon$ for $x, y \in E_k$,
- f is nearly affine: $|f(x) - f(y) - Df(z)(x - y)| < \epsilon |x - y|$ for $x, y, z \in E_k$.

On each E_k, f behaves like a linear map up to error ϵ, so the Jacobian formula holds up to $(1 + C\epsilon)$. Send $\epsilon \to 0$.

Step 5: Non-injectivity. The counting function $y \mapsto \#(A \cap f^{-1}(\{y\}))$ is measurable (this is not obvious; it follows from a structure theorem for Lipschitz maps). The multiplicity on the right accounts for non-injectivity. □

The decomposition in Step 4 goes by "Lipschitz linearization." It shows up everywhere in GMT. Lipschitz maps are well-approximated by their derivatives, and the error is controlled uniformly on compact subsets where the derivative varies little. I confused myself for an hour by trying to do Step 4 globally instead of on the pieces E_k. Do not do that. The global linearization fails because the derivative is only approximately constant, and the approximation error accumulates without the decomposition to contain it.

8.5 Surface area as an application

Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ a Lipschitz parametrization of a surface $\Sigma = f(U)$ with $n \leq m$. The area formula gives

$$\mathcal{H}^n(\Sigma) = \int_U J_f(x)\, dx,$$

provided f is injective (or counting multiplicity if not). For a graph $f(x) = (x, u(x))$ where $u : \mathbb{R}^n \to \mathbb{R}$, we get

$$J_f(x) = \sqrt{1 + |\nabla u(x)|^2},$$

which recovers the classical surface area formula from multivariable calculus. Six chapters of Evans machinery to justify something you learn in Math 226 as a given. Worth it? Ask me in December.

9 October 2, 2024

Midterm season. I have been living off coffee and the vending machine on the third floor of KAP. Today we finish Chapter 3 with the coarea formula.

9.1 The Coarea Formula

I found this harder to internalize than the area formula. Integrating over level sets instead of over the image makes the geometry less visual for me. I can draw the area formula on a napkin. The coarea formula I cannot.

The area formula handles $n \leq m$: a low-dimensional domain mapping into a higher-dimensional space. The coarea formula handles $n \geq m$: a high-dimensional domain mapping *down*. The level sets $f^{-1}(\{y\})$ are generically $(n-m)$-dimensional, and we integrate over them.

Theorem 9.1 (Coarea Formula (Evans–Gariepy, Thm. 2, p. 112))**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz, with $n \geq m$. Then for any Lebesgue measurable $g : \mathbb{R}^n \to [0, \infty]$,*

$$\int_{\mathbb{R}^n} g(x)\, J_f(x)\, d\mathcal{L}^n(x) = \int_{\mathbb{R}^m} \left(\int_{f^{-1}(\{y\})} g(x)\, d\mathcal{H}^{n-m}(x) \right) dy.$$

The outer integral on the right is over $\mathbb{R}^m$ with Lebesgue measure. The inner integral is over the level set $f^{-1}(\{y\})$ with $(n-m)$-dimensional Hausdorff measure.

9.2 Why this is a nonlinear Fubini theorem

Take $n = n_1 + n_2$, $m = n_2$, and let $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_2}$ be the projection $f(x_1, x_2) = x_2$. Then Df is the projection matrix, $J_f = 1$, and each level set $f^{-1}(\{y\})$ is $\mathbb{R}^{n_1} \times \{y\}$. The coarea formula becomes

$$\int_{\mathbb{R}^n} g(x)\, dx = \int_{\mathbb{R}^{n_2}} \left(\int_{\mathbb{R}^{n_1}} g(x_1, y)\, dx_1 \right) dy,$$

which is Fubini. So the coarea formula generalizes Fubini to nonlinear "projections." That actually is a nice way to think about it, and it is the viewpoint that finally made the formula click for me.

9.3 The case $m = 1$

The most common special case. Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz. Then $J_f = |\nabla f|$. The coarea formula becomes

$$\int_{\mathbb{R}^n} g(x)\,|\nabla f(x)|\,dx = \int_{-\infty}^{\infty} \left(\int_{f^{-1}(\{t\})} g\,d\mathcal{H}^{n-1} \right) dt.$$

Taking $g = \chi_A/|\nabla f|$ where $\nabla f \neq 0$:

$$\mathcal{L}^n(A) = \int_{-\infty}^{\infty} \int_{A\cap\{f=t\}} \frac{1}{|\nabla f|}\,d\mathcal{H}^{n-1}\,dt.$$

This "slices" the set A by level sets of f and reconstructs its volume. It shows up constantly in PDE and geometric analysis. We will need it in Chapter 5 for the BV coarea formula.

9.4 Proof sketch of the coarea formula

Proof sketch (Evans–Gariepy, pp. 112–117). Structurally similar to the area formula proof. Same architecture, different geometry.

Step 1: Linear case. For a surjective linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, verify that

$$J_L \cdot \mathcal{L}^n(A) = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap L^{-1}(\{y\}))\,dy$$

for all measurable A. The level sets of L are translates of $\ker L$, an $(n-m)$-dimensional subspace. So we are computing volume via "height times cross-section," weighted by the Jacobian. Use the polar decomposition and compute.

Step 2: Handle $\{J_f = 0\}$. Where $J_f(x) = 0$, the map $Df(x)$ is not surjective. Show that for a.e. $y \in \mathbb{R}^m$, the set $\{J_f = 0\} \cap f^{-1}(\{y\})$ has $\mathcal{H}^{n-m}$-measure zero. This uses Sard-type estimates: $f(\{J_f = 0\})$ has $\mathcal{L}^m$-measure zero. That is Sard's theorem for Lipschitz maps, which is nontrivial.

Step 3: Lipschitz linearization. Same trick as the area formula. Decompose $\mathbb{R}^n$ into pieces where Df is nearly constant, approximate by the linear formula on each piece, control errors, take limits.

Step 4: Measurability of the sliced integral. Verify that $y \mapsto \int_{f^{-1}(\{y\})} g\,d\mathcal{H}^{n-m}$ is measurable. Follows from the theory of rectifiable sets and the structure of Lipschitz level sets. □

The proof is harder than the area formula's, at least for me. The difficulty concentrates in Step 2: the critical set and how it interacts with level sets. Sard's theorem for smooth maps is standard differential topology (Milnor proves it in half a page for the C^∞ case). The Lipschitz version requires more care, and Evans spends about two pages on it.

Corollary 9.2. *Let $u : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz. Then for a.e. $t \in \mathbb{R}$, the level set $\{u = t\}$ has finite $\mathcal{H}^{n-1}$-measure, and*

$$\int_{-\infty}^{\infty} \mathcal{H}^{n-1}(\{u = t\} \cap A)\,dt \le \mathrm{Lip}(u) \cdot \mathcal{L}^n(A)$$

for every bounded measurable A.

This comes from the coarea formula with $g = \chi_A$ and $|\nabla u| \leq \mathrm{Lip}(u)$. It tells us "most" level sets of a Lipschitz function are $(n-1)$-dimensional surfaces with finite measure. The bad ones form a null set in the target. This connects to sets of finite perimeter in Chapter 5.

