Math 212a - Real Analysis

Taught by Yum-Tong Siu
Notes by Dongryul Kim

Fall 2017

This course was taught by Yum-Tong Siu. The class met on Tuesdays and Thursdays at 2:30–4pm, and the textbooks used were *Real analysis: measure theory, integration, and Hilbert spaces* by Stein and Shakarchi, and *Partial differential equations* by Evans. There were weekly problem sets, a timed take-home midterm, and a take-home final. There were 6 students enrolled, and detailed lecture notes and problem sets might be found on the course website.

Contents

1 August 31, 2017

We are going to use two books: Stein–Shakaarchi, *Real Analysis* and Lawrence Evans, *Partial Differential Equations*.

The motivation studying real analysis is solving differential equations. Fourier in 1822 wrote a book about the solving the heat equation. His idea was to convert the differential equation to algebraic equations. Solving differential equations is finding eigenfunctions for differential operators. For instance, e^{rx} is an eigenfunctions for d/dx with eigenvalue r. So if we let

$$f(x) = \sum_j c_j e^{r_j x},$$

then $df/dx = \sum_j c_j r_j e^{r_j x}$ and so this becomes an algebraic equation in the coefficients. A special case is when $f(x)$ is on $\mathbb{R}$ with period 2π. Then the eigenfunctions are e^{inx} with $n \in \mathbb{Z}$.

But one needs to argue that there is a unique representation $f(x) = \sum_{n\in\mathbb{Z}} c_n e^{inx}$. In case of a finite dimensional vector space, with an inner product, this is easy. You can easily choose an orthonormal basis of $\mathbb{C}^n$, and it is easy to check that

$$\vec{v} = \sum_{j=1}^{n} c_j \vec{e}_j, \text{ where } c_j = \langle \vec{v}, \vec{e}_j \rangle.$$

People tried to do a similar thing. It is easy to check that $e^{inx}/\sqrt{2\pi}$ are orthonormal with respect to the inner product

$$(f, g) = \int_{x=0}^{2\pi} f(x)\overline{g(x)} dx.$$

So it would be reasonable that

$$\gamma_n = (f, e^{inx}/\sqrt{2\pi}) = \int_{x=0}^{2\pi} f(x) \frac{e^{inx}}{\sqrt{2\pi}} dx.$$

Now the question is whether

$$f(x) = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} (f, e^{inx}) e^{inx}$$

is true. Lebesgue answered this question in 1902, using Lebesgue theory. Actually this didn't solve the problem entirely even in constant coefficients, and it remained unsolved until 1955 when Malgrange and Ehrenpreis solve it in constant coefficients.

1.1 Convergence of the Fourier series

We need to start with replacing the function f by $(c_n)_{n\in\mathbb{Z}}$. Consider the Nth partial sum

$$s_N = \sum_{n=-N}^{N} c_n e^{inx}.$$

We want to know if $\lim_{N\to\infty} s_N = f$. It can be computed that

$$\begin{aligned} s_n(x) &= \sum_{k=-n}^{n} c_k e^{ikx} = \sum_{k=-n}^{n} \frac{1}{2\pi}\int_{y=0}^{2\pi} f(y)e^{-iky}dy e^{inx} \\ &= \frac{1}{2\pi}\int_{y=0}^{2\pi} f(y)\biggl(\sum_{k=-n}^{n} e^{ik(x-y)}\biggr)dy. \end{aligned}$$

This sum in the integral is called the **Dirichlet–Dini kernel**. We can sum them as

$$D_n(x) = \sum_{k=-n}^{n} e^{ikx} = \frac{e^{-inx} - e^{i(n+1)x}}{1-e^{ix}} = \frac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x}.$$

Then we can write

$$s_n(x) = \frac{1}{2\pi}\int_0^{2\pi} f(y)D_n(x-y)dy,$$

as the convolution.

To argue about the limit of $s_n(x)$ as $n\to\infty$, we need to say about the commutation of limits. This is not very true of most functions, for instance characteristic functions of bad sets. Lebesgue had this big idea of partitioning the range instead of the domain. The base of the rectangles may be strange, though.

If we change the period from $[0,2\pi]$ to $[-L,L]$, we can rescale the length and use $e^{inx\pi/L}$ instead. In the limiting case $L\to\infty$, we will get another theory. Then

$$\hat{f}(\xi) = \int_{x=-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx, \quad f(x) = \int_{\xi=-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i\xi x}d\xi.$$

These are called the Fourier transform and the inversion formula.

This still doesn't solve the differential equations with constant coefficients. Let me tell you why. Suppose you have a differential equation like

$$\Bigl(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\Bigr)f = g.$$

If you take the Fourier transform, we get

$$-4\pi^2(\xi_1^2 + \cdots + \xi_n^2)\hat{f} = \hat{g}$$

and then divide by the polynomial. But this is a problem because there might be zeros in the polynomial. This was a big problem, and it took a long time to handle these.

2 September 5, 2017

So we start out with the details. We want to solve differential equations by eigenfunctions, but there are two problems here. The first is the convergence of the series like $f = \sum_{n=-\infty}^{\infty} c_n e^{inx}$. The second is the effect of differentiation. But here, it is easier to do integration, because it will be easier to exchange sums. So the tool we need here is the fundamental theorem of calculus. For Riemann integration, we need "uniformity", but this is too strong and so we will use Lebesgue theory. We are going to study the size of the set when a certain property holds or fails.

2.1 Measure of a set

We start with the reals $\mathbb{R}$. For a set $E \subseteq \mathbb{R}$, we are going to study this.

Theorem 2.1 (Structre theorem)**.** *Let O be an open subset of $\mathbb{R}$. Then O is the disjoint union of a countable number of open intervals, possibly of infinite length.*

Proof. The key is that a point disconnects an open set in $\mathbb{R}$. Then any connected component of O is an open interval. This is because, if $x \in O$, then

$$A_x = \{y \in \mathbb{R} : y > x \text{ and } [x, y] \subseteq O\}$$

is of the form (x, b). If you do the same thing on the other side, the connected component of x in O is (a, b). Countability follows from the fact that the rational numbers is dense and countable. □

This means that we can write $O = \bigcup_{k\in J}(a_k, b_k)$ as a disjoint union. This allows us to define its measure as

$$m(O) = \sum_{k\in J}(b_k - a_k).$$

Now given arbitrary set $E \subseteq \mathbb{R}$, Lebesgue tried to approximate E from the outside by open set, and from the inside by closed sets. But we are going to do something different.

Definition 2.2. Given a set $E \subseteq \mathbb{R}$, we define its **outer measure** as

$$m_*(E) = \inf_{O \supseteq E \text{ open}} m(O).$$

Lebesgue's original definition for the inner measure was also like $m_i(E) = \sup_{F\subseteq E} m(F)$ and call E measurable if the outer and inner measures coincide. But we are not doing this. Instead, we define

Definition 2.3. Given $E \subseteq \mathbb{R}$, E is **measurable** if for each $\epsilon > 0$, there exists an open set $O_\epsilon \supseteq E$ such that $m_*(O_\epsilon - E) < \epsilon$.

Why are these equivalent? Intuitively, we are trying to approximate $O_\epsilon - E$ from the outside, i.e., taking a neighborhood G of $O_\epsilon - E$. Here we can replace G with $G \cap O_\epsilon$. Then this is like approximating E from the inside by $F = O_\epsilon - G$, which is closed in E.

There are three important properties of the collection of all measurable sets. This is the abstract formulation.

(0) The empty set ϕ and the universe ($\mathbb{R}$) are both measurable.

(1) If A and B are measurable, then $A - B$ is measurable.

(2) Countable unions of measurable sets are measurable.

(3) Countable intersections of measurable sets are measurable.

If X is a set and a collection of subset of X have these three properties, then it is called a **σ-algebra**.

Proposition 2.4. *Countable unions of measurable sets are measurable.*

Proof. Let E_n be the collection. For each n, we approximate $E_n \subseteq O_{\epsilon,n}$ so that $m_*(O_{\epsilon,n} - E_n) < 2^{-n}\epsilon$. Now $E \subseteq O_\epsilon = \bigcup_n O_{\epsilon,n}$ and the excess would be

$$O_\epsilon - E \subseteq \bigcup_n (O_{\epsilon,n} - E_n),$$

and by countable sub-additivity of the outer measure, the right hand side has outer measure at most ϵ. □

Proving (1) is more complicated, because outer approximation and inner approximation get mixed up.

Proposition 2.5 (Additivity of outer measure)**.** *If $E_1, E_2 \subseteq \mathbb{R}$ have positive distance between them, i.e., $\inf_{x_j \in E_j} |x_1 - x_2| > 0$, then $m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2)$.*

Proof. Let δ be the distance. Let $G_j = \{y \in \mathbb{R} : \operatorname{dist}(y, E_j) < \delta/2\}$. This is an open set by the triangle inequality. Also G_1 and G_2 are disjoint. Given $O_\epsilon \supseteq E_1 \cup E_2$, we can assume that $O_\epsilon \subseteq G_1 \cup G_2$ by replacing with the intersection. Let $O_{\epsilon,j} = G_j \cap O_\epsilon$ so that O_ϵ is the disjoint union of $O_{\epsilon,1}$ and $O_{\epsilon,2}$. Then

$$m_*(E_1) + m_*(E_2) \leq m(O_{\epsilon,1}) + m(O_{\epsilon,1}) = m(O_\epsilon) < m_*(E_1 \cup E_2) + \epsilon.$$

Since ϵ is arbitrary, we get the inequality on the other side. □

Proposition 2.6. *Every closed subset $F \subseteq \mathbb{R}$ is measurable.*

Proof. By countable union, we may assume that F is bounded; $F = \bigcup_n F \cap [-n, n]$. Assume that $F \subseteq (a, b)$. Then we can again we can write

$$F = \bigcup_{j \in J} (a_j, b_j).$$

Now we can approximate (a_j, b_j) from the inside by closed sets, and then take the complement. Here, we actually need to select a finite number of big intervals. □

Proposition 2.7. *If E is measurable, then $E^c = \mathbb{R} - E$ is measurable.*

Proof. By definition, there are open sets $O_n \supseteq E$ such that $m_*(O_n - E) < n^{-1}$. If we let $S = \bigcup_n O_n^c$, then $E^c - S \subseteq O_n - E$ for all n, and so $m_*(E^c - S) = 0$. This implies that $E^c - S$ is measurable. On the other hand, S is a countable union of closed sets and so measurable. Hence E^c is also measurable. $\square$

This and (2) implies (3), and then (1) follows. Thus measurable sets form an σ-algebra.

2.2 More properties of the measure

Proposition 2.8 (Additivity of measure)**.** *If E_j are measurable and disjoint and $E = \bigcup_n E_n$, then*

$$m(E) = \sum_j m(E_j).$$

Proof. First reduce to the case when each E_j is bounded, by breaking them up into $E_j \cap [n, n+1)$. For the other side of the inequality, we approximate E_j for the inside like $F_j \subseteq E_j$ so that $m(E_j - F_j) < 2^{-j}\epsilon$. Here the point is that any two disjoint compact sets have positive distance. So we can use the previous additivity property to get additivity for F_j. $\square$

Proposition 2.9. *The measure of a limit of an increasing sequence is the limit of the measure.*

Proof. You just use the fact that $E_n = (E_n - E_{n-1}) \cup (E_{n-1} - E_{n-2}) \cup \cdots$. $\square$

Proposition 2.10. *The measure of a limit of a decreasing sequence is the limit of the measure, if the measure of at least one is finite.*

This $m(E_1) < \infty$ assumption is crucial. Here is a counterexample. If $E_n = (n, \infty)$, then $E = \emptyset$ and $m(E_n) = \infty$.

3 September 7, 2017

Let me make a remark about Borel measurable sets. Abstract measure theory is just taking some sets and taking countable unions, intersections, complements. Borel measurable sets are the sets that can be obtained by these operations from open sets. There are more Lebesgue measurable sets, namely the sets with measure zero.

Let me give you two counter-examples. We are going to construct a non-measurable set $E \subseteq [0,1]$ such that a countable disjoint union of E contains $[0,1]$ but is contained in $[-1,2]$. Consider the equivalence relation $x \sim y$ if and only if $x - y \in \mathbb{Q}$ on $[0,1]$, and pick one representative for each class. Then this set E has the property.

The second example is the Cantor function, which does not satisfy the fundamental theorem of calculus.

3.1 Integration

Let's look at the definition of Riemann integration. Riemann's idea was to cut the interval $[a,b]$ into smaller intervals, and take the limit as the mesh goes to zero. Here, we require that f is continuous, or almost continuous. We then approximate the function with the step function

$$\sum_j a_j \chi_{I_j}.$$

The definition depends on the fact that this step function approaches f uniformly. Then by definition,

$$\sum_j a_j m(I_j) = \int \sum_j a_j \chi_{I_j} \to \int f.$$

In Lebesgue's case, we assume that $0 \leq f \leq M$ and cut the target into pieces $0 = y_0 < y_1 < \cdots < y_p = M$. Let

$$\{x : y_{j-1} \leq f(x) < y_j\} = E_j.$$

Then you conclude that

$$\sum y_j \chi_{E_j} \to f$$

pointwise as the mesh goes to 0. Such a function is called a simple function. Then we define

$$\sum_j y_j m(E_j) = \int \sum y_j \chi_{E_j} \to \int f.$$

3.2 Egorov's theorem

Theorem 3.1 (Egorov, almost uniform convergence)**.** *Let E be measurable and $m(E) < \infty$. Let f_k be a sequence of measurable functions supported on E, and $f_k \to f$ almost everywhere on E. Then for every $\epsilon > 0$ there exists a closed subset $A_\epsilon \subseteq \mathbb{R}$ with $A_\epsilon \subseteq E$ such that $m(E - A_\epsilon) < \epsilon$ and $f_k \to f$ uniformly on A_ϵ.*

Definition 3.2. f is **measurable** if $\{f < c\}$ is measurable as a set for all $c \in \mathbb{R}$.

Proof. Uniform convergence $f_k \to f$ means that there exists an N such that $|f_k - f| < n^{-1}$ for all $k \geq N$. So we look at where this fails.

For n, N, define

$$E_{n,N} = \{x \in E : |f_k(x) - f(x)| < n^{-1} \text{ for all } k \geq N\}.$$

In words, this is the set on which f_k is uniformly n^{-1}-close to f for $k \geq N$. By definition $E_{n,N} \subseteq E_{n,N+1} \nearrow E$, because $f_k \to f$ pointwise. So given n, there exists a N_n such that

$$m(E - E_{n,N_n}) < 2^{-n}.$$

So f_k is n^{-1}-closed to f for $k \geq N_n$ outside a set of measure 2^{-n}.

Given $\epsilon > 0$, there exists a ℓ_ϵ such that

$$\hat{A}_\epsilon = \bigcap_{n \geq \ell_\epsilon} E_{n,N_n}$$

has $m(E - \hat{A}_\epsilon) < \epsilon/2$. This is the "good set", but we also have the condition that this has to be closed. Approximate $\hat{A}_\epsilon$ by a "good" closed subset $A_\epsilon \subseteq \hat{A}_\epsilon$ such that $m(\hat{A}_\epsilon - A_\epsilon) < \epsilon/2$. $\square$

Corollary 3.3. *If φ_k is a sequence of simple functions, with common finite-measure support, and $\varphi_k \to f$ pointwise, then $\varphi_k \to f$ almost uniformly.*

Corollary 3.4. *If φ_k is further uniformly bounded, then $\int \varphi_k$ is Cauchy.*

So we can define Lebesgue integration.

3.3 Convergence theorems

There are three convergence theorems and Fatou's lemma, all of which follow from Egorov's theorem.

Lemma 3.5 (Fatou)**.** *Let $f_k \geq 0$ be a sequence of functions on $\mathbb{R}$. Then*

$$\liminf_{k\to\infty} \int f_k \geq \int \liminf_{k\to\infty} f_k.$$

All the other theorems regard with the question of whether $\lim_{k\to\infty}\int f_k = \int f$ if $f_k \to f$.

Before I go on, I want to make some comments on the definition of Lebesgue integration. So far, we have defined the integral of $0 \le f \le M$ on E with $m(E) < \infty$. Now we relax the conditions, one by one. Suppose $f \ge 0$ but that it could be unbounded. What we can do is to truncate f at n so that it becomes $\max(f, n)$. The integral of this is well-defined and we can then take the limit.

Next, we would like to remove the condition $m(E) < \infty$. Here we truncate the domain and replace E with $E \cap [-n, n]$. But we still need the $f \ge 0$ condition, because otherwise we could get $\infty - \infty$.

Definition 3.6. We say that f is **integrable** if $\int f_+ = \int \max\{0, f\}$ is finite and $\int f_- = \int \max\{0, -f\}$ is also finite. In this case, we have $f = f_+ - f_-$ and we define

$$\int_f = \int f_+ - \int f_-.$$

The dominated convergence theorem concerns with the case $|f_k| \le g$ where g is integrable. Here we need absolute continuity, i.e., $\int_A g \to 0$ as $m(A) \to 0$.

Proof of Fatou's lemma. The proof is actually quite simple. Let $f = \liminf_k f_k$ and $\varphi_n = \inf_{k\ge n} f_k$ so that $\varphi_n \nearrow f$. Here we can use monotone convergence. On the other hand, $\varphi_n \le f_n$ implies $\int \varphi_n \le \int f_n$. Taking $\liminf$ of both sides gives the result immediately.

If things are not finite, you can truncate the functions and then take the limit. □

I'd also like to compare Riemann integration and Lebesgue integration. It turns out that f is Riemann integrable if and only if f is continuous outside a set of measure zero.