I still do not have sharp intuition for why the coarea formula involves the Jacobian of the gradient rather than the determinant of the full derivative. I think the Jacobian measures how much f "spreads" transverse to the level sets, and the component of Df along the level sets is irrelevant to the slicing. But that explanation is handwavy and I would not put it on an exam.

10 October 7, 2024

Rain all morning. The walk from my apartment on 30th Street to KAP was miserable but at least the topic is good. We are starting Sobolev spaces, which is where analysis starts having direct contact with PDE.

10.1 Weak Derivatives

Definition 10.1 (Weak derivative)**.** Let $U \subseteq \mathbb{R}^n$ be open. Suppose $u, v \in L^1_{\mathrm{loc}}(U)$. We say v is the **α-th weak partial derivative** of u, written $D^\alpha u = v$, provided

$$\int_U u\, D^\alpha \varphi\, dx = (-1)^{|\alpha|} \int_U v\, \varphi\, dx$$

for all test functions $\varphi \in C_c^\infty(U)$.

The idea: move all the derivatives onto the smooth test function via integration by parts, then *define* the derivative of u to be whatever v makes the identity work. If u happens to be classically differentiable, you recover the usual derivative. But now functions with corners and kinks are fair game. The absolute value function $|x|$ on $\mathbb{R}$ has weak derivative $\mathrm{sgn}(x)$, which is not a classical derivative at 0 but works perfectly in the weak sense.

Weak derivatives are unique up to a.e. equivalence. That is the fundamental lemma of the calculus of variations, nothing more.

10.2 Sobolev Spaces

Definition 10.2 (Sobolev space $W^{k,p}(U)$)**.** For $k \in \mathbb{N}$ and $1 \leq p \leq \infty$, the Sobolev space $W^{k,p}(U)$ consists of all $u \in L^p(U)$ such that for each multi-index α with $|\alpha| \leq k$, the weak derivative $D^\alpha u$ exists and belongs to $L^p(U)$.

The norm:

$$\|u\|_{W^{k,p}(U)} = \begin{cases} \left(\sum_{|\alpha| \leq k} \int_U |D^\alpha u|^p\, dx \right)^{1/p} & 1 \leq p < \infty, \\ \sum_{|\alpha| \leq k} \operatorname*{ess\,sup}_U |D^\alpha u| & p = \infty. \end{cases}$$

$W^{k,p}(U)$ is a Banach space. For $p = 2$ it is a Hilbert space, usually written $H^k(U)$.

$W_0^{k,p}(U)$ is the closure of $C_c^\infty(U)$ in $W^{k,p}(U)$. Think of these as Sobolev functions that "vanish on ∂U" in some appropriate sense. We will make that precise with traces later.

Product rules and chain rules work. If $u \in W^{1,p}(U)$ and $\Phi \in C^1$ with Φ' bounded, then $\Phi \circ u \in W^{1,p}(U)$ and $D(\Phi \circ u) = \Phi'(u) Du$.

One thing that confused me for a while: $W^{k,p}(U)$ is *not* just "C^k functions completed in the $W^{k,p}$ norm." That completion starting from C_c^∞ gives $W_0^{k,p}$, which is a different, smaller space. You need the weak derivative definition to get $W^{k,p}$ itself.

10.3 Approximation by Smooth Functions

The mollification argument is the template for everything in this chapter. Learn it once and you are set for all the approximation results.

Definition 10.3 (Mollifier)**.** Let $\eta \in C_c^\infty(\mathbb{R}^n)$ be the standard mollifier with $\eta \geq 0$, $\operatorname{spt}\eta \subseteq B(0,1)$, and $\int \eta = 1$. Set $\eta_\varepsilon(x) = \varepsilon^{-n}\eta(x/\varepsilon)$.

For $u \in L^1_{\text{loc}}(U)$, the mollification is $u^\varepsilon = \eta_\varepsilon * u$ on $U_\varepsilon = \{x \in U : \operatorname{dist}(x, \partial U) > \varepsilon\}$.

Theorem 10.4 (Local approximation)**.** *If $u \in W^{k,p}(U)$ with $1 \leq p < \infty$, then $u^\varepsilon \in C^\infty(U_\varepsilon)$ and $u^\varepsilon \to u$ in $W^{k,p}_{\text{loc}}(U)$ as $\varepsilon \to 0$.*

Proof sketch. The trick is showing $D^\alpha(\eta_\varepsilon * u) = \eta_\varepsilon * D^\alpha u$ for $|\alpha| \leq k$ on U_ε. Write out the convolution, differentiate under the integral (legal since η_ε is smooth and compactly supported), recognize the result as $(-1)^{|\alpha|} \int u\, D_y^\alpha \eta_\varepsilon(\cdot - y)\, dy$, then use the definition of weak derivative to flip the derivatives back. Once you have this identity, L^p convergence is the standard mollification result. □

Now the local-to-global step.

Theorem 10.5 (Meyers–Serrin)**.** *$C^\infty(U) \cap W^{k,p}(U)$ is dense in $W^{k,p}(U)$ for $1 \leq p < \infty$.*

The proof uses a partition of unity subordinate to an exhaustion of U. Mollify each piece with a small enough ε and sum. It is a careful $\varepsilon/2^k$ argument, not deep, but you have to track the supports. I messed up the support tracking the first time and got nonsense.

Can you approximate by $C^\infty(\overline{U})$ functions? In general, no.

Theorem 10.6 (Approximation up to the boundary)**.** *If ∂U is Lipschitz, then $C^\infty(\overline{U})$ is dense in $W^{k,p}(U)$ for $1 \leq p < \infty$.*

The Lipschitz condition lets you locally flatten the boundary and push the function slightly inward, where mollification is safe. Without it, you can build domains where the approximation fails. (Think of a domain with an inward cusp.)

Three approximation theorems, each stronger, each needing more hypotheses. I drew a diagram to keep them straight. Local → Meyers–Serrin → boundary.

11 October 9, 2024

Office hours ran long yesterday so I did not review traces before today. Walking in blind.

11.1 Traces

For smooth $u \in C^\infty(\overline{U})$, restricting to ∂U is trivial. For Sobolev functions you cannot just "evaluate on the boundary" because they are only defined a.e. But does the boundary even have positive measure? No. It has $\mathcal{L}^n$-measure zero. So restriction makes no sense without more work.

Theorem 11.1 (Trace theorem)**.** *Assume U is bounded and ∂U is Lipschitz. There exists a bounded linear operator*

$$T : W^{1,p}(U) \to L^p(\partial U)$$

such that $Tu = u|_{\partial U}$ for $u \in W^{1,p}(U) \cap C(\overline{U})$, and $\|Tu\|_{L^p(\partial U)} \le C\|u\|_{W^{1,p}(U)}$.

The trace characterizes $W_0^{1,p}(U)$: $u \in W^{1,p}(U)$ belongs to $W_0^{1,p}(U)$ if and only if $Tu = 0$ on ∂U. So $W_0^{1,p}$ really is the space of Sobolev functions vanishing at the boundary.

The sharper statement is that T maps into the fractional Sobolev space $W^{1-1/p,p}(\partial U)$. Evans and Gariepy do not go there. I looked it up in Adams and Fournier (Section 7.57) and decided I did not need it yet.

11.2 Extensions

Theorem 11.2 (Extension)**.** *Assume U is bounded and ∂U is Lipschitz. For each $1 \le p \le \infty$, there exists a bounded linear operator*

$$E : W^{1,p}(U) \to W^{1,p}(\mathbb{R}^n)$$

such that $Eu = u$ a.e. in U, $\operatorname{spt}(Eu)$ is bounded, and

$$\|Eu\|_{W^{1,p}(\mathbb{R}^n)} \le C\|u\|_{W^{1,p}(U)}.$$

Near the boundary, reflect the function across ∂U. Because ∂U is Lipschitz, you can locally flatten it, reflect, then patch with a partition of unity. Multiply by a cutoff for bounded support.

Extensions matter because lots of results are easier on all of $\mathbb{R}^n$ where there are no boundary issues. Prove it on $\mathbb{R}^n$, then transfer back via E. The Sobolev inequalities below are the prime example.

11.3 Gagliardo–Nirenberg–Sobolev ($p < n$)

This is the heart of the chapter. The notation is brutal.

The exponent $p^* = np/(n-p)$ looks arbitrary until you do dimensional analysis. $\|u\|_{L^q}$ has dimensions of $(\text{length})^{n/q}$ times pointwise size, and $\|Du\|_{L^p}$ has dimensions of $(\text{length})^{n/p-1}$ times pointwise size. For the inequality to be scale-invariant, we need $n/q = n/p - 1$, giving $q = np/(n-p) = p^*$. That is the only exponent that works. I wish Evans had explained this instead of just writing down p^* and moving on.

Theorem 11.3 (Gagliardo–Nirenberg–Sobolev)**.** *Assume $1 \le p < n$. There exists $C = C(p, n)$ such that*

$$\|u\|_{L^{p^*}(\mathbb{R}^n)} \le C\|Du\|_{L^p(\mathbb{R}^n)}$$

for all $u \in C_c^1(\mathbb{R}^n)$, where $p^* = \dfrac{np}{n-p}$.

The right-hand side only involves Du, not u itself. Because we are on all of $\mathbb{R}^n$ with compactly supported functions.

By density this extends to $W^{1,p}(\mathbb{R}^n)$, and via the extension operator to bounded domains with Lipschitz boundary:

$$\|u\|_{L^{p^*}(U)} \le C\|u\|_{W^{1,p}(U)}.$$

So $W^{1,p}(U) \hookrightarrow L^{p^*}(U)$ continuously. You trade one derivative for better integrability.

Corollary 11.4 (Poincaré inequality)**.** *If U is bounded and $1 \le p < n$, then for $u \in W_0^{1,p}(U)$:*

$$\|u\|_{L^q(U)} \le C\|Du\|_{L^p(U)} \quad \text{for all } 1 \le q \le p^*.$$

11.4 Morrey's Inequality ($p > n$)

Morrey's inequality is what makes $p > n$ special. You get continuity for free.

Theorem 11.5 (Morrey's inequality)**.** *Assume $n < p \leq \infty$. There exists $C = C(p, n)$ such that*

$$\|u\|_{C^{0,\gamma}(\mathbb{R}^n)} \leq C\|Du\|_{L^p(\mathbb{R}^n)}$$

for all $u \in C_c^1(\mathbb{R}^n)$, where $\gamma = 1 - n/p$.

So $W^{1,p}(U) \hookrightarrow C^{0,1-n/p}(\overline{U})$ for $p > n$. One derivative in L^p with p large enough, and the function *has* to be Hölder continuous. As $p \to \infty$, $\gamma \to 1$ and you approach Lipschitz. As $p \to n^+$, $\gamma \to 0$ and you barely get continuity.

The proof integrates $|u(x) - u(y)|$ along line segments, switches to polar coordinates, and applies Hölder's inequality. The key observation: the average of $|Du|$ over a ball controls the oscillation of u. I worked through this proof line by line and it is one of the most satisfying computations in the book.

11.5 The borderline $p = n$

$p = n$ is subtle. You do not get L^∞ (that would require $p > n$), and $p^* = np/(n - p)$ blows up. What you get is $W^{1,n}(U) \hookrightarrow L^q(U)$ for all $q < \infty$, but *not* $q = \infty$. Almost L^∞, but not quite. Trudinger's inequality gives an exponential integrability result, but Evans does not cover it.

11.6 Embedding summary

For U bounded with Lipschitz boundary:

Regime	**Embedding**	**Key feature**
$1 \leq p < n$	$W^{1,p}(U) \hookrightarrow L^{p^*}(U)$	Improved integrability
$p = n$	$W^{1,n}(U) \hookrightarrow L^q(U)$, all $q < \infty$	Almost L^∞
$p > n$	$W^{1,p}(U) \hookrightarrow C^{0,1-n/p}(\overline{U})$	Hölder continuity

The dimension n is the threshold. Everything turns on whether p is below, at, or above n.

12 October 14, 2024

I think I finally understand the Sobolev embeddings. Three passes through the proofs did it. My handwritten notes from the first pass are unreadable garbage; Brian's are worse. Today: compactness, capacity, and the fine properties of Sobolev functions.

12.1 Rellich–Kondrachov Compactness

Arzelà–Ascoli for Sobolev spaces.

Theorem 12.1 (Rellich–Kondrachov)**.** *Assume $U \subset \mathbb{R}^n$ is bounded and open with Lipschitz boundary. Then:*

1. *If $1 \leq p < n$: $W^{1,p}(U) \hookrightarrow\hookrightarrow L^q(U)$ for each $1 \leq q < p^*$.*

2. *If* $p = n$*:* $W^{1,p}(U) \hookrightarrow\hookrightarrow L^q(U)$ *for each* $1 \le q < \infty$*.*

3. *If* $p > n$*:* $W^{1,p}(U) \hookrightarrow\hookrightarrow C^{0,\beta}(\overline{U})$ *for each* $0 < \beta < 1 - n/p$*.*

Here $\hookrightarrow\hookrightarrow$ *means compact embedding: bounded sequences have convergent subsequences.*

Notice the strict inequalities. You get compactness into L^q for $q < p^*$, but *not* at the critical exponent $q = p^*$. The embedding $W^{1,p} \hookrightarrow L^{p^*}$ is continuous but not compact. This matters enormously in the calculus of variations. If you have ever heard someone say "loss of compactness at the critical exponent," this is the theorem behind it.