4 September 12, 2017

To define Lebesgue integration, we first looked at how to measure sets. Then we approximated the function with simple functions. We needed Egorov's theorem to give sense to it. Then for unbounded functions with non-compact support, we had to truncate it.

We were comparing Riemann integration with Lebesgue integration. We would want to characterize functions which are Riemann integrable.

Theorem 4.1. *If f is bounded on $[a, b]$ (with $a < b$), then f is Riemann integrable if and only if f is continuous outside a set of measure zero.*

Proof. ($\Rightarrow$) Let ψ_P and φ_P be the step functions on the partition P which gives the upper sum and the lower sum. If f is Riemann integrable, then there is a sequence of partitions such that

$$\varphi_{P_\nu} \leq \varphi_{P_{\nu+1}} \leq \cdots \leq f \leq \cdots \leq \psi_{P_{\nu+1}} \leq \psi_{P_\nu}.$$

Here, if we let $\varphi = \lim_{\nu\to\infty} \varphi_{P_\nu}$ and $\psi = \lim_{\nu\to\infty} \psi_{P_\nu}$, then

$$\int \varphi = \lim_{\nu\to\infty} \int \varphi_{P_\nu} = \int f = \int \psi,$$

in the Lebesgue sense. Since $\psi \geq \varphi$, we have $\psi = \varphi$ outside E, a measure zero set.

We claim that f is continuous outside E plus the countable number of partition points. This is because at any of the point outside E and not a partition point, the function f is approximated from above and below.

($\Leftarrow$) The function is Riemann integrable if it is continuous. Also, f is bounded, so if the non-continuous points have measure zero, then the upper sum and lower sum will have a small difference. □

4.1 Differentiability of nondecreasing dunctions

There are two forms of the fundamental theorem of calculus: you first differentiate, or you first differentiate.

$$\frac{d}{dx}\int_a^x f(t)dt = f(x), \quad \int_{x=a}^b F'(x)dx = F(b) - F(a)$$

For the first part, the left hand side is only defined by f up to "almost everywhere". So this is the best you can expect. But then we need to ask when dF/dx makes sense, almost everywhere.

Dini had the idea of defining 4 derivates. When you define the difference quotient

$$\lim_{h\to 0} \frac{f(h + x_0) - f(x_0)}{h},$$

you actually have limit from the left and right, and lim sup and lim inf.

$$\limsup_{h\to 0^+} = D^+f(x_0), \quad \liminf_{h\to 0^+} = D_+, \quad \limsup_{h\to 0^-} D^-, \quad \liminf_{h\to 0^-} = D_-$$

Under what condition are they the same? Vitali decided to look at the cumulative effect of the errors. Assume that f is nondecreasing. The "bad" sets we want to look at is the ones with

$$\liminf \frac{f(x_0+h)-f(x_0)}{h} < \limsup \frac{f(x_0+h)-f(x_0)}{h}.$$

Because this is hard to control, fix rationals $\alpha < \beta$ and only look at the x_0 such that the left hand side is smaller than α and the right hand side greater than β.

Lemma 4.2. *If $E_k \nearrow E$ in $\mathbb{R}$ then $\lim_{k\to\infty} m_*(E_k) = m_*(E)$.*

Proof. Approximate E_k by the open sets from the outside. ☐

Lemma 4.3 (Vitali covering)**.** *Let $E \subseteq \mathbb{R}$ be $m_*(E) < \infty$. For every $x \in E$, suppose there exists a nonempty A_x of positive numbers. Then for every $\epsilon > 0$, there exist $x_1, \dots, x_k \in E$ and $r_{x_j} \in A_{x_j}$ such that $(x_j, x_j + r_{x_j})$ are all disjoint and*

$$m_*\left(E \cap \bigcup_{j=1}^{k} (x_j, x_j + r_{x_j})\right) \geq m_*(E) - \epsilon.$$

The main difficulty is that we are only allowed to use finitely many intervals.

Proof. Let

$$E_n = \{x \in E : \sup A_x > 1/n, x \in [-n, n]\}.$$

Then $E_n \nearrow E$ and so at sufficiently large n, we can replace E by E_n with at most ϵ error. Now you use the at least $1/n$-length intervals to cover E almost entirely. ☐

Let f is nondecreasing on $[a, b]$ with $a < b$ in $\mathbb{R}$. We would like to show that $D_+f(x) = D^+f(x)$ almost everywhere. You need two other similar statements, to get differentiability entirely.

For $0 \leq \alpha < \beta$ rational numbers, let

$$E_\alpha = \{x \in [a, b) : D_+f(x) < \alpha\}.$$

Now for each $x \in E_\alpha$ there exists a $r_x > 0$ such that $f(x + r_x) - f(x) < \alpha r_x$. Cover this set E_α by $(x_j, x_j + r_{x_j})$ up to error ϵ. Then the total length of the f-image of the union is going to be at most

$$\alpha \sum_j r_{x_j}.$$

Now replace $[a, b]$ by $[x_j, x_j + r_{x_j}]$ for each j to apply Vitali again. Define

$$F_{j,\beta} = \{x \in [x_j, x_j + r_{x_j}) : D^+f(x) > \beta\}.$$

We can then cover the intervals $[x_j, x_j + r_{x_j}]$ with $[x_{j,l}, x_{j,l} + s_{x_{j,l}}]$ up to measure ϵ. It then follows that the total length of the f-image of $\{[x_{j,l}, x_{j,l} + s_{j,l}]\}$ is at least β times the total length. These two inequalities give a contradiction as $\epsilon \to 0$. This finishes the proof.

4.2 Fundamental theorem of calculus

We now want to show that
$$\frac{d}{dx}\int_a^x f(t)dt$$
exists almost everywhere if f is integrable. In the Riemann case, you can do this because it is literally defined using rectangles and so taking the difference gives the rectangle.

For the Lebesgue case, let f be integrable on $[a, b]$, and let
$$F(x) = \int_a^x f(t)dt.$$
We may assume that $f > 0$ since we can take the positive and negative part, and write f as the difference of the two. This also means that dF/dx is almost everywhere defined.

The difference quotient is, if $\epsilon_n > 0$,
$$\int_a^b \frac{F(x+\epsilon_n) - F(x)}{\epsilon_n} = \frac{1}{\epsilon_n}\Big(\int_a^b F(x+\epsilon_n) - \int_a^b F(x)\Big) = \frac{1}{\epsilon_n}\Big(\int_b^{b+\epsilon_n} F - \int_a^{a+\epsilon_n} F\Big).$$
We now use the convergence theorem for Lebesgue integrals. Because F is continuous, the right hand side goes to $F(b) - F(a)$ as $\epsilon_n \to 0$. This shows, by Fatou's lemma,
$$\int_a^b F' \leq \liminf_{n\to\infty} \int_a^b \frac{F(x+\epsilon_n) - F(x)}{\epsilon_n} = F(b) - F(a).$$
Actually, this is true for any increasing continuous F. (Just forget about f.) We would now like to show that $F(b) - F(a) \leq \int_a^b F'$ and that $F' = f$.

Assume now that f is bounded. The claim is that $\int_a^b F' = F(b) - F(a)$. Instead of Fatou, we use bounded convergence. Note that
$$\frac{F(x+\epsilon_n) - F(x)}{\epsilon_n} = \frac{1}{\epsilon_n}\int_x^{x+\epsilon_n} f$$
is bounded. So we really get
$$F(b) - F(a) = \int_a^b F'$$
if f is bounded. In general, you truncate the range to get $f_n = \min\{f, n\}$. In this case, let $F_n = \int_a^x f_n$ so that $\int_a^b F_n' = F_n(b) - F_n(a)$.

5 September 14, 2017

As we have seen before, if f is integrable, then

$$\frac{d}{dx}\int_a^x f(t)dt = f(x).$$

Here the key is to look at the cumulative errors and monotone functions. Define $F(x) = \int_a^x f(t)dt$, and then $F'(x)$ exists almost everywhere, and the question is whether $F' = f$ almost everywhere. The way to show this is to show that $\int_a^b (F' - f) = 0$ for every $a < b$.

If f is bounded, then we can use bounded convergence theorem to get

$$\lim_{\epsilon\to 0}\int_a^b \frac{F(x+\epsilon) - F(x)}{\epsilon} = \int_a^b f.$$

If f is not bounded, we can use the truncation $f_n = \min(f, n)$ and then for $F_n(x) = \int_a^x f_n$, we have $F_n'(x) = f_n(x)$ almost everywhere. Fatou's lemma tells us that for any F increasing and continuous,

$$\int_a^b F' \leq F(b) - F(a).$$

In this case, $F' \geq F_n'$ pointwise almost everywhere, and so

$$\int_a^b F' \geq \int_a^b F_n' = F_n(b) - F_n(a) \to F(b) - F(a)$$

by monotone convergence.

Theorem 5.1 (Fundamental theorem of calculus I)**.** *If f is integrable, then*

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

for almost every x.

5.1 Fundamental theorem of calculus II

For the other version of the fundamental theorem of calculus, we need the condition that F is absolutely continuous. It took a long time for people to nail down this condition.

The measure-theoretic condition is that $\int_E f \to 0$ if $m(E) \to 0$. But this doesn't make sense, and we use the approximate version, which replaces E by a finite number of disjoint open intervals.

Definition 5.2. A function $F : [a, b] \to \mathbb{R}$ is called **absolutely continuous** if

$$\sum_{j=1}^{\ell} |F(b_j) - F(a_j)| < \epsilon$$

for disjoint open intervals $\bigcup_{j=1}^{\ell}(a_j, b_j) \subseteq [a, b]$ with $\sum_{j=1}^{\ell}(b_j - a_j) < \delta_\epsilon$.

This is stronger than uniform continuity, because that is only for one interval.

The technique for proving this is again comparing the functions. Let's assume for now that we know that F' exists and is integrable. Let

$$G(s) = \int_a^x F'.$$

Then $G' = F'$ almost everywhere, because we have proved this already, and we want to check that $G(x) = F(x) - F(a)$. So if $(G-F)' = 0$ is almost everywhere, then is $G-F$ continuous? This is true only when $G-F$ is absolutely continuous. (Look at the Cantor function, which have almost everywhere zero derivative.)

So how do you prove that F absolutely continuous implies F' exists almost everywhere? If we manage to split F into a difference of increasing functions $F = F_+ - F_-$, then each of them are differentiable and F is also differentiable. Consider a partition P of $[a, b]$ into $a = x_0 < \cdots < x_n = b$. Define

$$\operatorname{Var}(P, f) = \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})|, \quad \operatorname{Var}_a^b f = \sup_P \operatorname{Var}(P, f).$$

If f is monotone, then clearly $\operatorname{Var}_a^b f = |f(b) - f(a)|$. On the other hand, we always clearly have $\operatorname{Var}_a^b(f+g) \le \operatorname{Var}_a^b f + \operatorname{Var}_a^b g$.

Theorem 5.3. *A function f on $[a, b]$ is the difference of two monotone functions if and only if $\operatorname{Var}_a^b f < \infty$.*

Proof. We have already proven the forward direction. For the other direction, define

$$\operatorname{Var}_+(P, f) = \sum_{j=1}^{n} \max(f(x_j) - f(x_{j-1}), 0),$$
$$\operatorname{Var}_-(P, f) = \sum_{j=1}^{n} \min(f(x_j) - f(x_{j-1}), 0).$$

Then $\operatorname{Var}(P, f) = \operatorname{Var}_+(P, f) - \operatorname{Var}_-(P, f)$ and $f(b) - f(a) = \operatorname{Var}_+(P, f) + \operatorname{Var}_-(P, f)$. It follows that

$$2 \operatorname{Var}_+(P, f) = \operatorname{Var}(P, f) + (f(b) - f(a))$$

and likewise for Var_-. Then we can take sup on both sides. After doing this, we can define

$$f_+ = \sup_P \operatorname{Var}_+(P_{[a,x],f})$$

and likewise for f_- so that $f = f_+ - f_-$. These are clearly increasing functions, almost by definition. $\square$

We now have to show that absolute continuity implies finite total variation, but this is obvious. So F is almost everywhere differentiable. We now want to to show that F' is integrable. We can do this for monotone functions, and we actually know that the derivative of a monotone continuous function is integrable, by Fatou's lemma. Thus we need to check that f_+ and f_- are both continuous, if f is absolutely continuous.

Let's check left-continuity. Given any ϵ, is there a δ such that $|\mathrm{Var}^y_{+,a}(f) - \mathrm{Var}^x_{+,a}(f)| < \epsilon$ if $x - \delta < y < x$? This you can just use the assumption for absolute continuity. For right-continuity, you reflect the whole picture by using additivity of Var_+.

The second statement is that if F is absolutely continuous, $F' = 0$ almost everywhere, then F is constant. This is because you can pick a very small set such that F is locally constant outside. Then all the variation is inside this little set. This shows that every two point has difference below epsilon. That is, F is actually constant.

Theorem 5.4 (Fundamental theorem of calculus II)**.** *If F is absolutely continuous, F' is integrable and*

$$\int_a^b F' = F(b) - F(a).$$

5.2 High-dimensional analogue

For some hundred years, people tried to make rigorous what Fourier did. To do this, we need the commutation between integration and differentiation, or more generally the commutation between integration and limits. This is the whole point of convergence theorems. Under the fundamental theorem of calculus, the statement

$$\frac{d}{dy}\int_a^b f(x,y)dx = \int_a^b \frac{\partial f}{\partial y}(x,y)dx$$

could be made into a commutation of integration.

If we are able to change order of integration, we can write

$$\begin{aligned}\int_{\eta=c}^y \int_{x=a}^b \frac{\partial}{\partial\eta} f(x,\eta)d\eta &= \int_{x=a}^b \int_{\eta=c}^y \frac{\partial}{\partial\eta} f(x,\eta)d\eta \\ &= \int_{x=a}^b f(x,y)dx - \int_{x=a}^b f(x,c)dx\end{aligned}$$

by integrating the fundamental theorem of calculus. Then by fundamental theorem again, we get the result.

Fubini was the one who managed to prove this, and his idea was to define the "double integral" that is just defined on functions on $\mathbb{R}^2$. To do this, we need measure theory and Lebesgue theory on $\mathbb{R}^d$. The main difference is that we no longer have the structure theorem for open subsets in $\mathbb{R}$. But why do we need the structure theorem anyways? We used it to define the measure of an open set, but if we can do this, we're fine.

The building blocks are the closed cubes in $\mathbb{R}^d$. An open set O is always an almost disjoint union of a countable number of cubes in $\mathbb{R}^d$. Here, almost disjoint means that their interiors are disjoint. Is the measure defined in this way unique? Actually you don't have to worry about this, because you're going to define the exterior measure as

$$m_*(E) = \inf \sum_j m(Q_j).$$

Now you can repeat everything again and get a measure theory on $\mathbb{R}^d$.

The next step is the fundamental theorem of calculus, and here you have a trouble, because you have many variables. One analogue of this can be thought of as

$$\lim_{x \in B, m(B) \to 0} \frac{1}{m(B)} \int_B f = f(x)$$

almost everywhere. (Note the one-dimensional case.) People call this an "averaging problem", and was solved by Lebesgue.

How are we going to approach this problem? In the one-dimensional case, we used a truncation of f and used the approximation $f_n \nearrow f$. Here, we are going to approximate f by a continuous function g of compact support, in the L^1 norm. And then you hope that a 3ϵ argument will give work out. The approximation between $f(x)$ and $g(x)$ is called Tchebychev's theorem, and the comparison between $\frac{1}{m(B)} \int_B f$ and $\frac{1}{m(B)} \int_B g$ is the inequality on Hardy–Littlewood maximal function.

6 September 19, 2017

We proved the fundamental theorem of calculus, and we verified to some extent the change of integration and differentiation by Fubini's theorem. Now we want to talk about the fundamental theorem of calculus in higher dimensions. In the 1-dimensional case, the fundamental theorem is saying that

$$\lim_{x \to a} \frac{1}{x-a} \int_a^x f(t)dt = f(a).$$

We can replace this set $[a, x]$ by a set B and write

$$\lim_{x \in B} \frac{1}{m(B)} \int_B f = f(x).$$

This is something that can be generalized.

If you have a function f, you can associate to it a measure

$$E \mapsto \int_E f = \mu(E).$$

Then what we are actually looking at is the **Radon–Nikodym** derivative

$$\lim_{B \to p^+} \frac{\mu(E)}{m(E)} = g.$$

The second part of the fundamental theorem of calculus can then be stated as $\mu(E) = \int_E g$. This is what we are going to do.

Example 6.1. Before starting, let me mention the **Cantor function**. This is function $f : [0, 1] \to [0, 1]$ is defined as

$$x = 0.a_1a_2a_3\ldots_{(3)} \mapsto f(x) = 0.b_1b_2b_3\ldots_{(2)} \qquad \text{where } b_j = \left\lfloor \frac{a_j}{2} \right\rfloor.$$

This satisfies $f' = 0$ almost everywhere.

6.1 The averaging problem

We want to show that, if f on $\mathbb{R}^d$ is integrable, then

$$\lim_{m(B) \to 0} \frac{1}{m(B)} \int_{x \in B} f = f(x)$$

for almost every x. The proof is done by looking at the stronger version

$$\limsup_{m(B) \to 0} \frac{1}{m(B)} \int_B |f(y) - f(x)| dy = 0$$

for almost every 0. (This is stronger by the triangle inequality.)