Proof idea for case (1). Take a bounded sequence (u_k) in $W^{1,p}(U)$. Extend to $\mathbb{R}^n$, then mollify. For fixed $\varepsilon > 0$, the mollified sequence (u_k^ε) is bounded in C^1 on compact sets (from the $W^{1,p}$ bound and properties of convolution), so Arzelà–Ascoli gives a convergent subsequence. Diagonal argument over $\varepsilon \to 0$. The $q < p^*$ restriction enters because you need spare integrability to interpolate between L^1 convergence of mollifications and the L^{p^*} bound. □

I will not pretend I understand every detail of the interpolation step. The mechanism is: Sobolev regularity gives equicontinuity after mollification, Arzelà–Ascoli gives convergence in C^0, interpolation converts that to L^q convergence for subcritical q. The interpolation inequality used is Hölder's applied to $|u_k - u_j|^q = |u_k - u_j|^\theta \cdot |u_k - u_j|^{q-\theta}$ with the right splitting.

12.2 Capacity

Capacity confused me for weeks. Here is what I eventually worked out: sets of capacity zero are the sets that Sobolev functions cannot "see." Measure zero sets are invisible to L^p functions. Capacity zero sets are invisible to $W^{1,p}$ functions. Capacity zero is a *finer* notion. There are measure-zero sets with positive capacity.

Definition 12.2 (p-capacity)**.** For $1 \le p < n$ and $A \subseteq \mathbb{R}^n$,

$$\operatorname{cap}_p(A) = \inf \left\{ \int_{\mathbb{R}^n} |Du|^p \, dx : u \in W^{1,p}(\mathbb{R}^n),\ u \ge 1 \text{ on a neighborhood of } A \right\}.$$

Properties:

1. Monotone: $A \subseteq B \implies \operatorname{cap}_p(A) \le \operatorname{cap}_p(B)$.
2. Countably subadditive: $\operatorname{cap}_p\big(\bigcup_k A_k\big) \le \sum_k \operatorname{cap}_p(A_k)$.
3. If $\operatorname{cap}_p(A) = 0$, then $\mathcal{H}^s(A) = 0$ for all $s > n - p$.
4. $\operatorname{cap}_p(A) = 0 \implies \mathcal{L}^n(A) = 0$, but not conversely.

Property (3) is the one that matters most in practice. It tells you the dimension of the exceptional set where Sobolev functions can misbehave. For $p = 2$ in $\mathbb{R}^3$, the exceptional set has Hausdorff dimension at most $3 - 2 = 1$, so it can be at worst a curve.

12.3 Quasicontinuity and Precise Representatives

A Sobolev function is only defined a.e., so talking about pointwise values is dangerous. But you can choose a representative that is well-behaved outside an arbitrarily small set in the capacity sense.

Definition 12.3 (Quasicontinuous)**.** A function $u : U \to \mathbb{R}$ is p**-quasicontinuous** if for every $\varepsilon > 0$ there exists an open set V with $\text{cap}_p(V) < \varepsilon$ such that $u|_{U\setminus V}$ is continuous.

Theorem 12.4 (Quasicontinuous representative)**.** *Every $u \in W^{1,p}(\mathbb{R}^n)$ $(1 \le p \le n)$ has a p-quasicontinuous representative $\tilde{u}$, unique up to sets of* cap_p*-zero.*

Sobolev functions are "almost continuous." The discontinuity set can be made as small as you want in the capacity sense. Not continuous. Not merely measurable. Quasicontinuous. That is the correct regularity notion for these functions.

The **precise representative** is defined pointwise:

$$u^*(x) = \lim_{r\to 0} ⨍_{B(x,r)} u\, dy,$$

whenever this limit exists. For $u \in W^{1,p}$, the limit u^* exists and equals $\tilde{u}$ outside a set of cap_p-zero.

This connects back to the differentiation-of-measures machinery from the Vitali covering lectures. Same idea: recover pointwise information from integral averages. But the exceptional set is measured by capacity instead of Hausdorff measure.

12.4 Differentiability on Lines

An equivalent way to think about Sobolev functions that avoids integration by parts entirely.

Theorem 12.5. *If $u \in W^{1,p}(U)$ for $1 \le p \le \infty$, then for each $i = 1, \dots, n$, the function $t \mapsto u(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n)$ is absolutely continuous on the relevant line segment for $\mathcal{L}^{n-1}$-a.e. choice of the remaining variables. Its classical derivative equals the weak derivative $D_i u$ a.e.*

The converse is also true: if $u \in L^p(U)$ is absolutely continuous on a.e. line in each coordinate direction, and the classical partials are in L^p, then $u \in W^{1,p}(U)$.

So Sobolev functions are exactly the L^p functions that are absolutely continuous on almost every axis-parallel line. Brian asked me why you can not just use this as the definition. You could. But the integration-by-parts definition plays better with PDE applications, and it generalizes to fractional and negative-order Sobolev spaces where ACL makes no sense. Still, the ACL characterization is the most concrete way to picture what a Sobolev function looks like.

13 October 21, 2024

Midterm grades came back. Could have been worse. Today we start Chapter 5: BV functions and sets of finite perimeter. Evans pp. 166–226 is one of the hardest stretches in the book, and I can feel it already from the density of the notation alone.

The whole point of BV is to handle jumps that $W^{1,1}$ cannot. Take the characteristic function of a ball. It has a perfectly good "derivative" in the distributional sense: a surface measure on the boundary. But $\chi_E \notin W^{1,1}$ because that derivative is not an L^1 function. BV says: fine, let the derivative be a measure. That is the entire motivation. I'm drinking lukewarm coffee from the thermos I forgot to wash out yesterday, and it tastes about as good as you'd expect.

13.1 Definitions and first properties

Let $U \subseteq \mathbb{R}^n$ be open.

Definition 13.1 (Variation)**.** For $f \in L^1(U)$, define the *variation* of f in U by

$$\|Df\|(U) \;=\; \sup\left\{ \int_U f \,\operatorname{div} \varphi \, dx \;\middle|\; \varphi \in C_c^1(U;\mathbb{R}^n),\; |\varphi| \le 1 \right\}.$$

Say $f \in BV(U)$ if $\|Df\|(U) < \infty$.

Test f against all smooth compactly supported divergence fields of magnitude at most 1 and take the sup. If finite, you are in BV. The sup is the total variation of Df, which by the Riesz representation theorem (the one from Chapter 1, Theorem 5.1) is an $\mathbb{R}^n$-valued Radon measure.

The norm on $BV(U)$ is $\|f\|_{BV} = \|f\|_{L^1} + \|Df\|(U)$.

Remark 13.2. $W^{1,1}(U) \subset BV(U)$. If $f \in W^{1,1}$, then $Df = \nabla f \, dx$ with $\nabla f \in L^1$, and $\|Df\|(U) = \int_U |\nabla f| \, dx < \infty$. But BV is strictly bigger: χ_E for E with smooth boundary is in BV but not $W^{1,1}$, because its derivative is a surface measure on ∂E.

Example 13.3. Some examples worth keeping in your head:

1. $f = \chi_E$ where E is a ball in $\mathbb{R}^n$. Then $f \in BV(\mathbb{R}^n)$ and $\|Df\|(\mathbb{R}^n) = \mathcal{H}^{n-1}(\partial E)$. The derivative is $(-\nu_E)\mathcal{H}^{n-1}\lfloor \partial E$ where ν_E is the outward unit normal.