The proof goes by approximation by g for which the statement is correct. For example, if g is continuous on $\mathbb{R}^d$ then the assertion is certainly true. The condition f integrable means that f can be approximated by a linear combination of characteristic functions of cubes, and then smooth out the corners to get a continuous function. So we get a sequence of functions g such that

$$\|f-g\|_{L^1} = \int_{\mathbb{R}^d} |f-g| \to 0.$$

Then we check that the statement carries out. We have

$$\frac{1}{m(B)}\int_B |f(y)-f(x)|dt \le \frac{1}{m(B)}\int_B |f(y)-g(y)| + \frac{1}{m(B)}\int_B |g(y)-g(x)|dy + |g(x)-f(x)|.$$

Applying lim sup on both sides, we get

$$\limsup_{m(B)\to 0} \frac{1}{m(B)}\int_B |f(y)-f(x)| \le \sup_{x\in B} \frac{1}{m(B)}\int_B |f(y)-g(y)|dy + |g(x)-f(x)|.$$

For $\alpha > 0$, let

$$\begin{aligned}
E_\alpha &= \Big\{x : \limsup_{m(B)\to 0} \frac{1}{m(B)}\int_B |f(y)-f(x)| > \alpha\Big\},\\
\tilde{E}_\alpha &= \Big\{x : \sup_{x\in B} \frac{1}{m(B)}\int_B |f(y)-g(y)|dy > \alpha\Big\},\\
\hat{E}_\alpha &= \{x : |g(x)-f(x)| > \alpha\}.
\end{aligned}$$

Then E_α is contained in $\tilde{E}_{\alpha/2} \cup \hat{E}_{\alpha/2}$.

Now we claim that

(1) $m_*(\tilde{E}_\alpha) \le \dfrac{3^d}{\alpha}\|f-g\|_{L^1}$,

(2) $m(\hat{E}_\alpha) \le \dfrac{1}{\alpha}\|f-g\|_{L^1}$.

This will finish the proof, because we have $\|f-g\|_{L^1} \to 0$ and then we can take the countable union over $\alpha > 0$.

The claim (2) is **Tchebychev's inequality**. In general, if $F \ge 0$ is integrable, then

$$m(\{F > \alpha\}) \le \frac{1}{\alpha}\int_{\mathbb{R}^d} F$$

because $F \ge \alpha\chi_{\{F>\alpha\}}$. (2) is simply this applied to $F = |f-g|$.

The claim (1) is more difficult. Define the **Hardy–Littlewood maximal function** of an integrable function F as

$$F^*(x) = \sup_{x\in B} \frac{1}{m(B)}\int_B |F|.$$

We are then claiming the weak-type inequality

$$m(\{F^* > \alpha\}) \leq \frac{3^d}{\alpha}\|F\|_{L^1}.$$

This is done by Vitali's covering technique.

Proposition 6.2. *Given a finite number of open balls $\mathcal{B}$, there exists a finite subcollection $\mathcal{B}'$ of disjoint open balls such that*

$$m\Big(\bigcup_{B\in\mathcal{B}} B\Big) \leq 3^d \sum_{B\in\mathcal{B}'} m(B).$$

Proof. Choose $B_0 \in \mathcal{B}$ with maximal radius. Consider the ball $3B_0$ with the same center and three-times radius. Among the balls in $\mathcal{B}$ that are not completely covered by $3B_0$, take the maximal radius ball $B_1 \in \mathcal{B}$. Note that B_0 and B_1 cannot intersect by maximality. Among the balls in $\mathcal{B}$ that are not completely covered by $3B_0 \cup 3B_1$, take the maximal radius ball $B_2 \in \mathcal{B}$. Repeat this until we exhaust of balls, and let $\mathcal{B}' = \{B_0, B_1, \dots\}$. $\square$

Theorem 6.3 (Weak-type inequality)**.** *If F is integrable, then*

$$m(\{F^* > \alpha\}) \leq \frac{3^d}{\alpha}\|F\|_{L^1}.$$

Proof. We first note that F^* is measurable because $\{F^* > \alpha\}$ is in fact open. Wiggling around doesn't change the integral much.

Consider any compact subset $K \subseteq E_\alpha = \{F^* > \alpha\}$. Then there is a finite covering of K by large balls, and then

$$m(K) \leq m\Big(\bigcup_{B\in\mathcal{B}} B\Big) \leq 3^d \sum_{B\in\mathcal{B}'} m(B) \leq \frac{3^d}{\alpha}\int_{B\in\mathcal{B}'}\int_B |F| \leq \frac{3^d}{\alpha}\|F\|_1.$$

This finishes the proof. $\square$

So we have shown that

$$\lim_{m(B)\to 0} \frac{1}{m(B)}|f(y) - f(x)|dy = 0 \tag{$*$}$$

for almost every x. This is actually not stronger that the first fundamental theorem of calculus

$$\lim_{m(B)\to 0} \frac{1}{m(B)} f(y)dy = f(x).$$

Definition 6.4. The set on which $(*)$ holds is called the **Lebesgue set** of f.

This has an interesting consequence. If $f = g$ almost everywhere, we regard them as more or less the same. The first fundamental theorem of calculus allows us to pick a representative of the equivalence class. Of course, the limit may not exist at some point, but we can then let them to take 0.

6.2 Convergence of Fourier series

This has something to do with the Lebesgue set. Start out with an integrable function f on $\mathbb{R}$ with period 2π. We define

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx,$$

and consider the sum

$$\begin{aligned} s_n(x) = \sum_{k=-n}^{n} c_k e^{ikx} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)\frac{\sin(n+\frac{1}{2})y}{\sin\frac{1}{2}y}dy \\ &= \frac{1}{2\pi}\int_{0}^{\pi} (f(x-y)+f(x+y))\frac{\sin(n+\frac{1}{2})y}{\sin\frac{1}{2}y}dy. \end{aligned}$$

(We are doing this last step so that there is a better chance of convergence.) So

$$s_n(x) - L = \frac{1}{2\pi}\int_{0}^{\pi} (f(x-y)+f(x+y)-2L)\frac{\sin(n+\frac{1}{2})y}{\sin\frac{1}{2}y}dy$$

and people want to see if this goes to 0 as $n \to \infty$.

Even if $f(x)$ is continuous, it is not true that $s_n \to f$. The Riemann–Lebesgue lemma tells us that if F is integrable and $a < b$ are finite, then

$$\int_a^b F(x)\sin\lambda x dx \to 0$$

as $\lambda \to \infty$. The way you do this is to approximate F by a smooth function G and estimate the difference. So people tried to group

$$\frac{f(x-y)+f(x+y)-2L}{\sin\frac{1}{2}y}$$

together. If f' exists, then we indeed let $L = f(x)$ and get an integrable function.

But this doesn't really work very nicely in general. Cesàro decided to look at the Lebesgue set.

Proposition 6.5. *If x is in the Lebesgue set, then*

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_{n-1}}{n} \to f(x)$$

as $n \to \infty$.

7 September 21, 2017

Let f be a integrable function on $\mathbb{R}$ with period 2π, and define

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx, \quad s_n = \sum_{k=-n}^{n} c_n e^{ikx}.$$

The first thing we want to establish is Dini's test, which is that the Fourier series for $f(x)$ converges at x to L if

$$\frac{f(x+y)+f(x-y)-2L}{y}$$

is integrable as a function of y near $y=0$.

There also Cesàro convergence. If x is in the Lebesgue set, then we can show that

$$\sigma_n = \frac{s_0+s_1+\cdots+s_{n-1}}{n} \to f(x).$$

Why is this interesting? We can interpret

$$\sigma_n = \frac{1}{n}\sum_{k=0}^{n-1} s_k = \sum_{k=0}^{n-1}\frac{n-k}{n}a_k$$

as a sum with some weight. We can also look at other weights, for instance, Abel's

$$A_r = \sum_{n=-\infty}^{\infty} r^{|n|}a_n.$$

Ordinary convergence implies Cesàro convergence, and this implies Abel's convergence. Can you go back? These are called Tauberian theorems, and is an interesting field.

7.1 Dirichlet test for Fourier series

Firstly, we have the Dirichlet kernel

$$D_n(x) = \sum_{k=-n}^{n} e^{ikx} = \frac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x},$$

and then

$$s_n - L = \frac{1}{2\pi}\int_{y=-\pi}^{\pi}(f(x-y)-L)D_n(y)dy = \frac{1}{2\pi}\int_{y=0}^{\pi}(f(x-y)+f(x+y)-2L)dy.$$

Does this go to 0 as $n\to\infty$?

The first thing done was to look at the function

$$\frac{f(x-y)+f(x+y)-2f(x)}{y}.$$

If this is integrable, then $s_n - f$ actually converge to 0.

Lemma 7.1 (Riemann–Lebesgue lemma)**.** *If $f(x)$ is integrable on $[a, b]$, then*

$$\int_a^b f(x) \sin \lambda x \to 0$$

as $\lambda \to \infty$.

Proof. You approximate f by a smooth function g in $L^1[a, b]$. We first check

$$\int_a^b g(x) \sin \lambda x = -\frac{\cos \lambda x}{\lambda} g(x)|_a^b + \int_a^b g'(x) \frac{\cos \lambda x}{\lambda} dx,$$

which goes to 0 as $\lambda \to \infty$. Now we have the estimate

$$\int_a^b f(x) \sin \lambda x dx = \int_a^b (f(x) - g(x)) \sin \lambda x + \int_a^b g(x) \sin \lambda x dx,$$

and we can use a 3ϵ argument to get the estimate. □

Corollary 7.2 (Dirichlet test)**.** *If $(f(x - y) + f(x + y) - 2f(x))/y$ is integrable as a function of y, then $s_n(x) \to f(x)$.*

7.2 Approximation to identity

Cesàro only defined the notion of convergence. The guy who actually proved this is Fejér. We have

$$F_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} D_k(x) = \frac{1}{n} \frac{1}{2\pi \sin \frac{1}{2} x} \sum_{k=0}^{n-1} \sin(k + \tfrac{1}{2})x = \frac{1}{2\pi n} \frac{\sin^2 \frac{1}{2} nx}{\sin^2 \frac{1}{2} x}.$$

This is the **Fejér kernel**. Then

$$\sigma_n(x) - f(x) = \frac{1}{2\pi n} \int_{y=0}^{\pi} \frac{\sin^2 \frac{1}{2} ny}{\sin^2 \frac{1}{2} y} (f(x + y) + f(x - y) - 2f(x)) dy.$$

If f is continuous at x, for instance, you can show that this goes 0 for $n \to \infty$, because we have $1/n$.

What is really the difference between D_n and F_n? The intuitive picture is that F_n is positive and approximates the Dirac delta for $n \to \infty$. On the other hand, D_n has some oscillation that does not disappear in magnitude as $n \to \infty$.

Let us make this more precise. A **good kernel** is a family of integrable functions K_δ for $\delta > 0$ satisfying

(1) $\displaystyle\int_{\mathbb{R}^d} K_\delta(x) dx = 1$ (unit interval),

(2) $\displaystyle\int_{\mathbb{R}^d} |K_\delta(x)| \leq A$ independent of δ (uniform integrability),

(3) for every $\eta > 0$, $\displaystyle\int_{|x|>\eta} |K_\delta(x)|dx \to 0$ as $\delta \to 0$ (small L^1 norm outside the origin).

This give convergence almost everywhere, but we don't know if things converge on Lebesgue sets.

So we strengthen the condition, and say that K_δ is an **approximation to identity** if it satisfies (1) and the following two strengthenings:

(2s) $|K_\delta(x)| \le \dfrac{A}{\delta^d}$ for all $\delta > 0$ (parameter pole order estimate),

(3s) $|K_\delta(x)| \le \dfrac{A\delta}{|x|^{d+1}}$ for all $\delta > 0$ (coordinate pole order estimate).

Proposition 7.3. *If K_δ is a good kernel, and f is an integrable function, then there is a sequence $\delta_\nu \to 0$ such that $f * K_{\delta_\nu} \to f$ almost everywhere.*

Proof. Write $f_y(x) = f(x - y)$. Note that

$$(f * K_\delta)(x) - f(x) = \int_{\mathbb{R}^d} (f_y(x) - f(x))K_\delta(y)dy.$$

So if we take the L^1-norm,

$$\begin{aligned} \|f * K_\delta - f\|_{L^1} &\le \int_{\mathbb{R}^d} \|f_y - f\|_{L^1} |K_\delta(y)|dy \\ &= \int_{|y|>\eta} \|f_y - f\|_{L^1} |K_\delta(y)|dy + 2\|f\|_{L^1} \int_{|y|<\eta} |K_\delta| \\ &\le A \sup_{|y|\le\eta} \|f_y - f\|_{L^1} + 2\|f\|_{L^1}\epsilon. \end{aligned}$$

Then we can use approximation of f by smooth functions to show $\lim_{\eta\to 0} \sup_{|y|\le\eta} \|f_y - f\|_{L^1} = 0$.

So we get $\lim_{\delta\to 0} \|f * K_\delta - f\| = 0$. You have shown in you homework that then there is a subsequence δ_ν such that $f * K_{\delta_\nu} \to f$ almost everywhere. □

Let us now look at the approximation to identity case. Consider an approximation by identity $K_\delta(x)$. If x is in the Lebesgue set, we have

$$A_x(r) = \frac{1}{r^d} \int_{|y|<r} |f(x - y) - f(x)|dy \to 0$$

as $r \to 0$. Also, $A_x(r)$ is continuous on $r > 0$ and uniformly bounded since it is at most $2\|f\|_{L^1}$ for $r > 1$.

Proposition 7.4. *If K_δ is an approximation to identity, f is measurable and x is a Lebesgue point, then $f * K_\delta(x) \to f(x)$.*

Proof. We have

$$\begin{aligned}
|(f * K_\delta)(x) - f(x)| &\le \int_{\mathbb{R}^d} |f(x-y) - f(x)||K_\delta(y)dy| \\
&= \int_{|y|<\delta} |f(x-y) - f(x)||K_\delta(y)|dy \\
&\qquad + \sum_{k=0}^{\infty} \int_{2^k\delta<|y|\le 2^{k+1}\delta} |f(x-y) - f(x)||K_\delta(y)|dy \\
&\le cA(\delta) + \sum_{k=0}^{\infty} \frac{c\delta}{(2^k\delta)^{d+1}} A(2^{k+1}\delta)
\end{aligned}$$

From the properties we had for $A(r)$, it follows that this goes to 0 as $\delta \to 0$. $\square$

7.3 Fourier transform

If f is integrable on $\mathbb{R}^d$, we can define the **Fourier transform**

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i x\cdot\xi} dx.$$

The inverse Fourier transform formula is

$$f(x) = \int_{\mathbb{R}^d} \hat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi.$$

You can ask when is the inverse Fourier transform actually the inverse of the Fourier transform? It turns out this is easier to deal with than the Fourier series, because f and $\hat{f}$ both are defined in $\mathbb{R}^d$.

The idea is that if f and $\hat{f}$ are in L^1, then the Fourier transform is formally symmetric. That is, if $f, g \in L^1$, then $\int f\hat{g} = \int \hat{f}g$. The verification is Fubini's theorem and rescaling the Gauss distribution to use it as a good kernel.

8 September 26, 2017

The difference between the Dirichlet kernel and the Fejér kernel is that the Dirichlet kernel has oscillation with constant amplitude, while the Fejér kernel has oscillation going to zero. So for convergence with the Dirichlet kernel, you need some nice properties for f. The Fejér kernel is nice, so the Fourier series converges to f on Lebesgue sets. So f can be approximated by some very nice functions. Moreover, $s_n = f * D_n$ so f can be "smoothed out" by convoluting.

8.1 Fejér–Lebesgue theorem

Theorem 8.1 (Fejér–Lebesgue theorem). *If $f \in L^1([0,2\pi])$, and x is in its Lebesgue set, then $\sigma_n(x) \to f(x)$ as $n \to \infty$.*

Proof. We have

$$\sigma_n(x) - f(x) = \int_{y=0}^{2\pi} F_n(y)(f(x+y) + f(x-y) - 2f(x))dy,$$

where F_n is the Fejér kernel. Here, we are going to take care of $0 < y < \eta$ with the Lebesgue point condition and handle $\eta < y < \pi$ with the oscillation. Let us write $\phi(y) = \frac{1}{2}(f(x+y) + f(x-y) - 2f(x))$ and

$$\Phi(t) = \int_{y=0}^{t} |\phi(y)|dt.$$

Then $\Phi(t) \leq 2\|f\|_{L^1}$ and the Lebesgue point property gives $\Phi(t)/t \to 0$ as $t \to 0$.

Given $\epsilon > 0$, choose $0 < \eta < \pi$ such that $0 \leq \Phi(t) < \epsilon t$ for all $t \leq \eta$. Now we want to show that

$$\frac{1}{2\pi n}\int_{y=0}^{\pi} \frac{\sin^2 \frac{\pi n y}{2}}{y^2}\phi(y)dy \to 0$$

as $n \to \infty$. We have from $0 < y < 1/n$,

$$\frac{1}{2\pi n}\int_0^{1/n} \frac{\sin^2 \frac{\pi n y}{2}}{y^2}\phi(y)dy \leq \frac{1}{2\pi n}n^2\int_0^{1/n}|\phi(y)|dy = \frac{n}{2\pi}\Phi\Big(\frac{1}{n}\Big) \to 0,$$

because $\sin^2\theta \leq \theta^2$.

For the other interval, we note that integration by parts give

$$\int_{y=a}^{b} \frac{\sin^2 \frac{\pi n y}{2}}{y^2} \leq \int_a^b \frac{1}{y^2}|\phi(y)|dy = \frac{\Phi(b)}{b^2} - \frac{\Phi(a)}{a^2} + 2\int_{y=a}^{b}\frac{1}{y^3}\Phi(y)dy.$$

So we have

$$\begin{aligned}\left|\frac{1}{2\pi n}\int_{1/n}^{\eta} \frac{\sin^2 \frac{\pi n y}{2}}{y^2}\phi(y)dy\right| &\leq \frac{1}{2\pi n}\frac{\Phi(\eta)}{\eta^2} - \frac{n}{2\pi}\Phi(\Big(\frac{1}{n}\Big) + \frac{1}{\pi n}\int_{y1/n}^{\eta}\frac{1}{y^3}\Phi(y)dy \\ &\leq \frac{\epsilon}{2\pi} + \frac{\epsilon}{\pi n}\int_{y=1/n}^{\eta}\frac{dy}{y^2} \leq \frac{3\epsilon}{2\pi}.\end{aligned}$$

Finally, we have

$$\frac{1}{2\pi n}\int_{y=\eta}^{\pi}\frac{\sin^2\frac{ny}{2}}{y^2}\phi(y)dy$$

goes to zero, because the amplitude of the oscillation goes to zero. $\square$

Consider the function given by

$$sf(x)=\frac{1}{2i}\sum_{n\neq 0}\frac{e^{inx}}{n}.$$

This the Fourier series of the sawtooth function, which looks something like

$$sf(x)=\begin{cases}-\frac{x+\pi}{2} & -\pi<x<0\\ -\frac{x-\pi}{2} & 0<x<\pi.\end{cases}$$

Let $P_n(\theta)$ be the principal sum

$$P_N(\theta)=\frac{1}{2i}\sum_{|n|<N}\frac{e^{in\theta}}{n}=\sum_{n=1}^{N}\frac{\sin n\theta}{n}.$$

This is actually uniformly bounded for all $N\in\mathbb{N}$ and all $\theta\in\mathbb{R}$. On the other hand, the non-principal sum

$$Q_N(\theta)=\sum_{n=-N}^{-1}\frac{e^{in\theta}}{n}$$

cannot converge because $|Q_N(0)|=\sum_{n=1}^{N}\frac{1}{n}\geq\log N$. To show that $P_N(\theta)$ is uniformly bounded, you use Abel summation.

Choose $\alpha_k>0$ and $N_k\in\mathbb{N}$ so that

(1) $\sum_{k=1}^{\infty}\alpha_k<\infty$,

(2) $N_{k+1}>3N_k$,

(3) $\alpha_k\log N_k\to\infty$ as $k\to\infty$,

e.g., $\alpha_k=k^{-2}$, $N_k=2^{k^3}$, and then construct the function

$$f(\theta)=\sum_{k=1}^{\infty}\alpha_k e^{i2N_k\theta}P_{N_k}(\theta).$$

The condition (2) ensures that the coefficients do not interfere. Then if we cut the sum at $2N_k$, then we get one large non-principal sum and some small principal sums that are uniformly bounded. This means that the Fourier series cannot converge.

8.2 Fourier inversion formula

For $f \in L^1(\mathbb{R})$, consider the function

$$\hat{f}(\xi) = \int_{x\in\mathbb{R}} f(x)e^{-2\pi i\xi x}dx.$$

We further assume that $\hat{f} \in L^1(\mathbb{R})$, and we want to prove that

$$f(x) = \int_{\xi\in\mathbb{R}} \hat{f}(\xi)e^{2\pi i\xi x}d\xi.$$

Here we are going to approximate the identity by Gauss distributions. These are given by

$$K_\delta(y) = \frac{1}{\delta^{d/2}}e^{-\pi|y|^2/\delta}.$$

The mean is 0 and the variance is $\delta/2\pi$. It is a good kernel with parameter $t = \sqrt{\delta}$.

The key argument is that if $f, g, \hat{f}, \hat{g} \in L^1(\mathbb{R}^d)$, then

$$\int_{y\in\mathbb{R}^d} f(y)\hat{g}(y)dy = \int_{\xi\in\mathbb{R}^d} \hat{f}(\xi)g(\xi)d\xi,$$

i.e., $f \mapsto \hat{f}$ is formally symmetric with respect to the pairing. This is just Fubini, because we can replace $\hat{g}$ and $\hat{f}$ with integrals.

We first note that $e^{-\pi x^2}$ is its own Fourier transform. This can be verified directly. Now take

$$g(\xi) = e^{-\pi\delta|\xi|^2}e^{2\pi i x\cdot\xi}, \quad \hat{g}(y) = \frac{1}{\delta^{d/2}}e^{-\pi|x-y|^2/\delta}.$$

Then as $\delta \to 0$, we immediately get by dominated convergence,

$$f(x) = \int_{\xi\in\mathbb{R}^d} \hat{f}(\xi)g(\xi)d\xi$$

for x a Lebesgue point.

8.3 Fubini's theorem

We haven't proved this, so I should talk about it. Let $f(x, y)$ be a function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and $f^y(x) = f(x, y)$.

Theorem 8.2 (Fubini's theorem)**.** *Suppose f is integrable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then f^y is integrable on $\mathbb{R}^{d_1}$ for almost all $y \in \mathbb{R}^{d_2}$, $y \mapsto \int_{\mathbb{R}^{d_1}} f(x)$ is integrable on $\mathbb{R}^{d_2}$, and*

$$\int_{\mathbb{R}^{d_2}}\left(\int_{\mathbb{R}^{d_1}} f dx\right)dy = \int_{\mathbb{R}^{d_1+d_2}} f.$$

Proof. Check this for a linear combination of characteristic functions of cubes. Then use monotone convergence theorem to check that this is correct for simple functions. □

9 September 28, 2017

Since we have everything, we can look at

$$\frac{d}{dy}\int_{x=a}^{b} f(x,y)dx = \int_{x=a}^{b} \frac{\partial f}{\partial y}(x,y)dy.$$

This holds when $\partial f/\partial y$ is $L^1([a,b]\times[c,d])$. As I have said, we can use Fubini and fundamental theorem of calculus.

9.1 Mean value theorem

There are two ways of looking at this. The first is to look through derivatives. The statement is that

$$F(b) - F(a) = F'(\xi)(b-a)$$

for some ξ. The integration formulation is

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

The interpretation is that the left hand side is the weighted average of the constant function 1. This can be replaced by a constant weight somewhere. You can also look at the Stieltjes version

$$\int_a^b f(x)\varphi(x)dx = f(\xi)\int_a^b \varphi(x)dx.$$

But what if you want to have the weight concentrated at the end? The second mean value theorem is, if φ monotone and f is an integrable function on (a,b), there exists a ξ such that

$$\int_a^b f(x)\varphi(x)dx = \varphi(b-)\int_a^\xi f(x)dx + \varphi(a+)\int_\xi^b f(x)dx.$$

Proof. Suppose φ is increasing. We may assume that $\varphi(a^+) = 0$. Now let $m = \min_\xi \int_\xi^b f(x)dx$ and $M = \max_\xi \int_\xi^b f(x)dx$. Then by Abel summation, we have the inequality

$$\varphi(b)m \le \int_a^b f(x)\varphi(x)dx \le \varphi(b)M.$$

Then there exists a ξ such that $\varphi(b)\int_\xi^b f(x)dx = \int_a^b f(x)\varphi(x)dx$. □

When we were talking about the fundamental theorem of calculus in 1-dimension, the platform for the functions was integrable and absolutely continuous. In higher dimensions, the platform the first part is integrable. Then we can talk about Lebesgue points and so forth. What about the second part? If you think about $\int_a^b F' = F(b) - F(a)$, these are two different ways of giving an identical "measure" to an interval. So the higher dimensional analogue will also be equating two measures.

This is how abstract measure theory work. People look at abstract measure spaces and also define signed measures. For two measures μ and ν, we say that μ is **absolutely continuous** with respect to ν if $\nu(E) = 0$ implies $\mu(E) = 0$. Then there is a **Radon–Nikodym derivative**, which is a measurable function, such that

$$\mu(E) = \int_E \frac{d\mu}{d\nu} d\nu.$$

9.2 Hölder and Minkowski inequalities

We now want to solve linear differential equations with compatibility conditions. The theme is generalized Cramer's rule applied to infinite dimensional spaces. We want to solve $Ax = b$ subject to $Sb = 0$. In the finite dimensional case, we get that the minimal solution is

$$x_{\min} = A^*(AA^* + S^*S)^{-1}b.$$

To get the inverse, we need that the map is invertible, i.e., has nonzero eigenvalues. We have

$$\langle (AA^* + S^*S)y, y\rangle = \|A^*y\|^2 + \|Sy\|^2,$$

and so is automatically invertible. But for this to be true for function spaces, we need an estimate. And to formulate this, we need to set up the notion of Hilbert spaces.

In an inner product space, we have the parallelogram law

$$\|u - v\|^2 + \|u + v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

Definition 9.1. Let E be a measurable set. For $p \geq 1$, the space $L^p(E)$ is the space of measurable functions f such that $\int_E |f|^p < \infty$. The L^p-norm is defined as

$$\|f\|_{L^p} = \|f\|_p = \left(\int_E |f|^p\right)^{1/p}.$$

The first thing you need to worry about is the triangle inequality. This is Minkowski's inequality, and it comes from concavity. The most basic inequality is $\sqrt{ab} \leq (a+b)/2$, which can be shown easily, and its general case is $a^\alpha b^\beta \leq \alpha a + \beta b$ for $\alpha + \beta = 1$ and $a, b, \alpha, \beta > 0$. The condition $\alpha + \beta$ ensures homogeneity and so we may set $b = 1$ and show $a^\alpha \leq \alpha(a - 1) + 1$. The left hand side is a concave function in a, and the right hand side is a tangent line at $a = 1$.

From this, we can deduce other inequalities. Suppose we have numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, and by normalization assume $\sum a_j = \sum b_j = 1$. Using

$a_j^\alpha b_j^\beta \leq \alpha a_j + \beta b_j$, we get

$$\sum_{j=1}^n a_j^\alpha b_j^\beta \leq 1 = \Big(\sum_{j=1}^n a_j\Big)^\alpha \Big(\sum_{j=1}^n b_j\Big)^\beta.$$

If we rescale a_j and b_j, and let $p = 1/\alpha$, $q = 1/\beta$, then we get the standard **Hölder inequality**

$$\sum_{j=1}^n a_j b_j \leq \Big(\sum_{j=1}^n a_j^p\Big)^{1/p} \Big(\sum_{j=1}^n b_j^q\Big)^{1/q}.$$

We can do this with functions, and then we will get

$$\|fg\|_{L^1} \leq \|f\|_{L^p}\|g\|_{L^q}.$$

Minkowski's inequality is Hölder's inequality with handling the coefficients. We have

$$\begin{aligned}\int_E |f+g|^p &= \int_E |f+g||f+g|^{p-1} \leq \int_E |f||f+g|^{p-1} + \int_E |g||f+g|^{p-1} \\ &\leq \Big(\int_E |f|^p\Big)^{1/p}\Big(\int_E |f+g|^p\Big)^{1-1/p} + \Big(\int_E |g|\Big)^{1/p}\Big(\int_E |f+g|^p\Big)^{1-1/p}.\end{aligned}$$

Then we can move things to the other side to get the **Minkowski inequality**

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

9.3 Hilbert space

The prototype for our differential equation is

$$\begin{cases} \frac{\partial u}{\partial x}(x,y) = P(x,y) \\ \frac{\partial u}{\partial t}(x,y) = Q(x,y) \end{cases} \qquad \text{subject to } \frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0.$$

This can be handled using Poincaré's lemma, by integrating. But we want to do things more generally.

Definition 9.2. A **Hilbert space** (over $\mathbb{C}$) with a Hermitian inner product that is

(i) complete, i.e., it is complete respect to the metric defined by the norm,

(ii) separable, i.e., there exists a countable dense subset.

Example 9.3. For any measurable space E, the space $L^2(E)$ is a Hilbert space. We know that it is separable, because every function can be approximated by simple functions and we can make them have rational coefficients.

To show completeness, consider a Cauchy sequence f_n and then choose a subsequence such that $\|f_n - f_{n+1}\| \leq 2^{-n}$. Then $\sum_n \|f_n - f_{n+1}\| < \infty$. Look at the telescopic series

$$f_n = f_1 + (f_2 - f_1) + \cdots + (f_{n-1} - f_{n-2}) + (f_n - f_{n-1}).$$

The sequence of functions

$$g_n = |f_1| + |f_2 - f_1| + \cdots + |f_n - f_{n-1}|$$

converges to some g_∞, which is L^2 by Minkowski and monotone. This shows that f_n converges almost everywhere, and the limit is L^2.

We are trying to solve differential equations $Tf = g$ where $Sg = 0$. But in general T is going to be a differential operator, and $T : L^2 \to L^2$ is not going to be always defined. But it is densely defined, and we will be able to apply Cramer's rule here.

We would want to make sense of the adjoint operator. This is supposed to be something that satisfy

$$(Tf, g) = (f, T^* g).$$

Given g, we want to define $T^* g$ such that the identity is true for all f. That is, we want to represent the map $f \mapsto (Tf, g)$ by the inner product with some element. This step is called the Riesz representation theorem.

10 October 3, 2017

We will be looking at linear differential equations. We can ask how to get $A^*(AA^* + S^*S)^{-1}b$ in function spaces. So we look at Hilbert spaces, which are complete and separable. One difficulty, which help people until Friedrichs(1944), was that some operators are not defined on the domain.

The operators like d/dx are not defined on $L^2(\mathbb{R})$. Rather, it is defined on differentiable functions, which is dense. The operator T^*means that

$$(Tx, y) = (x, T^*y).$$

That means that we fix a y and are trying to find a T^*y.

10.1 Riesz representation theorem

Proposition 10.1. *Let X be a Hilbert space and Y be a closed subspace (so it's also a Hilbert space). For $v \in X - Y$, there exists a unique $w \in Y$ such that $v - w \perp Y$.*

Proof. Get $w \in Y$ to minimize $\|v - w\|$. Rigorously, let $\mu = \inf_{y \in Y} \|v - y\|$ and there is a sequence $\|v - y_n\| \to \mu$. Then

$$\|y_n - y_m\|^2 + \|2v - y_n - y_m\|^2 = 2\|v - y_n\|^2 + 2\|v - y_m\|^2$$

by the parallelogram law. Then $\|2v - y_n - y_m\|^2 \geq 4\mu^2$ and the right hand side goes to $4\mu^2$. This shows that $\|y_n - y_m\| \to 0$ and that this is a Cauchy sequence. Then $y_n \to w$.

Now we need to show that it is perpendicular. Let $0 \neq y \in Y$. We have

$$\mu^2 \leq \|v - w + \lambda y\|^2 = \|v - w\|^2 + 2\lambda(v - w, y) + \lambda^2 \|y\|^2.$$

If we take $\lambda = -(v - w, y)/\|y\|^2$, we get $(v - w, y) = 0$. □

Theorem 10.2 (Riesz representation)**.** *Let X be a Hilbert space and let $f : X \to \mathbb{R}$ be an $\mathbb{R}$-linear continuous functional. (Then there exists a c_f such that $\|f(x)\| \leq c_f \|x\|$. The minimal such c_f is written as $\|f\|$.) Then there exists a unique $v_f \in X$ such that $f = (-, v_f)$.*

Proof. Assume $f \neq 0$. Let $Y = \ker f$, so that $Y \subsetneq X$ is a closed subspace. Let $v \in X$ such that $f(v) \neq 0$, and take $w = \mathrm{proj}_Y x$. Replace v by $v - w$, and we may assume that $v \perp Y$ and also that $\|v\| = 1$ by normalization.

We seek $v_f = \lambda v$ for some $\lambda \in \mathbb{R}$. We do this by testing at v. We need $(v, v_f) = f(v)$, so we use $\lambda = f(v)$. Now we claim that

$$v_f = f(v)v$$

does the job. Because x decomposes into a linear multiple of v and an element of Y we only need to show $(x, f(v)v) = f(x)$ for $x \in Y$ and $x = v$. We have already checked it for $x = v$. For $x \in Y$, $(x, f(v)v) = 0 = f(x)$. □

Now it can be proved that $\|v_f\| = \|f\|$. First,

$$|f(x)| \leq \|x\|\|v_f\|$$

by Cauchy–Schwartz and so $\|f\| \leq \|v_f\|$. On the other hand, testing at v_f gives $\|f\|\|v_f\| = f(v_f) = \|v_f\|^2$.

10.2 Adjoint operator

Let $T : X \to X$ be an $\mathbb{R}$-linear continuous map. We define the **adjoint** T^* as

$$(Tx, y) = (x, T^*y).$$

If we fix y, then

$$|(Tx, y)| \leq \|Tx\|\|y\| \leq \|T\|\|y\|\|x\|,$$

and so there exists this element T^*y that represents the map. Moreover, we would have the inequality

$$\|T^*y\| = \|x \mapsto (Tx, y)\| \leq \|T\|\|y\|.$$

This shows that T^* is well-defined and that $\|T^*\| \leq \|T\|$. By symmetry and since $(T^*)^* = T$, we get $\|T^*\| = \|T\|$.

Now let us look at the real situation. We have

$$H_1 \xrightarrow{T} H_2 \xrightarrow{S} H_3,$$

where H_1, H_2, H_3 are Hilbert spaces but T and S are only densely defined. We can assume that the graphs of T and S are closed.

What do I mean by this? We want to make this function defined on a domain as large as possible, so we look at the closure of the graph

$$\{(x, Tx) : x \text{ in domain of } T\} \subseteq H_1 \times H_2.$$

But there might be an ambiguity, so that $x_\nu \to x^* = 0$ but $y_\nu \to y^* \neq 0$. It turns out that differential operators have this property, and this is included in the definition.

Consider the space C_0^∞ of compactly supported smooth functions. Here we have integrating by parts. So if $T = d/dx$ and $f_\nu \to 0$ and $f'_\nu \to g$ in L^2, then

$$(f'_\nu, h) = (f_\nu, h') \to 0$$

for all $h \in C_0^\infty$ and so $g = 0$.