2. The Cantor function on $[0,1]$: in BV but not $W^{1,1}$. Its derivative is entirely singular, concentrated on the Cantor set. No jump part either. All Cantor part.

3. Any monotone function $f : \mathbb{R} \to \mathbb{R}$ is BV_{loc}. Connects to the classical theory of functions of bounded variation on the real line.

13.2 The structure theorem

This makes everything else in the chapter go.

Theorem 13.4 (Structure Theorem for BV)**.** *Let $f \in BV(U)$. Then Df is an $\mathbb{R}^n$-valued Radon measure on U, admitting the decomposition*

$$Df \;=\; D^a f + D^s f,$$

where:

- *$D^a f$ is absolutely continuous with respect to $\mathcal{L}^n$. By Radon–Nikodym, $D^a f = \nabla f \, \mathcal{L}^n$ for some $\nabla f \in L^1(U;\mathbb{R}^n)$.*

- *$D^s f$ is singular with respect to $\mathcal{L}^n$: concentrated on a set of Lebesgue measure zero.*

This is Lebesgue–Radon–Nikodym decomposition applied to the measure Df. Nothing deep yet, but naming the pieces is what lets you talk precisely about the structure.

For χ_E, the absolutely continuous part vanishes and the singular part is the surface measure on ∂E. For a smooth function, it is all absolutely continuous. For the devil's staircase, it is all singular.

$D^s f$ decomposes further into $D^j f + D^c f$: a *jump part* concentrated on a countably $(n-1)$-rectifiable set and a *Cantor part* that is singular but diffuse. We will need this in the pointwise properties section later.

13.3 Approximation and compactness

Two big results. First, lower semicontinuity.

Theorem 13.5 (Lower semicontinuity)**.** *If $f_j \to f$ in $L^1(U)$, then*

$$\|Df\|(U) \;\le\; \liminf_{j\to\infty} \|Df_j\|(U).$$

Why? $\|Df\|(U)$ is a supremum of continuous linear functionals of f. Each functional passes to the limit, and the sup of limits is $\le$ the liminf of sups. I had to sit down and write it out to believe it. Standard argument, but I keep getting the quantifiers backward until I write it slowly.

Theorem 13.6 (Approximation by smooth functions)**.** *If $f \in BV(U)$, there exist $f_j \in C^\infty(U) \cap BV(U)$ with*

$$f_j \to f \text{ in } L^1(U) \quad \text{and} \quad \|Df_j\|(U) \to \|Df\|(U).$$

Here is the catch. This is convergence in the *strict* topology: L^1 convergence of functions plus convergence of total variation. You do NOT get $f_j \to f$ in the BV norm. That would require $\|Df_j - Df\|(U) \to 0$, which is impossible in general. Why? BV functions can have singular parts and smooth functions cannot. The mollification $f * \rho_\varepsilon$ converges in L^1 and the variation converges, but $D(f * \rho_\varepsilon) - Df$ does not go to zero as a measure. I confused this with norm convergence for two weeks.

BV is not the closure of C^∞ in the BV norm. That closure is $W^{1,1}$.

Theorem 13.7 (Compactness)**.** *Let $\{f_j\}$ be a sequence in $BV(U)$ with $\sup_j \|f_j\|_{BV(U)} < \infty$, where U is bounded with Lipschitz boundary. Then a subsequence converges in $L^1(U)$ to some $f \in BV(U)$.*

In Sobolev spaces you would get weak convergence by reflexivity, but $W^{1,1}$ is not reflexive and neither is BV. The compactness is only in L^1. Combined with lower semicontinuity, it is enough for the direct method in the calculus of variations. That combination (compactness plus l.s.c.) shows up so often that I should probably tattoo it on my arm.

13.4 Traces and extensions

Theorem 13.8 (Trace for BV)**.** *Let $U \subset \mathbb{R}^n$ be bounded open with Lipschitz boundary. There's a bounded linear operator*

$$T : BV(U) \to L^1(\partial U; \mathcal{H}^{n-1})$$

with $Tf = f|_{\partial U}$ for $f \in BV(U) \cap C(\overline{U})$.

The trace only lands in $L^1(\partial U)$. No fractional Sobolev regularity. Makes sense: BV functions can jump, so the trace just records the boundary values. The proof uses the same flattening machinery as Sobolev traces but needs the BV-specific approximation results from above.

Theorem 13.9 (Extension)**.** *Let $U \subset \mathbb{R}^n$ be bounded with Lipschitz boundary. There's a bounded linear operator $E : BV(U) \to BV(\mathbb{R}^n)$ with $Ef = f$ a.e. in U, $\operatorname{spt}(Ef)$ compact, and $\|Ef\|_{BV(\mathbb{R}^n)} \le C\|f\|_{BV(U)}$.*

Mirrors the Sobolev case. The Lipschitz boundary condition is doing real work.

14 October 23, 2024

Today: coarea formula for BV and isoperimetric inequality. I think the coarea formula is the single best result in the chapter. Brian said the same thing last week, independently, which either means we are both right or we have been talking to each other too much.

14.1 Coarea formula for BV

Theorem 14.1 (Coarea formula)**.** *Let $f \in BV(U)$. For $\mathcal{L}^1$-a.e. $t \in \mathbb{R}$, the level set $\{f > t\}$ has finite perimeter in U. And for every Borel set $A \subseteq U$,*

$$\|Df\|(A) \;=\; \int_{-\infty}^{\infty} \|\partial\{f > t\}\|(A)\, dt,$$

where $\|\partial\{f > t\}\|$ is the perimeter measure of the superlevel set $\{x \in U : f(x) > t\}$.

Let me unpack this. The left side is the total variation of Df on A. The right side slices this into contributions from each level: at height t, the superlevel set $\{f > t\}$ contributes its perimeter restricted to A. Integrate over all heights and you recover the total variation.

For $f \in W^{1,1}(U)$, the formula becomes

$$\int_A |\nabla f|\, dx \;=\; \int_{-\infty}^{\infty} \mathcal{H}^{n-1}(A \cap \partial^*\{f > t\})\, dt,$$

which generalizes the classical coarea formula from Chapter 3.

Remark 14.2. The formula also gives a converse: $f \in L^1(U)$ is in $BV(U)$ if and only if $\int_{-\infty}^{\infty} \|\partial\{f > t\}\|(U)\, dt < \infty$. Sometimes that is the easiest way to check BV membership. I used it on a problem set for a radial function and it was far cleaner than the definition.

The proof confused me on first read because Evans reduces to smooth f first, proves the smooth case by the classical coarea formula, then passes to the limit using lower semicontinuity. The $\leq$ and $\geq$ directions are proved separately. The $\geq$ direction is harder.

14.2 Isoperimetric inequalities

Among all sets of given volume, the ball has the smallest perimeter. Known since antiquity, which makes you wonder what the ancient Greeks would have done with distributional derivatives.

Theorem 14.3 (Isoperimetric inequality)**.** *Let $E \subset \mathbb{R}^n$ have finite perimeter. Then*

$$\|\partial E\|(\mathbb{R}^n) \;\geq\; n\, \alpha(n)^{1/n} \left(\mathcal{L}^n(E)\right)^{(n-1)/n},$$

where $\alpha(n) = \mathcal{L}^n(B(0,1))$. Equality iff E is a ball (up to measure zero).