If T is densely defined with closed graph, then we can define its **adjoint** T^*. Here, $g \in \operatorname{dom}(T^*)$ if and only if there exists an h such that

$$(f, h)_{H_1} = (Tf, g)_{H_2}$$

for all $f \in \operatorname{dom}(T)$. In this case, we say $h = T^*g$.

Proposition 10.3. *Assume that $ST = 0$, i.e.,* $\mathrm{dom}(S) \supseteq \mathrm{im}(T)$ *and* $ST = 0$. *Also assume we have the a priori estimate*

$$\|T^*g\|^2 + \|Sg\|^2 \geq c\|g\|^2$$

for all $g \in \mathrm{dom}(T^*) \cap \mathrm{dom}(S)$, *for some* $c > 0$. *The conclusion is that if* $g \in \mathrm{dom}\, S$ *such that* $Sg = 0$, *then there exists a* $f \in \mathrm{dom}\, T$ *such that* $Tf = g$ *and* $f \perp \ker T$.

You could try to define $(T^*T + SS^*)^{-1}$, but it becomes very complicated because the domain is bad. Instead, we do this directly.

Proof. Take any $g \in H_2$. Write $g = g_1 + g_2$ with $g_1 \in \ker S$ and $g_2 \in (\ker S)^\perp$. ($\ker S$ is closed because the graph is closed.) Then we have

$$g_2 \in (\ker S)^\perp \subseteq (\mathrm{im}\, T)^\perp \subseteq \ker T^*.$$

The idea is that we want to solve $Tf = g$ weakly, so that $(Tf, u) = (g, u)$ for all test functions u. We'll finish next time. $\square$

11 October 5, 2017

We need to prove the following proposition.

Proposition 11.1. *Let $S : H_1 \to H_2$ and $T : H_2 \to H_3$ be densely defined closed operators between Hilbert spaces and $ST = 0$ and*

$$\|T^* g\|_{H_1}^2 + \|Sg\|_{H_3}^2 \geq c\|g\|_{H_2}^2$$

for all $g \in \operatorname{dom}(T^*) \cap \operatorname{dom}(S)$. Then the equation $Tu = f$ with $f \in \ker S$ can be solved uniquely with $u \perp \ker T$ and moreover there is the estimate

$$\|u\|_{H_1} \leq \frac{1}{\sqrt{c}}\|f\|_{H_2}.$$

Proof. We want to do this in the weak sense, so that we want to make so that $(u, T^*g) = (Tu, g) = (f, g)$. To get u, we would need to use the Riesz representation theorem to $T^*g \mapsto (g, f)$. This is not defined on all of H_1. To do this, we are going to prove that this is bounded. Then we can take the closure.

We have

$$|(g, f)| \leq \|g\|\|f\| \leq \frac{\|f\|}{\sqrt{c}}\sqrt{\|T^*g\|^2 + \|Sg\|^2},$$

but only when $g \in \operatorname{dom} S$. We also have the additional term $\|Sg\|^2$.

The idea is to decompose $g = g_1 + g_2$ with $g_1 \in \ker S$ and $g_2 \in (\ker S)^\perp$. Here, $\ker S$ is closed because S is a closed operator. Because $ST = 0$, $\operatorname{im} T \subseteq \ker S$. Then $h \mapsto (g_2, Th) = 0$ because $g_2 \in (\ker S)^\perp$ and $Th \in \ker S$. This shows that $g_2 \in \operatorname{dom} T^*$ with $T^* g_2 = 0$. Therefore if $g \in \operatorname{dom} T^*$ then $g_1, g_2 \in \operatorname{dom} T^*$.

Now for any $g \in \operatorname{dom} T^*$, we get

$$|(g, f)| = |(g_1, f)| \leq \frac{\|f\|}{\sqrt{c}}\|T^* g_1\| = \frac{\|f\|}{\sqrt{c}}\|T^* g\|.$$

So $T^*g \mapsto (g, f)$ is bounded by $\|f\|/\sqrt{c}$. This is a map $\operatorname{im} T^* \to \mathbb{R}$ that is bounded, and so we can extend it to its closure $\overline{\operatorname{im} T^*}$ and the to H_1 by orthogonal projection. Now by Riesz representation theorem, there exists a $u \in H_1$ such that

$$\|u\| \leq \frac{\|f\|}{\sqrt{c}}$$

such that $(T^*g, u) = (g, f)$ for all $g \in \operatorname{dom} T^*$.

This implies that $f = (T^*)^* u$ with $u \in \operatorname{dom}(T^*)^*$. But it can be shown that $(T^*)^* = T$ and so $f = Tu$. $\square$

So let's prove that $(T^*)^* = T$.

Theorem 11.2. *Let $T : H_1 \to H_2$ be densely defined with closed graph. Then T^* is densely defined with closed graph and $(T^*)^* = T$.*

Proof. The idea is to go to the product space $H_1 \times H_2$. The orthogonal complement of the graph is

$$(\operatorname{Graph} T)^\perp = \{(-T^* f, f) : f \in \operatorname{dom} T^*\},$$

just by definition of T^*. Then by closedness of the graph, we get

$$H_1 \times H_2 = (\operatorname{Graph} T) \oplus (\operatorname{Graph} T)^\perp.$$

If $\operatorname{dom} T^*$ is not dense, there exists a nonzero $h \in H_2$ that is perpendicular to $\operatorname{dom} T^*$. Then $(h, f)_{H_2} = 0$ for all $f \in \operatorname{dom} T^*$ and so

$$((0, h), (-T^* f, f)) = (0, -T^* f) + (h, f) = 0$$

for all $f \in \operatorname{dom} T^*$. Then $(0, h) \in \operatorname{Graph}(T)$ which is not possible unless $h = 0$. This shows that $\operatorname{dom} T^*$ is dense.

Finally, $T = (T^*)^*$ because $\operatorname{Graph}((T^*)^*)$ is going to be the orthogonal complement of $\operatorname{Graph}(T^*)$, which is $\operatorname{Graph}(T)$. □

11.1 Approximating by convolutions

So we have the basic theorem we can apply to real situations. The problem is that we really need this approximation $TT^* + S^* S \geq c > 0$. The technique here will be sum of squares.

Consider a differential operator

$$L = \sum_{j=1}^n a_j^{(k)}(x_1, \dots, x_n) \frac{\partial}{\partial x_j} + b_j^{(k)}(x_1, \dots, x_n).$$

We can clearly apply it to smooth compactly supported $u \in C_0^\infty$. Then if $(u, Lu) \to (v, w)$ in L^2, we have to define $w = Lv$. Friedrichs' idea was to do things in $u \in C_0^\infty$ and then pass to general (v, w). To approximate (v, w) by such (u, Lu), you take a convolution and smooth out.

For instance, we consider a $\chi(x) \geq 0$ in C_0^∞ with $\int_{\mathbb{R}} \chi = 1$. Then we scale

$$\chi_\epsilon(x) = \frac{1}{\epsilon} \chi\Big(\frac{x}{\epsilon}\Big),$$

so that the support has order ϵ.

Now assume $u \in \operatorname{dom} L$ and let $u_\epsilon = u * \chi_\epsilon$. Now we are good if we could prove that

$$(u_\epsilon, Lu_\epsilon) \to (u, Lu)$$

in the L^2 norm of the product space.

Theorem 11.3. *Let $u, Lu \in L^2$ for some differential operator L. Then for any $u \in C_0^\infty$, $L(u * \chi_\epsilon) \to Lu$ in L^2 as $\epsilon \to 0$.*

Proof. The main idea is the 3ϵ argument applied to the sequence. We have

$$u_\epsilon = u * \chi_\epsilon \to u, \quad (Lu) * \chi_\epsilon \to Lu$$

in L^2, by the good kernel argument. So what we are trying to show is

$$L(u * \chi_\epsilon) - (Lu) * \chi_\epsilon \to 0$$

in L^2.

For simplicity, consider the operator

$$L = a(x)\frac{d}{dx} + b(x).$$

Suppose we show that

$$\|\chi_\epsilon * Lu - L(\chi_\epsilon * u)\| \leq C\|u\| \tag{†}$$

independent of ϵ. Given any $\delta > 0$, $v \in C_0^\infty$, we have

$$\|(\chi_\epsilon * Lu) - L(\chi_\epsilon * u)\| \leq \|\chi_\epsilon * (L(u-v)) - L(\chi_\epsilon * (u-v))\| + \|\chi_\epsilon * Lv - L(\chi_\epsilon * v)\|.$$

But the first term is bounded by $C\|u-v\|$ and the second part goes to 0 as $\epsilon \to 0$. So if we prove this statement, we get

$$\chi_\epsilon * Lu - L(\chi_\epsilon * u) \to 0$$

as $\epsilon \to 0$.

Now let us show the inequality (†). We have

$$\begin{aligned}
\chi_\epsilon * Lu - L(\chi_\epsilon * u) &= \chi_\epsilon\Big(a\frac{\partial u}{\partial x}\Big) - a\frac{\partial}{\partial x}(\chi_\epsilon * u) \\
&= \chi_\epsilon * \Big(\frac{\partial}{\partial x}(au)\Big) - \chi_\epsilon * \Big(u\frac{\partial a}{\partial x}\Big) - a\frac{\partial}{\partial x}(\chi_\epsilon * u) \\
&= \Big(\frac{\partial}{\partial x}\chi_\epsilon\Big) * (au) - a\Big(\Big(\frac{\partial}{\partial x}\chi_\epsilon\Big) * u\Big) - \chi_\epsilon * \Big(u\frac{\partial a}{\partial x}\Big).
\end{aligned}$$

The last term goes to zero, so the first two terms is

$$\begin{aligned}
\int \frac{\partial \chi_\epsilon}{\partial y}(y)\, a(x-y)u(x-y) - a(x)\frac{\partial \chi_\epsilon}{\partial y}(y)u(x-y) \\
= \int \frac{\partial \chi_\epsilon(y)}{\partial y}(a(x-y) - a(x))u(x-y)dy.
\end{aligned}$$

Now as $\epsilon \to 0$, the first derivative grows with order ϵ^{-1} and $a(x-y) - a(x)$ decreases with order ϵ. $\square$

12 October 10, 2017

We have seen how to solve the equation $Tu = f$, when there is an estimate

$$\|T^* g\|^2 + \|Sg\|^2 \geq c\|g\|^2$$

and $Sf = 0$. Now if we actually solve the problem, we want to have something like regularity. If $Tu = f$ is smooth and u is orthogonal to $\ker T$, is u smooth? To do this, we are going to look at Sobolov spaces and Gårding's inequality.

To actually get the solution, we need to know how to invert operators. Here, we invert $TT^* + S^*S$, which is self-adjoint. Self-adjoint operators can be diagonalized, and this is how we invert, as long as there is no zero eigenvalue. The diagonalization statement is called the spectral theorem.

12.1 Compact operators

Let H be a Hilbert space, and let $T : H \to H$ be a bounded self-adjoint operator. We want to diagonalize this, i.e., find orthonormal $\varphi_n \in H$ such that $T\varphi_n = \lambda_n \varphi_n$. Here, orthonormality means $(\varphi_m, \varphi_n) = \delta_{mn}$ and every $f \in \mathcal{H}$ can be approximated by $\sum_{n=1}^{N} (f, \varphi_n)\varphi_n$.

Here, we need the additional assumption that the map is almost finite-dimensional. That is, there exist finite rank operators T_n such that

$$\lim_{n\to\infty} \|T - T_n\| = 0.$$

Definition 12.1. An operator $T : H \to H$ is called **compact** if for every sequence f_n bounded in H, there exists a subsequence f_{n_k} such that Tf_{n_k} is Cauchy.

In other words, this means that the image of the unit ball is sequentially compact. The condition that T can be approximated by finite rank operators is equivalent to T being compact.

Theorem 12.2 (Spectral theroem)**.** *If $T : H \to H$ is compact self-adjoint operator, then there exists an orthonormal basis φ_n with $T\varphi_n = \lambda_n \varphi_n$, and for each $c > 0$, there exists finitely many n such that $|\lambda_n| \geq c$.*

How can this be used? We can, for instance, look at **Sturm–Liouville equations**

$$(Lf)(x) = \frac{1}{r(x)}\Big(\frac{d}{dx}\Big(p(x)\frac{d}{dx}f(x)\Big) - q(x)f(x)\Big)$$

This is actually self-adjoint, because we formally have (forgetting about r for a moment)

$$(Lf, g) = ((pf')' - qf, g) = -(pf', g') - (qf, g).$$

Then $r(x)$ is just gives a weight on the inner product. This only holds formally, because we have some terms on the boundary. To make this actually work, we need to work with the space of all L^2 functions on $[a, b]$ vanishing at a and b.

Let H be a Hilbert space, and let $L(H)$ be all bounded linear operators from H to H. This is an algebra, i.e., there is bilinear multiplication. Let $L_c(H)$ be the space of all compact operators from H to itself.

Theorem 12.3. *$L_c(H)$ is a two-sided, closed, adjoint-invariant ideal $L(H)$. In $L_c(H)$, the finite-rank elements of $L_c(H)$ is dense.*

Proof. We need to show that if $T \in L(H)$ and $S \in L_c(H)$, then $TS \in L_c(H)$ and $ST \in L_c(H)$. This is clear, just by mapping the sequence.

To show closedness, we need to check that if $\|T - T_n\| \to 0$ and $T_n \in L_c(H)$ then $T \in L_c(H)$. This can be done by diagonal subseqeunce from a sequence of nested subsequences. The usual 3ϵ argument works.

Now let us show that elements in $L_c(H)$ can be approximated by finite rank operators. Let $\{e_n\}$ be an orthonormal basis. Then define

$$T_n f = \sum_{k \le n} (Tf, e_k) e_k, \quad Q_n f = \sum_{k > n} (f, e_k) e_k$$

so that $T = T_n + Q_n T$. We need to show that $\|Q_n T\| \to 0$. Suppose not. Then there exists a sequence f_n with $\|f_n\| = 1$ and $\|Q_n T f_n\| \ge c > 0$. Then let $Tf_{n_k} \to g$. Then

$$Q_{n_k} g = Q_{n_k} T f_{n_k} - Q_{n_k}(T f_{n_k} - g)$$

we get a contradiction as we send $k \to \infty$.

Finally, we need to show that $T \in L_c(H)$ implies $T^* \in L_c(H)$. But if we write $I = P_n + Q_n$ with $\|P_n T - T\| \to 0$, then $\|T^* P_n^* - T^*\| \to 0$. Here P_n^* has finite rank. □

12.2 Spectral theorem

The idea for proving the spectral theorem is to find one eigenvector v, and then replace the whole thing to $v^\perp$ to get the next eigenvector.

Theorem 12.4 (Spectral theroem)**.** *If $T : H \to H$ is compact self-adjoint operator, then there exists an orthonormal basis φ_n with $T\varphi_n = \lambda_n \varphi_n$, and for each $c > 0$, there exists finitely many n such that $|\lambda_n| \ge c$.*

Lemma 12.5. $\|T\| = \sup_{\|x\| \le 1} |(Tx, x)|$.

Proof. First we see that $\|T\| = \sup_{\|x\|, \|y\| \le 1} |(Tx, y)|$. One direction is Cauchy–Schwartz and the other direction is by specifying $y = Tx / \|Tx\|$. Now to change (Tx, y) to (Tx, x), we use polarization

$$(Tx, y) = \frac{1}{4} \sum_{k=0}^{3} (T(x + i^k y), x + i^k y).$$ □

Proof of the spectral theorem. Let $\mu = \|T\|$. Then there exist x_n with $\|x_n\| = 1$ such that $(Tx_n, x_n) \to \pm\mu$. (This is from the lemma.) By compactness, there exists a subsequence of Tx_{n_k} that converges, and let

$$\lim_{k \to \infty} T x_{n_k} = y.$$

The claim is that y is an eigenvector. We have

$$0 \leq (Tx_n - \mu x_n, Tx_n - \mu x_n) \leq \mu^2 - 2\mu(Tx_n, x_n) + \mu^2 \to 0$$

as $n \to \infty$. This shows that

$$x_n = \frac{1}{\mu}(Tx_n - (Tx_n - \mu x_n)) \to \frac{1}{\mu}y.$$

Then because T is bounded, $Ty = \mu \lim_{n\to\infty} Tx_n = \mu y$.

As we take away eigenvectors, the eigenvalues have to go down to zero, because otherwise the eigenvectors themselves form a sequence in the unit ball whose image doesn't contain a Cauchy subsequence. □

In general, we have an equation like $Lf = g$, where L is not compact and not even bounded. But in many cases $1 + L^*L$ is bounded, and we can hope that it is bounded. Let's look at the equation, and for instance,

$$Lf = \frac{d^2}{dx^2} f = g$$

with $f(a) = f(b) = 0$. This operator L is not compact, but Green figured out that L^{-1} is compact. He constructed the kernel

$$K(x, y) = \begin{cases} \frac{(x-a)(b-y)}{a-b} & a \leq x \leq y \leq b, \\ \frac{(b-x)(y-a)}{a-b} & a \leq y \leq x \leq b. \end{cases}$$

Then the solution is given by

$$f(x) = \int_{y=a}^{b} K(x, y) g(y) dy.$$

13 October 12, 2017

Last time we have proved the spectral theorem for self-adjoint compact operators. For instance, let us look at integral operators. Let $L^2(\mathbb{R}^d)$ and $K(x,y) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$. Then we can look at the integral operator, or **Hilbert–Schmidt operator**

$$(Tf)(x) = \int_{y\in\mathbb{R}^d} K(x,y)f(y)dy.$$

Then we know that

$$\|T\| \leq \|K\|_{L^2(\mathbb{R}^d,\mathbb{R}^d)}.$$

13.1 Sturm–Liouville equation

These can help us in solving differential equations. For instance, take $f(x)$ defined on $[a,b]$, and look at the toy model

$$(Lf)(x) = f''(x).$$

The motivation is that if you integrate this to remove the derivative, you get a single integral. By integration by parts, you can replace multiple integration by single integration, like

$$\begin{aligned} f(x) = f(a) + \int_a^x f'(t)dt &= f(a) + (x-a)f'(a) + \int_a^x (x-t)f''(t)dt \\ &= f(a) + (x-a)f'(x) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \int_a^x \frac{(x-t)^{n-1}}{(n-1)!}f^{(n)}(t)dt. \end{aligned}$$

So

$$\frac{d^n}{dx^n}f(x) = \frac{d^n}{dx^n}\int_a^x \frac{(x-t)^{n-1}}{(n-1)!}f^{(n)}(t)dt$$

and so we can set $K(x,y) = (x-y)^{n-1}/(n-1)! \cdot \chi_{[a,x]}$.

In this case, the operator f'' is self adjoint, so we want $K(x,y)$ to be symmetric in x and y. We also want it to vanish at the boundary, so that the boundary terms vanish. After experimenting, people got

$$K(x,y) = \begin{cases} \frac{(x-a)(b-y)}{a-b} & a \leq x \leq y \leq b \\ \frac{(b-x)(y-a)}{a-b} & a \leq y \leq x \leq b. \end{cases}$$

You can differentiate to actually verify this. This $K(x,y)$ is called the **Green's kernel**.

The Sturm–Liouville equation is solved similarly. Consider the operator

$$Lf = (pf')' - qf.$$

If we can find φ_1 and φ_2 such that

$$\begin{aligned} &L(\varphi_-) = 0, && \varphi_-(a) = 0, && \varphi_-'(a) \neq 0, \\ &L(\varphi_+) = 0, && \varphi_+(b) = 0, && \varphi_+'(b) \neq 0. \end{aligned}$$

Then we can write down Green's kernel as

$$K(x,y) = \begin{cases} \frac{\varphi_-(x)\varphi_+(y)}{w} & a \le x \le y \le b, \\ \frac{\varphi_-(y)\varphi_+(x)}{w} & a \le y \le x \le b. \end{cases}$$

Here, the function

$$w(x) = p(x) \begin{vmatrix} \varphi_-(x) & \varphi_+(x) \\ \varphi_-'(x) & \varphi_+'(x) \end{vmatrix}$$

is actually constant in x.

13.2 Fredholm's alternative

In the toy model, we use $\vec{x}_{\min} = A^*(AA^* + S^*S)^{-1}\vec{b}$. Here, $AA^* + S^*S$ is not going to be nice for general Hilbert spaces, so we were not allowed to use this directly. This is why we needed all the things like $\|T^*g\|^2 + \|Sg\|^2 \ge c\|g\|^2$.

But maybe just inverting the operator works for the equations we care about, i.e., differential equations. The **Fredholm alternative** is the assumption that every $\lambda \neq 0$ is either an eigenvalue or $(T - \lambda I)^{-1}$ is bounded.

If the range is finite dimension, this is fine. If the operator T is compact and self-adjoint, this is still fine, because we have the spectral theorem. We can pick a orthonormal basis $\vec{e}_k$ and write

$$(T - \lambda)^{-1} f = \sum_k \frac{1}{\lambda_k - \lambda} (f, \vec{e}_k)\vec{e}_k.$$

This is bounded by the fact that there are finitely many eigenvalues outside a neighborhood of 0.

Actually the Fredholm alternative holds for all compact operators T. The techniques we will use are the lower bound technique and the iterated approximation to get surjectivity.

Theorem 13.1. *If T is compact and $\lambda \neq 0$ is not an eigenvalue, then $(T-\lambda)^{-1}$ is bounded.*

This actually works for Banach spaces too.

Proof. Let's first prove that $\|(T - \lambda)x\| \ge c\|x\|$. If this is not true, there is a sequence of x_n with $\|x_n\| = 1$ such that $(T - \lambda)x_n \to 0$. Assume that $Tx_n \to y$. Then

$$y \leftarrow Tx_n = \lambda x_n + (T - \lambda)x_n.$$

So $x_n \to y/\lambda$ and $Tx_n \to Ty/\lambda$. So y is an eigenvalue.

Now we need to show that $T - \lambda$ is surjective, so that we can invert it. Let $\|T_n - T\| \to 0$ with T_n finite dimension range. Now the claim is that there exists a $c > 0$ such that

$$\|(T_n - \lambda)x\| \ge c\|x\|$$

for all sufficiently large n. This is by the same technique. If $(s_n - \lambda)x_n \to 0$, then by the diagonal argument we can assume $s_n x_k \to y_n$ as $k \to \infty$ for all n. Also assume $Tx_k \to y$. Then

$$\|y_n - y\| \leq \|y_n - S_n x_n\| + \|S_n x_k - Tx_k\| + \|Tx_k - y\|.$$

Now $(T_n - \lambda)^{-1}$ is defined an is bounded, because T_n has finite-dimensional range. Take an arbitrary y and let $(T_n - \lambda)x_n = y$ with $\|x_n\| \leq C|y|$. Then $(T - \lambda)x_n = y + (T - T_n)x_n$. So

$$\|(T - \lambda)x_n - y\| \leq \|T - T_n\| C \|y\|.$$

Now if $x_n \to x$ then $(T - \lambda)x = y$. □

13.3 Banach spaces

When you solve a differential equation, you only get a weak solution. But now, you would want to know if the solution is differentiable, or even continuous. We can solve an equation like $d^2 f/dx^2 = g$, with g smooth, but can we know that f is smooth? That is, is f the same, up to a measure zero set, as a smooth function? This is known as the Sobolev embedding theorem.

Here, we will look at L^p spaces as well as L^2 spaces. This means that we will have to look at **Banach spaces**, which are vector spaces with a complete metric.

Generally, looking at complements to subspaces is not possible in Banach spaces. This is a big problem. So we only have the technique of iterated approximation to hit some element.

Theorem 13.2 (Baire category theroem)**.** *Let X be a complete metric space. If $\mathcal{O}_n$ are open and dense in X, then*

$$E = \bigcap_{n \in \mathbb{N}} \mathcal{O}_n$$

is dense in X.

Proof. Let U be a nonempty open set. Then each $U \cap O_1$ is nonempty, and so there is a open ball $B_1 \subseteq U \cap O_1$. Then by the same argument there is an open ball $B_2 \subseteq U \cap O_2$, and so forth. Then the intersection of B_n is nonempty by completeness, and so $U \cap E \neq \emptyset$. □

Theorem 13.3 (Open mapping theorem)**.** *Let X, Y be Banach spaces and $T : X \to Y$ be continuous, i.e., bounded. If it is surjective, then the map is open.*

14 October 17, 2017

We were dealing with Banach spaces because we want to look at L_k^p in the regularity problem.

14.1 Useful facts about Banach spaces

Theorem 14.1 (Baire category theorem)**.** *A countable intersection of open dense sets is dense. Equivalently, a countable union of closed nowhere dense sets is nowhere dense.*

Theorem 14.2 (Banach–Steinhaus)**.** *Let X be a Banach space and let $\mathcal{T}$ be a collection of linear operators on X. Suppose that for each $x \in X$, $\sup_{T\in\mathcal{T}} \|Tx\| < \infty$. Then $\sup_{T\in\mathcal{T}} \|T\| < \infty$.*

Proof. For each $n \in \mathbb{N}$, the set

$$F_n = \{x \in X : \sup_{T\in\mathcal{T}} \|Tx\| \leq n\}$$

is closed, and we also have $\bigcup_{n\in\mathbb{N}} F_n = X$. So some F_n contains an open ball. □

Theorem 14.3 (Open mappint theorem)**.** *Let $T : X \to Y$ is a bounded and surjective map of Banach spaces. Then T is open.*

Proof. For B_X the unit open ball, we need to show that $T(B_X)$ contains some open ball in Y. But we have $\bigcup_n T(nB_X) = Y$, and so $\bigcup_n \overline{T(nB_X)} = Y$. The Baire category theorem implies that some $\overline{T(nB_X)}$ contains an open ball in Y. We may assume that $\overline{T(nB_X)} \supseteq B_Y$.

The whole point is to get rid of the closure. The condition we have obtained says that for each $y \in Y$ and $\epsilon > 0$, there exists an $x \in X$ such that

$$\|x\|_X \leq n\|y\|_Y \text{ and } \|y - Tx\|_Y < \epsilon.$$

Then we replace y by $y - Tx$. Then you can approximate $y = Tx_1 + Tx_2 + Tx_3 + \cdots$. □

Theorem 14.4 (Closed graph theorem)**.** *Let $T : X \to Y$ be a linear map such that the graph is closed. Then T is bounded.*

Proof. The space $\Gamma_T \subseteq X \times Y$ is a Banach space. Now the projection $\Gamma_T \to X$ is surjective, so it is open. This shows that any inverse image of an open set in Y to Γ_T is open, and the projection to X is open. □

Theorem 14.5. *$(L^p)^* = L^q$, when $\frac{1}{p} + \frac{1}{q} = 1$.*

The main tool we use is the characterization of the indefinite integral of L^p functions. This was in the assignment.

Proof. We have $L^q \subseteq (L^p)^*$ by Hölder's inequality. The difficult part is $(L^p)^* \subseteq L^q$. Suppose $\Phi : L^p \to \mathbb{R}$ is a linear functional. To determine the function g that represents Φ, it suffices to know its value at $f = [\xi, \eta]$. This goes through the whole process of obtaining measurable sets from open sets, and then using simple functions to approximate measurable functions. Then we can define

$$g = \frac{d}{dx}\Phi(\chi_{[a,x]}).$$

It suffices to show that $F(\chi_{[a,x]})$ is the indefinite integral of a L^q function.

In the homework, you have shown that if F satisfies

$$\sum_{j=1}^{k} \frac{|F(\eta_j) - F(\xi_j)|^q}{(\eta_j - \xi_j)^{q-1}} \leq M$$

for any disjoint $(\xi_1, \eta_1), \ldots, (\xi_k, \eta_k) \subseteq [a, b]$, then F is the indefinite integral of a L^q function. Let

$$f = \sum_{j=1}^{k} \frac{|F(\eta_j) - F(\xi_j)|^{q-1} \operatorname{sgn}(F(\eta_j) - F(\xi_j))}{(\eta_j - \xi_j)^{q-1}} \chi_{[\xi_j, \eta_j]}$$

so that $\Phi(f)$ is the sum we want to estimate. Now if you compute $\|f\|_{L^p}$, it is

$$\|f\|_{L^p} = \left(\sum_{j=1}^{k} \frac{|F(\eta_j) - F(\xi_j)|^q}{|\eta_j - \xi_j|^{q-1}}\right)^{1/p} = (\Phi(f))^{1/p}.$$

But we have $\Phi(f) \leq \|\Phi\| \|f\|_{L^p}$, so we get the right bound. $\square$

14.2 Hahn–Banach theorem

When you do the Fredholm alternative, you prove that $\|(\lambda I - T)x\| \geq c\|x\|$, and you needed to extend the map by using an orthogonal projection. But the thing is that you cannot use orthogonal projection in a Banach space. The Hahn–Banach theorem deals with this situation.

Theorem 14.6 (Hahn–Banach theorem)**.** *Let X be a linear space with a seminorm, over $\mathbb{R}$. (That is, $\|-\| : X \to \mathbb{R}_+ \cup \{0\}$ satisfies $\|\lambda x\| = |\lambda|\|x\|$ and $\|x + y\| \leq \|x\| + \|y\|$.) Given Y a subspace of X, assume that $f : Y \to \mathbb{R}$ is linear and bounded. Then f extends to a linear function on X with the same bound.*

Proof. We extend one dimension at a time, from Y to $Y + \mathbb{R}z$. To do this, you only need to know how to assign $f(z)$. Assume $|f(y)| \leq \|y\|$ for all $y \in Y$ by rescaling. We want to make

$$-\|y + z\| \leq f(y) + f(z) \leq \|y + z\|$$

for all y. That is, we want

$$f(z) \le \|y+z\| - f(y), \quad f(z) \ge -f(y') - \|y'+z\|$$

for all y and y'. This can be done, because

$$f(y-y') \le \|y-y'\| \le \|y+z\| + \|y'+z\|$$

by the triangle inequality. □

We also have a $\mathbb{C}$-linear version. Let f be a $\mathbb{C}$-linear function. We can always write $f = u+iv$, where u and v are $\mathbb{R}$-linear. Then f being $\mathbb{C}$-linear is equivalent to $v(x) = -u(ix)$. That is, given any u, we can define $f(x) = u(x) - iu(ix)$.

Proof of $\mathbb{C}$-linear version. Suppose $|f(x)| \le \|x\|$. Then clearly $|u(x)| \le \|x\|$. We extend $u(x)$ to $U(x)$, and then define $F(x) = U(x) - iU(ix)$. The norm estimate follows automatically. We have

$$|F(x)| = |F(e^{i\alpha}x)| = |U(x)| \le \|x\|.$$ □

Let X be a compact smooth differentiable manifold. Hodge theory is about generalizing the simple prototype to $\omega = Pdx + Qdy$, $d\omega = 0$ and solving $\omega = d\eta$. This is not generally solvable, and our goal is to look at these problems with k-forms.

15 October 19, 2017

Theorem 15.1 (Hahn–Banach). *Let X be a vector space with a semi-norm, and let Y be a subspace with a bounded functional $f : Y \to \mathbb{C}$. Then there exists an extension to X with the same bound.*

Note that a semi-norm is the same thing as a convex set A such that $\bigcup_n nA = X$ and $tA = A$ for $|t| = 1$. Then f bounded by 1 means that A contains the ball B or radius 1. What we are doing is to separating this set with another point by a hyperplane. Using the difference of convex sets $A - B$, you can show that if two convex sets are disjoint, then they can be separated by a hyperplane.

15.1 Calculus on manifolds

To discuss regularity, we need the Sobolev spaces. There is the Gårding's inequality and Rellich's lemma. The motivating example is Hodge theory. This started out with

$$\begin{cases} \frac{\partial u}{\partial x} = P, \\ \frac{\partial u}{\partial t} = Q \end{cases} \qquad \text{subject to } \frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0.$$

This is sort of the fundamental theorem of calculus over $\mathbb{C}$, i.e., Cauchy's theory. But we can go in another direction, which is to interpret it as $d(Pdx+Qdy) = 0$. There is the problem with boundaries, so Hodge looked at compact manifolds without boundary.

We have discussed solving $Tu = f$ subject to $Sf = 0$. But for compact manifolds there is a local question and a global question. The local one is called the Poincaré lemma. We can let $T = d$ on $(p-1)$-forms and $S = d$ on p-forms. This satisfies the estimate, but globally it fails up to finite dimension. These finite dimensional exceptions are called harmonic forms and have important consequences in geometry and topology.

Let us look at the p-forms. If M is a smooth (C^∞) manifold, there is the notion of a **tangent vector**. This is relates to each smooth function its directional derivative. The tangent vector maps the germs of smooth functions to $\mathbb{R}$. This map is $\mathbb{R}$-linear and satisfies the Leibniz formula

$$v(fg) = v(f)g(p) + f(p)v(g).$$

So the space of tangent vectors is a 1-st order approximation to X at P.

Differential forms are then defined. A 0-form is just a function. A 1-form at a point p is an element $(T_{X,p})^*$, and smooth 1-form can be defined. Each 0-form f can be made into a 1-form as

$$(df)_p = \sum_i \Big(\frac{\partial f}{\partial x_i}\Big)_p \Big(\frac{\partial}{\partial x_i}\Big)_p^* = \sum_i \Big(\frac{\partial f}{\partial x_i}\Big)_p (dx_i)_p.$$

What are the 2-forms then? If you just take the derivative, the second order terms will enter. To keep them out, you take the skew-symmetrization. Define

$$d\left(\sum_j \varphi_j x_j\right) = \sum_{j,k} \frac{\partial \varphi_j}{\partial x_k} dx_k \wedge dx_j.$$

15.2 Poincaré lemma

So let us look at the local compatibility condition. Take a smooth p-form

$$\omega = \frac{1}{p!} \sum_{j_1,\dots,j_p} \omega_{j_1,\dots,j_p} dx_{j_1} \wedge \cdots \wedge dx_{j_p},$$

where $\omega_{j_1,\dots,j_p}$ is alternating. Assume that $d\omega = 0$ locally. Then we would like to say that $\omega = d\gamma$ for some γ a local smooth $(p-1)$-form.

Lemma 15.2 (Poincaré lemma)**.** *Let Ω be a star-like domain. If ω is a p-form on Ω such that $d\omega = 0$, then there is a $(p-1)$-form γ on Ω such that $d\gamma = \omega$.*

There are two ways of doing this: one is using polar coordinates and the other is using Cartesian coordinates, one at a time.

Proof. Let us take a radial coordinate t, on a star-like domain Ω. There is a map

$$\Phi : \Omega \times [0,1]; \quad (x,t) \mapsto tx.$$

Write

$$\Phi^*\omega = \alpha + dt \wedge \beta,$$

where dt does not occur in α or β. This is closed, so

$$0 = d\Phi^*\omega = dt \wedge \frac{\partial \alpha}{\partial t} + d_x\alpha - dt \wedge d_x\beta.$$

The coefficient of dt is 0, so $\frac{\partial \alpha}{\partial t} = d_x\beta$. Then

$$\omega - 0 = \alpha|_{t=1} - \alpha|_{t=0} = d_x \int_0^1 \beta dt. \qquad \square$$

If you want to do this on Cartesian coordinates, you use descending induction on the number of $dx_1, \dots, dx_n$ occuring in ω.