The argument uses the coarea formula. First establish the BV Sobolev inequality: for $f \in BV(\mathbb{R}^n)$,

$$\|f\|_{L^{n/(n-1)}(\mathbb{R}^n)} \;\leq\; \frac{1}{n\, \alpha(n)^{1/n}}\, \|Df\|(\mathbb{R}^n).$$

Then plug in $f = \chi_E$. The left side gives $\mathcal{L}^n(E)^{(n-1)/n}$ and the right gives $\|\partial E\|(\mathbb{R}^n)$ times the constant.

Actually, that is only the inequality. The equality case requires a symmetrization argument, which is a different beast entirely. But the inequality itself follows from the coarea formula in about half a page.

There is also a *relative isoperimetric inequality*: if U is bounded Lipschitz and $\mathcal{L}^n(E \cap U) \leq \frac{1}{2}\mathcal{L}^n(U)$, then

$$\mathcal{L}^n(E \cap U)^{(n-1)/n} \;\leq\; C(U)\;\|\partial E\|\,(U).$$

Useful for compactness arguments.

15 October 28, 2024

This chapter is harder than Chapter 4. I am sure of it now. Today we get into the geometric measure theory proper: reduced boundary, blow-up, and the generalized divergence theorem. I sat in the front row for this one.

15.1 Sets of finite perimeter and the reduced boundary

For a set E of finite perimeter, the "boundary" in the BV sense is not the topological boundary ∂E (which can be wild). It is a well-behaved subset called the *reduced boundary* $\partial^* E$.

Definition 15.1 (Reduced boundary)**.** Let E have finite perimeter in $\mathbb{R}^n$. Write $\mu_E = D\chi_E$ for the Gauss–Green measure and $\|\mu_E\| = \|\partial E\|$ for its total variation. A point $x \in \mathbb{R}^n$ belongs to the *reduced boundary* $\partial^* E$ if:

1. $\|\mu_E\|\,(B(x,r)) > 0$ for all $r > 0$.
2. The limit $\nu_E(x) = \lim_{r\to 0} \frac{\mu_E(B(x,r))}{\|\mu_E\|(B(x,r))}$ exists.
3. $|\nu_E(x)| = 1$.

The vector $\nu_E(x)$ is called the *measure-theoretic outer unit normal*.

$\partial^* E$ consists of points where the boundary "looks like a hyperplane" at small scales. At each such point there is a meaningful notion of which way is out. The reduced boundary sits inside the topological boundary ∂E, and the two can differ wildly. Modify E on a null set and the topological boundary can change beyond recognition, but $\partial^* E$ does not move.

15.2 The blow-up theorem

This is the hard theorem of the chapter. I do not have a clean one-paragraph summary for it.

Theorem 15.2 (Blow-up at the reduced boundary)**.** *Let E have finite perimeter, $x \in \partial^* E$ with outer normal $\nu_E(x)$. Define*

$$E_{x,r} \;=\; \frac{E - x}{r}.$$

Then $\chi_{E_{x,r}} \to \chi_{H^-(\nu_E(x))}$ in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ as $r \to 0^+$, where $H^-(\nu) = \{y : y \cdot \nu < 0\}$.

Zoom in at a point of $\partial^* E$ and the set converges to a half-space. The boundary looks flat at infinitesimal scales. The normal to the half-space is $\nu_E(x)$ from the definition. The proof uses Besicovitch differentiation, the definition of the reduced boundary, and BV compactness. Evans pp. 190–200. I went through it once and got the main ideas. Reproducing it from scratch would take me a full afternoon.

Remark 15.3. A major consequence: $\partial^* E$ is countably $(n-1)$-rectifiable, and the perimeter measure satisfies $\|\partial E\| = \mathcal{H}^{n-1} \lfloor \partial^* E$. The perimeter measure is $(n-1)$-dimensional Hausdorff measure restricted to the reduced boundary.

15.3 Measure-theoretic boundary and Gauss–Green

Definition 15.4 (Measure-theoretic boundary)**.** The *measure-theoretic boundary* $\partial_* E$ consists of all $x \in \mathbb{R}^n$ where neither E nor its complement has density 1:

$$\partial_* E \; = \; \left\{ x : \limsup_{r \to 0} \frac{\mathcal{L}^n(E \cap B(x,r))}{\alpha(n) r^n} > 0 \text{ and } \limsup_{r \to 0} \frac{\mathcal{L}^n(B(x,r) \setminus E)}{\alpha(n) r^n} > 0 \right\}.$$

We always have $\partial^* E \subseteq \partial_* E \subseteq \partial E$. For sets of finite perimeter, $\mathcal{H}^{n-1}(\partial_* E \setminus \partial^* E) = 0$, so the reduced and measure-theoretic boundaries agree up to $\mathcal{H}^{n-1}$-null sets.

Now the payoff.

Theorem 15.5 (Generalized Gauss–Green theorem)**.** *Let $E \subset \mathbb{R}^n$ have finite perimeter, $\varphi \in C^1_c(\mathbb{R}^n; \mathbb{R}^n)$. Then*

$$\int_E \operatorname{div} \varphi \, dx \; = \; \int_{\partial^* E} \varphi \cdot \nu_E \, d\mathcal{H}^{n-1}.$$

The classical divergence theorem needs E to have smooth or at least Lipschitz boundary. This version only needs finite perimeter. The boundary integral is over $\partial^* E$ with the measure-theoretic normal. The cleanest way to see why: $D\chi_E = -\nu_E \mathcal{H}^{n-1} \lfloor \partial^* E$ by the structure theory, and the formula is just the definition of distributional derivative applied to χ_E. Once you have the blow-up theorem and rectifiability, this falls out.

15.4 Pointwise properties of BV functions

What can you say about individual points? More than I expected.

Theorem 15.6 (Approximate limits and jump set)**.** *Let $f \in BV(U)$. Then:*

1. *For $\mathcal{H}^{n-1}$-a.e. $x \in U$, f has an approximate limit $\tilde{f}(x) = \operatorname{ap\,lim}_{y \to x} f(y)$ (the precise representative).*

2. *The jump set J_f is countably $(n-1)$-rectifiable.*

3. *$D^s f = D^j f + D^c f$, where $D^j f = (f^+ - f^-)\nu_f \, \mathcal{H}^{n-1} \lfloor J_f$ is the jump part and $D^c f$ is the Cantor part.*

So the full decomposition is:

$$Df \; = \; \underbrace{\nabla f \, \mathcal{L}^n}_{\text{abs. cont.}} \; + \; \underbrace{(f^+ - f^-)\nu_f \, \mathcal{H}^{n-1} \lfloor J_f}_{\text{jump}} \; + \; \underbrace{D^c f}_{\text{Cantor}} \, .$$

Three parts, each living on a different kind of set. The ac part is spread over U. The jump part sits on the $(n-1)$-dimensional jump set. And the Cantor part lives on something stranger, like the Cantor set itself. Three flavors of derivative, each with its own geometry. I spent an evening staring at this decomposition and thinking about which real-valued BV functions on $\mathbb{R}$ realize each part. Monotone + continuous + singular = all three at once, but constructing an explicit example in $\mathbb{R}^n$ for $n \geq 2$ seems harder.