15.3 Rellich's lemma

Now let us look at the global situation. Let X be a compact smooth manifold. Let f be a smooth p-form on X with $df = 0$. Let us try to solve $du = f$ for some u a smooth $(p-1)$-form. We want to look at the same thing

$$\|T^*g\|^2 + \|Sg\|^2 \geq ?.$$

This is the same thing as looking at the operator $TT^* + S^*S$. You ask if this allows diagonalization. Here it is easier to use $1 + TT^* + S^*S$.

One way is to use Ascoli–Arzelà. Let us look at the L^2 space and the L_1^2 space. What is the L^2 space of global p-forms? Note that locally p-forms look like linear combinations of $dx_{j_1} \wedge \cdots \wedge dx_{j_p}$ with some coefficients. So we can take the L^2 norm of this after defining the metric and volume form on the manifold. I am going to do this.

Now L_1^2 space consists of measurable functions whose first derivative is measurable in the weak sense. This means that there exists a $g \in L^2$ such that

$$-\int f\varphi' = \int g\varphi$$

for all $\varphi \in C_c^\infty$.

Lemma 15.3. *Let H_s be the L_s^2 space of p-forms, and let H_r be the L_r^2 space of p-forms, with $r > s$. Then $H_r \subset H_s$. If a map*

$$T : H_s \to H_r$$

satisfies $\|Tf\|_r \leq c\|f\|$, then $T : H_s \to H_s$ is compact.

Proof. We can only look at the coefficients. Let ω_n be a bounded sequence in the L_s^2-norm. We want to extract a subsequence of $T\omega_n$ that converges in the L_s-norm. Taking a partition of unity, we may assume that we are working on Euclidean space, and the function has compact support.

We now assume that the chart is $(-1, 1)^n$, and assume that the period is 2π. You look at the Fourier series, and you can show that $\sum |\nu c_\nu|^2$ is bounded. $\square$

16 October 24, 2017

We were developing Hodge theory for compact smooth manifolds. Locally there is no obstruction to every closed p-form being exact. This is Poincaré's lemma.

16.1 L^2 space of differential forms

To make everything rigorous, we first need to define the L^2 space of differential forms. We do this by introducing a Riemannian metric g_{ij}. This is a symmetric 2-form

$$g = \sum_{i,j=1}^{n} g_{ij} dx_i \otimes dx_j$$

where g_{ij} is symmetric and positive-definite. This gives a pointwise inner product, and to make this global, we need a volume form

$$\sqrt{\det g_{ij}}\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n.$$

Now to a p-form

$$\varphi = \frac{1}{p!} \sum \varphi_{j_1 \cdots j_p} dx_{j_1} \wedge \cdots \wedge dx_{j_p},$$

we can define its L^2 norm

$$\|\varphi\|_{L^2}^2 = \sum \int_X \varphi_{j_1 \cdots j_p} \varphi_{k_1 \cdots k_p} g^{j_1 k_1} \cdots g^{j_p k_p} \sqrt{\deg g}\, dx_1 \wedge \cdots \wedge dx_n.$$

In the old days, people tried to solve this locally using Poincaré, but then they don't agree at the boundary. This difference is then going to have zero differential, so can be written as d of something. This motivated Leray to define sheaves.

Let us call $\mathscr{A}^p$ the space of all smooth p-forms on a smooth compact manifold X. Also, let $\mathscr{A}^p_{L^2}$ be the space of all p-forms with locally L^2 coefficients. We have a sequence

$$\cdots \to \mathscr{A}^{p-2} \xrightarrow{d} \mathscr{A}^{p-1} \xrightarrow{d} \mathscr{A}^p \xrightarrow{d} \mathscr{A}^{p+1} \xrightarrow{d} \mathscr{A}^{p+2} \to \cdots.$$

We are then trying to solve the obstructions to $\ker(\mathscr{A}^p \to \mathscr{A}^{p+1}) \supseteq \operatorname{im}(\mathscr{A}^{p-1} \to \mathscr{A}^p)$. But out method of solving this is involving limits, so we are forced to look instead at

$$\mathscr{A}^{p-1}_{L^2} \xrightarrow{d_{p-1}} \mathscr{A}^p_{L^2} \xrightarrow{d_p} \mathscr{A}^{p+1}.$$

It is clear that $d_p \circ d_{p-1} = 0$. So the only thing we need is the a priori estimate. We ask if $d_{p-1} d_{p-1}^* + d_p^* d_p \geq c$ is true. Then we have all our machinery that allows us to solve the equation. But this is not all we want, because we are interested in $\mathscr{A}^p$. So we have another regularity problem to solve.

It turns out that $d_{p-1}d_{p-1}^* + d_p^* d_p \geq c$ up to a finite dimension. That is, there is some finite dimension that behaves badly, and if we exclude the bad guys by taking the orthogonal complement, we get the right inequality.

We will show that $(I + d_{p-1}d_{p-1}^* + d_p^* d_p)^{-1}$ is compact. This will prove that we have the right estimate up to finite dimension, by the spectral theorem. There are two ingredients, and the first one is Gårding's inequality.

Theorem 16.1 (Gårding's inequality)**.** *If $\varphi \in \mathscr{A}^p_{L^2}$ is in $\operatorname{Dom} d_p \cap \operatorname{Dom} d_{p-1}^*$ then*

$$\lVert d_{p-1}^*\varphi\rVert^2_{\mathscr{A}^{p-1}_{L^2}} + \lVert d_p\varphi\rVert^2_{\mathscr{A}^{p+1}_{L^2}} + \lVert\varphi\rVert^2_{\mathscr{A}^p_{L^2}} \geq c\lVert\nabla\varphi\rVert_{\mathscr{A}^p_{L^2}}.$$

Note that this also implies that φ is in L^2_1. The second ingredient is Rellich's lemma.

Lemma 16.2 (Rellich's lemma)**.** *The map $L^2_r \hookrightarrow L^2_s$ is compact for $r > s$.*

Proof. Let φ_ν be a sequence of L^2 p-forms on X where the coefficients are in L^2_1 and are bounded uniformly. We want to select a subsequence that converges in L^2. We can localize by using a partition of unity. Then assume $\operatorname{supp}\varphi_\nu \subseteq (-\pi,\pi)^n$.

Look at the Fourier series and write

$$\varphi_\nu = \sum a_{m_1,\ldots,m_n,\nu} e^{im_1x_1x_1+\cdots+im_nx_n}.$$

Note that g is bounded both above and below, and hence we may just assume g is the standard metric, because we only need convergence. Then up to constant,

$$\lVert\varphi_\nu\rVert^2 = \sum_{m_1,\ldots,m_n} \lvert a_{m_1,\ldots,m_n,\nu}\rvert^2 \leq 1.$$

If we take the derivative, Gårding's inequality will give

$$\left\lVert \frac{\partial}{\partial x_j}\varphi_\nu\right\rVert^2 = \sum \lvert m_j\rvert^2 \lvert a_{m_1,\ldots,m_n,\nu}\rvert^2 \leq 1.$$

This shows that the tail part $\sum_{\lvert m_j\rvert \leq N} \lvert a_{m_1,\ldots,m_n,\nu}\rvert^2$ is bounded by $1/N$, independent of ν. Then we can truncate this tail part, get a finite-dimensional space, and find a subsequence. Now let $N \to \infty$ and use the diagonal. ☐

16.2 Gårding's inequality

The statement is that

$$\lVert d_{p-1}^*\varphi\rVert^2 + \lVert d_p\varphi\rVert^2 + \lVert\varphi\rVert^2 \geq c\lVert\nabla\varphi\rVert^2.$$

Let us look at the case $p = 1$ first. If

$$\varphi = \sum_{j=1}^n \varphi dx_j$$

then

$$d\varphi = \sum_{j,k} \partial_k \varphi_j dx_k \wedge dx_j = \sum_{j<k} (\partial_j \varphi_k - \partial_k \varphi_j) dx_j \wedge dx_k.$$

But what is $d^*\varphi$? Let us work in $\mathbb{R}^n$ for simplicity. If $d^*\varphi = \psi$, then

$$\int \psi u = \sum_j \int \varphi_j \partial_j u = -\int \Big(\sum_j \partial_j \varphi_j \Big) u$$

if everything has compact support. So $d^*\varphi = -\sum_j \partial_j \varphi_j$. Therefore

$$\|d\varphi\|^2 = \sum_{j<k} \int (\partial_j \varphi_k - \partial_k \varphi_j)^2, \quad \|d^*\varphi\|^2 = \int \Big(\sum_j \partial_j \varphi_j \Big)^2.$$

Also

$$\|\nabla \varphi\|^2 = \sum_{j,k} \int (\partial_j \varphi_k)^2.$$

Here, we are adding the sum of squares of linear combinations of partial derivatives of coefficients. This already shows that you can't bound $\|\nabla\varphi\|^2$ by $\|d\varphi\|^2$ and $\|d^*\varphi\|^2$. This is because there are n squares in $\|\nabla\varphi\|^2$ but $n(n-1)/2$ squares in $\|d\varphi\|^2$ and 1 square in $\|d^*\varphi\|^2$.

But this is not linear algebra and we have the squares. The idea is $|\vec{u} \cdot \vec{v}|^2 + |\vec{u} \times \vec{v}|^2 = |\vec{u}|^2 |\vec{v}|^2$. Then

$$|\mathrm{tr}(\vec{u} \otimes \vec{v})|^2 + \|\vec{u} \wedge \vec{v}\|^2 = \|\vec{u} \otimes \vec{v}\|^2$$

and if you let $u = (\partial_1, \dots, \partial_n)$ and $v = (\varphi_1, \dots, \varphi_n)$, then you formally get the inequality.

So formally, pointwise,

$$\begin{aligned} \sum_{j<k} (\partial_j \varphi_k - \partial_k \varphi_j)^2 &= \frac{1}{2} \sum_{j,k} (\partial_j \varphi_k - \partial_k \partial_j)^2 \\ &= \sum_{j,k} (\partial_j \varphi_k)^2 - \sum_{j,k} (\partial_k \varphi_j)(\partial_j \varphi_k). \end{aligned}$$

But when we integrate, we can move ∂ around, so $\int (\partial_k \varphi_j)(\partial_j \varphi_k) = \int (\partial_j \varphi_j)(\partial_k \varphi_k)$.

This is the case for Euclidean space. But we have a Riemannian metric. Here,

$$\|d\varphi\|^2 = \int_X (\partial_j \varphi_k - \partial_k \varphi_j)(\partial_l \varphi_m - \partial_m \varphi_l) g^{jl} g^{km} \sqrt{\det g}.$$

Here, you can mostly ignore $g_{jl} g^{km}$ part because we are working up to constant. But when you do integration by parts, you get up taking the derivative of g. This is why you need an additional $\|\varphi\|^2$ in the general case.

What about for p-forms? Let me explain a bit about covariant differentiation in this case. Because difference quotient depends on the coordinate, you need a connection on the manifold. Levi-Cevita gave a way to get a connection from a Riemannian metric.

17 October 26, 2017

We were looking at p-forms and $\mathscr{A}^p_{L^2}(X) \to \mathscr{A}^{p+1}_{L^2}(X)$. Our actual goal is $\mathscr{A}^p_{C^\infty}(X) \to \mathscr{A}^{p+1}_{C^\infty}(X)$. The two ingredients are

Lemma 17.1 (Rellich's lemma)**.** *For $r > s$, the inclusion $L^2_r(X) \to L^2_s(X)$ is compact.*

Proof. You localize, and then use Fourier series. $\square$

Theorem 17.2 (Gårding's inequality)**.** *If $\varphi \in \mathscr{A}^p_{C^\infty}(X)$ then*

$$\|d^*_{p-1}\varphi\|^2 + \|d_p\varphi\|^2 + \|\varphi\|^2 \geq c\|\nabla\varphi\|^2.$$

Here ∇ is covariant differentiation which can be defined by the Riemannian metric.

Theorem 17.3 (Gårding's inequality, L^2-version)**.** *If $\varphi \in \mathscr{A}^p_{L^2}(X)$ in $\operatorname{Dom} d_p \cap \operatorname{Dom} d^*_{p-1}$ then*

$$\|d^*_{p-1}\varphi\|^2 + \|d_p\varphi\|^2 + \|\varphi\|^2 \geq c\|\nabla\varphi\|^2.$$

In particular, $\partial_k\varphi_{j_1,\dots,j_p}$ is L^2.

This is a bit complicated, and we need to use Fredrichs' lemma, proven in 1944. This says that if L is a differential operator, and χ_ϵ is kernel with $\operatorname{supp}\chi_\epsilon$ size of order ϵ, then $u * \chi_\epsilon \to u$ and $L(u\chi_\epsilon) \to Lu$ in L^2.

Proof. First localize, so that we have a compactly supported φ. Then Fredrich's lemma tells us that

$$\varphi * \chi_\epsilon \to \varphi, \quad d_p(\varphi * \chi_\epsilon) \to d_p\varphi, \quad d^*_{p-1}(\varphi * \chi_\epsilon) \to d^*_{p-1}\varphi.$$

Then we use the C^∞ version Gårding's inequality. To take care of $\nabla\varphi$, use Fatou. $\square$

If you want to solve the problem with boundary, then things get easily complicated. You have done the D^2 case in the homework.

17.1 Solving the equation

Now we want to solve the differential equation. The spectral theorem will be applied to $(I + d^*_p d_p + d_{p-1}d^*_{p-1})^{-1}$. Now you may object that this $d^*_p d_p$ or maybe their sum are not defined.

This is Friedrichs' contribution. Let $T : H \to H$ be a closed densely defined operator. He showed that given any $u \in H$, there exists a v such that $(1 + T^*T)v = u$. This means both that Tv is defined and then T^*Tv is defined.

In our case, we are looking at the map

$$\mathscr{A}^p_{L^2}(X) \to \mathscr{A}^{p-1}_{L^2}(X) \oplus \mathscr{A}^{p+1}_{L^2}(X); \quad \varphi \mapsto d^*_{p-1}\varphi \oplus d_p\varphi.$$

In this case, we have $T : H_1 \to H_2$, and we make it into

$$S = \begin{pmatrix} 0 & 0 \\ T & 0 \end{pmatrix} : H_1 \oplus H_2 \to H_1 \oplus H_2.$$

If we apply the theorem, we just inverse $(1+SS^*)^{-1}$, which is just $(1+TT^*)^{-1}$. So we just don't have to worry about these double domain containment conditions.

Given any u_ν a bounded sequence, we have v_ν such that

$$(1 + d_p^* d_p + d_{p-1} d_{p-1}^*) v_\nu = u_\nu.$$

Here, $v_\nu \in \operatorname{Dom} d_p \cap \operatorname{Dom} d_{p-1}^*$. This sequence v_ν, by Gårding's inequality, is bounded in L_1^2. So by Rellich there exists a subsequence converging in L^2. This shows that $(1 + d_p^* d_p + d_{p-1} d_{p-1}^*)^{-1}$ is compact.

Applying the spectral theorem gives eigenvalues $\lambda_n \geq 0$ converging to 0 as $n \to \infty$, with eigenspaces E_n finite-dimensional except for $\lambda_n = 0$. This is for $(1 + d_p^* d_p + d_{p-1} d_{p-1}^*)^{-1}$. Now if we look at the map $d_p^* d_p + d_{p-1} d_{p-1}^*$, the same eigenspaces E_n are eigenspaces with eigenvalue

$$\mu_n = \frac{1}{\lambda_n} - 1.$$

Because $0 < \lambda_n \leq 1$ with $\lambda_n \to 0$, we get $0 \leq \mu_n$ with $\mu_n \to \infty$. The only problem is when $\lambda_n = 1$ and $\mu_n = 0$. This is a problem because we want to solve the differential equation and so we want to invert it.

So let H be the eigenspace for

$$d_p^* d_p + d_{p-1} d_{p-1}^* = \Delta$$

with eigenvalue 0. We can decompose $\mathscr{A}_{L^2}^p(X) = H \oplus H^\perp$. Then we can invert

$$(\Delta|_{H^\perp})^{-1} \vec{e}_j = \frac{1}{\mu(\vec{e}_j)} \vec{e}_j,$$

which has bound $1/\mu'$ where μ' is the minimal nonzero μ_n. We can extend $(\Delta|_{H^\perp})^{-1}$ to 0 on H. This operator G is called the **Green's operator**.

Now we are looking at

$$\mathscr{A}_{L^2}^{p-1} \xrightarrow{d_{p-1}} \mathscr{A}_{L^2}^{p} \xrightarrow{d_p} \mathscr{A}_{L^2}^{p+1}.$$

Given any $\varphi \in \ker d_p$, we want to obtain the minimal solution so that

$$\psi = d_{p-1}^* (d_{p-1} d_{p-1}^* + d_p^* d_p)^{-1} \varphi.$$

But of course Δ is not invertible because of H. So the best we can do is to solve it in the case $\varphi \in H^\perp$, and then we can just write

$$\psi = d_{p-1}^* G \varphi.$$

17.2 Regularity of harmonic forms

But I haven't done anything about regularity although I have advertised it a lot. We want to show that if φ is smooth then the solution φ is also smooth. Here we are going to use the fact that

$$d_{p-1}\psi = \varphi, \quad d^*_{p-2}\psi = 0.$$

Proposition 17.4. *If $f = d_p\varphi$ and $g = d^*_{p-1}\varphi$ are both C^∞, then φ is C^∞.*

The technique is to use the commutators of differential operators with difference operators.