For χ_E with E having finite perimeter: $\nabla \chi_E = 0$, $D^c \chi_E = 0$, and $D^j \chi_E = -\nu_E \, \mathcal{H}^{n-1} \lfloor \partial^* E$. All jump.

15.5 Essential variation on lines

Theorem 15.7 (Characterization via sections)**.** *$f \in L^1(U)$ is in $BV(U)$ if and only if for each coordinate direction e_i, the one-dimensional slices $t \mapsto f(x' + te_i)$ have finite essential variation for $\mathcal{L}^{n-1}$-a.e. x', and the total of these variations is finite.*

BV analogue of the ACL characterization for Sobolev functions. For Sobolev, the slices are absolutely continuous; for BV, the slices have bounded essential variation. The parallel is clean.

15.6 Criterion for finite perimeter

Theorem 15.8 (Criterion)**.** *$E \subset \mathbb{R}^n$ measurable has finite perimeter in U iff*

$$\sup\left\{ \int_E \operatorname{div} \varphi \, dx \;\middle|\; \varphi \in C_c^1(U; \mathbb{R}^n),\ |\varphi| \le 1 \right\} < \infty,$$

which is $\chi_E \in BV(U)$.

Sets of finite perimeter include: sets with Lipschitz boundary, convex sets, sublevel sets $\{f < t\}$ of BV functions for a.e. t (by the coarea formula). Sets that do NOT have finite perimeter: fractal boundaries with infinite $\mathcal{H}^{n-1}$-measure, like the Koch snowflake in $\mathbb{R}^2$. The Koch snowflake has finite area but its boundary has infinite length, so it fails the criterion.

16 November 4, 2024

I'm tired. Four chapters of measure theory will do that. But this chapter is where it pays off: Sobolev functions, BV functions, convex functions, all the rough objects from earlier, turn out to be differentiable in various precise senses almost everywhere. And then the punchline: Sobolev functions are "almost C^1." I'm writing these notes at 11pm with a bag of Trader Joe's peanut butter pretzels and a highlighter I keep clicking open and shut.

16.1 L^p differentiability

Instead of asking that f be differentiable pointwise, you ask that the Taylor remainder is small *on average* over shrinking balls.

Definition 16.1 (L^p differentiability)**.** Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $1 \le p \le \infty$. We say f is L^p *differentiable* at x if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$\left(\frac{1}{r^n} \int_{B(x,r)} |f(y) - f(x) - L \cdot (y - x)|^p \, \mathrm{d}y \right)^{1/p} = o(r) \quad \text{as } r \to 0.$$

When this holds, L is the L^p derivative of f at x.

Note the scaling: the left side has units of f, and $o(r)$ means the remainder decays faster than linear. Classical differentiability is the L^∞ case. L^p differentiability is weaker for finite p, and how much weaker depends on p in a way that connects directly to the Sobolev embedding exponents.

Theorem 16.2 (L^{p^*} differentiability of $W^{1,p}$ functions)**.** *If $f \in W^{1,p}_{\text{loc}}(\mathbb{R}^n)$ with $1 \le p < n$, then f is L^{p^*} differentiable at $\mathcal{L}^n$-a.e. x, where $p^* = np/(n-p)$. The L^{p^*} derivative equals $Df(x)$.*

The proof uses the Morrey-type estimate from Chapter 4. Write f as its average plus an integral of the gradient, apply Poincaré on balls, and pin down the base point via Lebesgue differentiation. The exponent p^* is sharp. You do not get L^q for $q > p^*$.

16.2 Approximate differentiability

This one is strange on first contact. The approximate derivative says: f looks differentiable if you are willing to ignore a set of density zero at the point.

Definition 16.3 (Approximate differentiability)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$.

1. The *approximate limit* $\ell = \operatorname{ap}\lim_{y\to x} f(y)$ exists if for every $\varepsilon > 0$,

$$\lim_{r\to 0} \frac{\mathcal{L}^n(\{y \in B(x,r) : |f(y) - \ell| \geq \varepsilon\})}{\alpha(n)r^n} = 0.$$

   The bad set has density zero at x.

2. f is *approximately differentiable* at x if there's a linear $L : \mathbb{R}^n \to \mathbb{R}$ with

$$\operatorname{ap}\lim_{y\to x} \frac{f(y) - f(x) - L \cdot (y - x)}{|y - x|} = 0.$$

   Write $\operatorname{ap} Df(x) = L$.

You replace the ordinary limit with an approximate limit. The set where the difference quotient is large could be nonempty, even uncountable. You just need it to be thin in a density sense. Is this a real derivative? Sort of. It agrees with the classical one when the classical one exists, and it is unique, but it does not satisfy the chain rule in general. Actually, wait. It does satisfy the chain rule for Lipschitz outer functions. I confused myself on this point for a day.

Theorem 16.4. *If f is approximately differentiable at x, then $\operatorname{ap} Df(x)$ is unique.*

Proof. Suppose L_1, L_2 both work. For each $\varepsilon > 0$, the sets $A_i = \{y : |f(y)-f(x)-L_i(y-x)|/|y-x| \geq \varepsilon\}$ each have density zero at x. Their union also has density zero. On $B(x,r) \setminus (A_1 \cup A_2)$ we get $|(L_1 - L_2)(y - x)| \leq 2\varepsilon|y - x|$, and this set has density one. Pick directions and send $\varepsilon \to 0$. □

Now the big result for BV.

Theorem 16.5 (BV functions are approximately differentiable a.e.)**.** *Let $f \in BV_{\text{loc}}(\mathbb{R}^n)$. Then f is approximately differentiable at $\mathcal{L}^n$-a.e. x, with*

$$\operatorname{ap} Df(x) = \nabla f(x) \quad \mathcal{L}^n\text{-a.e.},$$

where ∇f is the Radon–Nikodým derivative of Df with respect to $\mathcal{L}^n$.

Surprisingly strong. BV functions can jump across hypersurfaces, but away from those jumps they are approximately differentiable with the "expected" gradient. The jumps form a set of $\mathcal{L}^n$-measure zero (they live on $(n-1)$-rectifiable sets) and hence density zero at $\mathcal{L}^n$-a.e. point. The approximate derivative just ignores them.

Remark 16.6. The hierarchy, strongest to weakest:

$$\text{classically diff.} \implies L^\infty \text{ diff.} \implies L^p \text{ diff.} \implies \text{approx. diff.}$$

None of the arrows reverse.

17 November 6, 2024

Short lecture today. We do the $p > n$ case, Rademacher redux, and convex functions.

17.1 Differentiability a.e. for $W^{1,p}$, $p > n$

When $p > n$ you get classical differentiability. This is where the Sobolev embedding from Chapter 4 really pays off.

Theorem 17.1. *Let $f \in W^{1,p}_{\text{loc}}(\mathbb{R}^n)$ with $n < p \leq \infty$. Then f is differentiable at $\mathcal{L}^n$-a.e. x, with derivative $Df(x)$.*

Proof. By Morrey's inequality, for $y \in B(x,r)$:

$$|f(y) - f(x) - Df(x)\cdot(y-x)| \leq Cr\left(\frac{1}{r^n}\int_{B(x,r)} |Df(y) - Df(x)|^p \,\mathrm{d}y\right)^{1/p}.$$

Lebesgue differentiation applied to $|Df - Df(x)|^p$ gives the right side as $o(r)$ for a.e. x. That's classical differentiability. □

Short. Morrey does the work.

Theorem 17.2 (Rademacher's Theorem — second proof)**.** *Every locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable $\mathcal{L}^n$-a.e.*

Proof. Locally Lipschitz implies $f \in W^{1,\infty}_{\text{loc}}(\mathbb{R}^n)$. Since $\infty > n$, apply Theorem 17.1. □

Two lines. The first proof, back in Chapter 3, was this long argument with directional derivatives, Fubini, countable dense sets, and covering arguments. This proof hides all that pain inside the Chapter 4 prerequisites. Fair trade.

17.2 Convex functions

Theorem 17.3. *Every convex function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz.*

Convexity is a global geometric condition but it forces local analytic regularity. A convex function cannot oscillate wildly: supporting hyperplanes box it in. Combined with Rademacher, every convex function is differentiable a.e. But Aleksandrov goes one derivative further.

17.3 Aleksandrov's theorem

A convex function has second derivatives almost everywhere. No smoothness assumed. I found this hard to believe the first time I read it.

Theorem 17.4 (Aleksandrov)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. For $\mathcal{L}^n$-a.e. x, there exists a symmetric $D^2 f(x) \in \mathbb{R}^{n\times n}$ with*

$$\lim_{r\to 0}\frac{1}{r^n}\int_{B(x,r)} \frac{\left|f(y) - f(x) - Df(x)\cdot(y-x) - \frac{1}{2}(y-x)^T D^2 f(x)(y-x)\right|}{r^2}\,\mathrm{d}y = 0.$$

And $D^2 f(x) \geq 0$ a.e.

Think about what this says. The second-order Taylor remainder, averaged over a ball, is $o(r^2)$. Convex functions have a second-order Taylor expansion a.e., and the Hessian is nonneg. You would *expect* nonneg from convexity, sure. But proving the Hessian exists without any smoothness assumption is the hard part.

Proof sketch. The argument takes all of Evans Section 6.4. Three steps:

1. f is locally Lipschitz by Theorem 17.3, so differentiable a.e. by Rademacher. The gradient Df is BV (the distributional Hessian is a nonneg. matrix-valued Radon measure).

2. Decompose $D^2 f$ into ac and singular parts using BV structure.

3. The Taylor expansion holds at Lebesgue points of the ac part, which is a.e.

The jump from first to second derivatives is the deepest result in the chapter. □

Consider $f(x) = |x|$ on $\mathbb{R}$. It has $f''(x) = 0$ for $x \neq 0$ and is not twice differentiable at $x = 0$. But $\{0\}$ is a single point. Aleksandrov says this is always the situation for convex functions: the bad set has measure zero. The example is trivial, but the theorem is not.

17.4 Whitney's extension theorem

Theorem 17.5 (Whitney extension). *Let $C \subseteq \mathbb{R}^n$ be closed. Suppose $f : C \to \mathbb{R}$ and $d : C \to \mathbb{R}^n$ satisfy the compatibility condition: for every $\varepsilon > 0$ and compact $K \subseteq \mathbb{R}^n$, there exists $\delta > 0$ such that*

$$|f(y) - f(x) - d(x) \cdot (y - x)| \leq \varepsilon\, |y - x|$$

for $x, y \in C \cap K$ with $|x - y| < \delta$.

Then there exists $F \in C^1(\mathbb{R}^n)$ with $F|_C = f$ and $DF|_C = d$.

I will not go through the proof. It is a partition-of-unity construction with Whitney cubes, and the bookkeeping is long. The point is the statement: if your data (f, d) on C is *consistent* with being a function and its derivative, then a C^1 extension exists.

17.5 Approximation by C^1 functions

Here is the punchline of the entire book.

Theorem 17.6 (C^1 approximation). *Let $f \in W^{1,p}(\mathbb{R}^n)$, $1 \leq p \leq \infty$. For every $\varepsilon > 0$, there's a $g \in C^1(\mathbb{R}^n)$ with*

$$\mathcal{L}^n(\{x : f(x) \neq g(x) \text{ or } Df(x) \neq Dg(x)\}) < \varepsilon.$$

Read that again. It does not say f is close to g in some norm. It says f literally *equals* a C^1 function except on a set of measure less than ε. The function and its gradient agree with a smooth function outside a tiny exceptional set. The first time I read this theorem I thought I was misreading it.

Proof sketch. Three ingredients:

1. **Approximate differentiability a.e.** By Theorem 16.2 or Theorem 16.5, f is approximately differentiable a.e. with approximate derivative equal to Df.

2. **Lusin-type extraction.** Egorov's theorem plus the definition of approximate differentiability gives a closed set C with $\mathcal{L}^n(\mathbb{R}^n \setminus C) < \varepsilon$ such that $f|_C$ and $Df|_C$ satisfy Whitney's compatibility condition.

3. **Whitney extension.** Apply Theorem 17.5 to $(f|_C, Df|_C)$ to get $g \in C^1(\mathbb{R}^n)$ with $g = f$ and $Dg = Df$ on C.

The exceptional set is $\mathbb{R}^n \setminus C$, which has measure $< \varepsilon$. □

Lusin's theorem for Sobolev functions. Lusin says a measurable function is continuous outside a small set. This says a Sobolev function is C^1 outside a small set. The analogy is exact, and the proof structure mirrors Lusin's: extract a good set, check compatibility, extend.

Remark 17.7. This is NOT mollification. Mollification gives $f_\varepsilon \in C^\infty$ with $f_\varepsilon \to f$ in $W^{1,p}$, but $f_\varepsilon \neq f$ everywhere. The C^1 approximation theorem gives pointwise equality on a big set. Much stronger, but you only get C^1, not C^∞.

Why does this matter? It gives an intrinsic characterization of Sobolev spaces: $W^{1,p}$ functions are exactly the L^p functions that agree with C^1 functions outside sets of arbitrarily small measure, with the right gradient control. In GMT it lets you transfer smooth results to Sobolev functions. In PDE it sometimes lets you reduce regularity arguments to the C^1 case.

Everything in Chapter 6 depends on the analysis from Chapters 3–5. The Sobolev embeddings, the BV theory, the fine properties of measures. This is where those investments pay off. The C^1 approximation theorem is, I think, the best result in the book. Six chapters of machinery, and this is what falls out at the end. I walked home from KAP at midnight after writing up the proof sketch and felt like the semester finally made sense.

Acknowledgments. Thanks to Brian Fan for working through most of this with me.