Proof. First you localize. Here this is a problem because

$$d(\rho_\nu\varphi) = \rho_\nu d\varphi + E,$$

where E depends on φ linearly. So $d_p\varphi = f + E$ is no longer smooth but is just in L^2. This implies that

$$d_p\varphi = f + E, \quad d^*_{p-1}\varphi = g + E'$$

for some $E, E' \in L^2$. By Gårding's inequality, we immediately get $\varphi \in L^2_1$. Then we look at an arbitrary partial derivative D. Here

$$Dd_p\varphi = d_pD\varphi + [D, d_p]\varphi.$$

This $[D, d_p]$ is first-order 1 by the product formula. Because $Dd_p\varphi$ and $[D, d_p]$ is L^2, we get we get $d_pD\varphi \in L^2$. Then likewise we get $d^*_{p-1}D\varphi \in L^2$. Again by Gårding, we get $D\varphi \in L^2_1$. Then we can bootstrap this iteratively to get $\varphi \in C^\infty$.

Here, we don't know if ∇ is in L^2. So we use a difference quotient instead of a partial derivative, and apply Gårding's inequality. Then we can take the limit using Fatou's lemma. This problem here because we really can't to integration by parts because everything is a weak derivative. $\square$

18 October 31, 2017

I think we finished Hodge theory. Now I will go and do the other technique of solving differential equations. This was developed by Malgrange and Ehrenspreis. We are going to use Fourier transform to solve linear partial differential equations with constant coefficients. The point is to overcome the difficulty of division by 0.

The point is that the Fourier transform is complex-analytic. We have

$$\hat{f}(\xi) = \int_{x \in \mathbb{R}} f(x) e^{-2\pi i \xi x} dx.$$

This is the weighted average of the exponential function. So after taking the weighted average, it is still going to be complex-analytic. Then we have the mean-value property. Even though f might have a zero, it might not have a zero on a circle.

We want to solve differential equations in a weak sense. So we want to use test functions. We can use the compactly supported smooth functions C_c^∞. But using the Schwartz space is good enough. These are the functions whose $D^\alpha f$ decay faster than any polynomial order.

When people tried to solve Laplace's equation

$$\sum_{j=1}^{n} \frac{\partial^2 u}{\partial x_j^2} = f,$$

they took the Fourier transform. Here,

$$\widehat{f'}(\xi) = \int_{\mathbb{R}} f'(x) e^{-2\pi i \xi x} dx = (2\pi i \xi) \int_{\mathbb{R}} f'(x) e^{-2\pi i \xi x} dx.$$

So we have

$$\widehat{\Big(\frac{\partial}{\partial x}\Big)^\alpha} f(\xi) = (2\pi i \xi)^\alpha \hat{f}(\xi).$$

So we would have

$$\hat{u} = \frac{\hat{f}}{-4\pi^2 |\xi|^2}.$$

Here you have a problem because you're dividing by zero.

18.1 Smearing out by polynomials

This is the key lemma. We want to write $F(z) = G(z)/P(z)$.

Lemma 18.1. *Let us write $P(z) = z^m + \sum_{k=0}^{m-1} b_k z^k$. If the function $F(z)$ is holomorphic on $|z| \leq 1$, then*

$$|F(0)|^2 \leq \frac{1}{2\pi} \int_{\theta=0}^{2\pi} |P(e^{i\theta}) F(e^{i\theta})|^2 d\theta.$$

Proof. The idea is to reflect the zeros outside the unit circle. Let us write $P(z) = P_1(z)P_2(z)$ where the roots of $P_1(z)$ are outside the unit disk and those of $P_2(z)$ are inside the unit disk. For $|\beta| < 1$, there is a Möbius transformation

$$z \mapsto \frac{z-\beta}{1-\bar{\beta}z}$$

which maps the circle to itself. The polynomial P_2 is the problem, so we reflect to define

$$\tilde{P}_2(z) = \prod_{|\beta|<1, P(\beta)=0} (1-\bar{\beta}z).$$

So if we let $\tilde{P}(z) = P_1(z)\tilde{P}_2(z)$ then all roots of $\tilde{P}$ are outside the unit disk and

$$|\tilde{P}(0)| = \prod_{\alpha\geq 1, P(\zeta)=0} |\alpha| \geq 1, \quad |P(e^{i\theta})F(e^{i\theta})| = |\tilde{P}(e^{i\theta})F(e^{i\theta})|.$$

But the mean-value property with Cauchy–Schwartz imply

$$|H(0)|^2 \leq \frac{1}{2\pi}\int_{\theta=0}^{2\pi} |H(e^{i\theta})|^2 d\theta.$$

Applying it to $\tilde{P}F$ gives

$$|\tilde{P}(0)F(0)|^2 \leq \frac{1}{2\pi}\int_{\theta=0}^{2\pi} |\tilde{P}(e^{i\theta})F(e^{i\theta})|^2 d\theta.$$

Then we are done. $\square$

18.2 Differential equations with constant coefficients

Suppose we want to solve $Lu = f$, so that taking the Fourier transform gives $\widehat{Lu}(\xi) = Q(\xi)\hat{u}(\xi)$. We want to replace by Riesz representation theorem and estimate.

Let us first look at the formal adjoint of

$$L = \sum_\alpha a_\alpha \Big(\frac{\partial}{\partial x}\Big)^\alpha,$$

which is

$$L^* = \sum_\alpha (-1)^{|\alpha|}\bar{a}_\alpha \Big(\frac{\partial}{\partial x}\Big)^\alpha.$$

We assume for now that $\Omega \subseteq \mathbb{R}^n$ is a bounded domain where we are solving the equation.

So it suffices to find a $c > 0$ such that

$$\|\psi\|_{L^2(\Omega)} \leq c\|L^*\psi\|_{L^2(\Omega)}.$$

If this is the case, we will be able to use Riesz representation to show that for all $f \in L^2(\Omega)$ there exists an $u \in L^2(\Omega)$ such that $Lu = f$ and $\|u\|_{L^2(\Omega)} \leq c\|f\|_{L^2(\Omega)}$.

How do we go from L^2 to holomorphic? Assume that $\Omega = (-M, M)$ with $M > 0$. Assume g is $L^2(\mathbb{R})$ with $\operatorname{supp} g \subseteq \Omega$. Then I claim that

$$\hat{g}(\xi + i\eta) = \int_{-M}^{M} e^{-2\pi i(\xi+i\eta)x} dx$$

is holomorphic on $\mathbb{C}$. To show this, we need to justify passing the differential operator past g. This can be justified by using the fundamental theorem of calculus and Fubini's theorem.

Now we want to check that $\|\psi\|_{L^2(\Omega)} \leq c\|L^*\psi\|_{L^2(\Omega)}$. We do this by using the Fourier transform. This is done by using the fact that $\hat{}$ preserves the L^2 norm. It is also done in the context of the Schwartz space $\mathscr{S}$. Let $\Omega \subseteq \mathbb{R}^{n+1}$ with coordinates $(x, y_1, \ldots, y_m)$. Assume that L takes the form of

$$L = B\frac{\partial^m}{\partial x^m} + \sum_{v=0}^{m-1} L_v \frac{\partial^v}{\partial x^v}, \quad L_v = \sum_{\lambda_1,\ldots,\lambda_n} a_{\lambda_1,\ldots,\lambda_n} \frac{\partial^{\lambda_1+\cdots+\lambda_n}}{\partial y_1^{\lambda_1} \cdots \partial y_n^{\lambda_n}}.$$

We now check that

$$\|\psi\|_{L^2(\Omega)} \leq c\|L^*\psi\|_{L^2(\Omega)}$$

for all $\psi \in \mathscr{S}$. We can extend by zero and instead show

$$\|\hat{\psi}\|_{L^2(\mathbb{R}^{n+1})} \leq c\|Q(\xi)\widehat{\psi}\|_{L^2(\mathbb{R}^{n+1})}.$$

But note that this is exactly in the situation of the key lemma. We need

$$\int_{\xi,\sigma_1,\ldots,\sigma_n} |Q(\xi + i\eta, \sigma_1, \ldots, \sigma_k)\hat{\psi}(\xi + i\eta, \sigma_1, \ldots, \sigma_n)|^2$$

at $\eta = 0$ to dominate some constant times $\int |\hat{\psi}(\xi, \sigma_1, \ldots, \sigma_n)|^2$. Let $z = \xi + re^{i\theta}$. Then the key lemma gives

$$|\hat{\psi}(\xi, \sigma_1, \ldots, \sigma_n)|^2 \leq C\frac{1}{2\pi}\int_{\theta=0}^{2\pi} |(Q\hat{\psi})(\xi + \cos\theta + i\sin\theta, \sigma_1, \ldots, \sigma_n)|^2 d\theta.$$

So we have

$$|\hat{\psi}(\xi, \sigma_1, \ldots, \sigma_n)|^2 \leq \frac{C}{2\pi}\int_{\theta=0}^{2\pi} |\widehat{L^*\psi}(\xi + \cos\theta + i\sin\theta, \sigma_1, \ldots, \sigma_n)|^2 d\theta.$$

But

$$\widehat{L^*\psi}(\xi + \cos\theta + i\sin\theta, \sigma_1, \ldots, \sigma_n) = \widehat{L^*\psi}(\xi, \sigma_1, \ldots, \sigma_n)e^{-2\pi i(\xi+\cos\theta+i\sin\theta)x}.$$

Because $\Omega \subseteq [-M, M] \times \mathbb{R}^n$, the $e^{-2\pi i(\xi+\cos\theta+i\sin\theta)x}$ is bounded. This shows that we have our inequality.

19 November 2, 2017

Last time we looked at the method of Malgrange–Ehrenpreis. Here, we proved the estimate

$$\|u\|_{L^2(\Omega)} \leq c\|L^* u\|_{L^2(\Omega)}$$

for all $u \in C_0^\infty(\Omega)$. Then we were able to conclude that L is surjective and so $Lu = f$ is always solvable.

Note that this is not a special case of the a priori estimate

$$c^2(\|T^* g\|^2 + \|Sg\|^2) \geq \|g\|^2.$$

The a priori estimate needs to hold for all $g \in \operatorname{Dom} S \cap \operatorname{Dom} T^*$. But we have only shown this for g a compactly supported smooth function.

When we say that $Lu = f$ in the weak sense, we are simply saying that $(L^* g, u) = (g, f)$ for all $g \in C_0^\infty(\Omega)$. This is not saying that u is in the domain of L and $Lu = f$. To do this, we need some Friedrich thing and we have a problem for higher order derivatives.

19.1 Distributions

Definition 19.1. A **distribution** or a **generalized function** is a continuous linear functional on $C_0(\Omega) = \mathcal{D}(\Omega)$. The space is called $\mathcal{D}'(\Omega)$.

But what does continuous mean? This should mean that $\varphi_n \to 0$ implies $T(\varphi_n) \to 0$.

Definition 19.2. We introduce a metric such that $\varphi_n \to 0$ means that there exists a compact K such that $\operatorname{supp} \varphi_n \subseteq K$ for all n, and $D^\alpha \varphi_n \to 0$ uniformly for all α.

So when $u \in L^2$, the function Lu makes sense as a distribution. We define it as

$$(Lu, \varphi) = (u, L^* \varphi).$$

This is going to be a continuous functional.

This also comes up when you want Ω unbounded in the Malgrange–Ehrenpreis. If $\Omega \subseteq [-M, M] \times \mathbb{R}^n$ all is good because the Fourier transform is bounded, but once Ω is too big we get into trouble. Then we are forced to use distributions. Using this setting, we can use language like

$$\Delta\Big(c_n \frac{1}{|x|^{n-2}}\Big) = \delta(x).$$

Here, we want to have Fourier analysis in the setting of distributions. We would have, for all $\varphi \in \mathcal{D}(\mathbb{R})$,

$$\int_\xi \hat{f}(\xi)\varphi(\xi) = \int_\xi \int_x f(x) e^{-2\pi i x\xi} \varphi(\xi) d\xi dx.$$

So we would like to say $(\hat{f})(\varphi) = f(\hat{\varphi})$. But here we have trouble because $\hat{\varphi}$ doesn't have compact support. I will talk about this more, but you can justify this.

Let us write $E = c_n/|x|^{n-2}$ for the fundamental solution. Then

$$\hat{E}(\xi) = -\frac{1}{4\pi|\xi|^2}.$$

We want to say something like this, but this can't be justified even if we know the answer.

19.2 Tempered distributions

We had this problem about $\hat{\varphi}$ not being compactly supported even if φ is compactly supported. This is a problem when defining $\hat{T}$, so we are going to change our definition of a distribution.

Definition 19.3. The **Schwartz space** $\mathcal{S}(\mathbb{R}^n)$ is the set of all functions $C^\infty(\mathbb{R}^n)$ such that

$$\sup_{x\in\mathbb{R}^n} |x|^a |D^\alpha f(x)| < \infty$$

for all a, α. A **tempered distribution** is a distribution that extends to a continuous functional on $\mathcal{S}(\mathbb{R}^n)$.

So far we've been looking at solving differential equations. If we directly take the Fourier transform, then this is bad because we get division by zero. The other thing we've been doing is to use Riesz representation. Here, we estimate the L^2 norm using the fact that the Fourier transform fixes the L^2 norm.

Because we can't directly apply the Fourier transform, we are going to take some nice properties.

(0) Df corresponds to $\hat{f}$ times a polynomial.

(1) $\widehat{f * g} = \hat{f} \cdot \hat{g}$.

(2) $\int \hat{f} g = \int f \hat{g}$.

(3) $\sum_n f(n) = \sum_n \hat{f}(n)$.

Recall that Plancherel follows from (1). If we let $g(x) = \overline{f(-x)}$ and evaluate $f * g$ at 0, then

$$f * g(0) = \int \widehat{(f * g)}(\xi) d\xi = \int |\hat{f}|^2 d\xi.$$

The inversion formula follows from $e^{-\pi x^2}$ being its own Fourier transform and the identity (2).

Theorem 19.4. *Let $P(D)$ be a (constant coefficient) polynomial in the partial differential operators on $\mathbb{R}^n$. There exists a fundamental solution E for $P(D)$, which means that E is a (tempered) distribution on $\mathbb{R}^n$ with*

$$P(D)E = \delta_0$$

is the Dirac delta at 0.

We can basically repeat the same argument. But we are going to refine the trick of Malgrange–Ehrenpreis. Recall that we were only using this trick when $\Omega \subseteq [-M, M] \times \mathbb{R}^n$. Also, there was this one direct x such that the highest term is just $(\partial/\partial x)^m$. This is not true anymore, so we are going to average over all he special one coordinates. That is, we let $z_j = e^{i\theta_j}$ and average over the torus T^n.

Let us write $P = P_0 + P_1 + \cdots + P_N$ where P_ν is homogeneous of degree ν. We are going to use the mean-value property of a holomorphic function. For f a holomorphic function on $\mathbb{C}^n$, consider

$$F(\lambda) = f(z + r\lambda w)$$

for fixed center $z \in \mathbb{C}^n$ and $w \in T^n$. The trick of Malgrange–Ehrenpreis then gives

$$r^N |P_N(w)||f(w)| \leq \frac{1}{2\pi} \int_{\theta=0}^{2\pi} |(fP)(z + re^{i\theta}w)| d\theta$$

because Q has leading coefficient $r^N P(w)$. If average, we get

$$|f(z)| \leq \frac{A}{r^N} \int_{w \in T^n} |(fP)(z+w)| \frac{d\theta_1}{2\pi} \cdots \frac{d\theta_n}{2\pi},$$

where

$$\frac{1}{A} = \int_{w \in T^n} |P_N(w)| \frac{d\theta_1}{2\pi} \cdots \frac{d\theta_n}{2\pi}.$$

20 November 7, 2017

We have looked at Malgrange–Ehrenpreis, and here we need the monic assumption and also bounded domain $\Omega \subseteq [-M, M] \times \mathbb{R}^n$. To get rid of the boundedness assumption in one coordinate, we were forced to look at distributions.

Distributions is the dual of the space of test functions. The test functions on compact sets form a Fréchet spaces, and so the space $\mathcal{D}(\mathbb{R}^n)$ of all test functions is a direct limit of Fréchet space. Then the dual is going to be the distributions. Duals of Fréchet spaces are DF spaces, but we are looking at the dual of a direct limit of Fréchet space.

Other than Malgrange–Ehrenpreis, there is a way of solving a differential equation with probability. When I was at Yale, I once heard this ingenious idea from Kakutani and was very impressed. Suppose you want to solve Dirichlet's problem with some boundary condition. What you can do here is to find a value in the middle is to consider a random walk starting at this point. It is going to hit the boundary at some finite time, and look at the expected value of the function value of the hitting point. This enjoys the harmonic property because it has to first go somewhere, and then the expected value is the expected value of starting at the nearby points.

20.1 Malgrange–Ehrenpreis on unbounded region

We remove the monic assumption. Let $P(z_1, \dots, z_n)$ be a homogeneous polynomial of degree m. Then leading coefficient is

$$C = \left(\frac{1}{(2\pi)^n} \int_{\theta_j=0}^{2\pi} |P_m(e^{i\theta_1}, \dots, e^{i\theta_n})| d\theta_1 \cdots d\theta_n \right)^{-1}$$

and then we have the inequality

$$|F(0)| \leq \frac{1}{(2\pi)^n} \int_{\theta_j=0}^{2\pi} |(FP)(e^{i\theta_1}, \dots, e^{i\theta_n})| d\theta_1 \cdots d\theta_n.$$

After you do this, we want to repeat the same technique. Instead of solving $Lu = f$, we solve $LE = \delta_0$ and then we would get $u = E * f$. This E is called a fundamental solution, and is called a tempered distribution. Note that for a distribution T, we are going to define

$$\Big(\frac{\partial}{\partial x_j} T \Big) \varphi = T\Big(-\frac{\partial \varphi}{\partial x_j} \Big),$$

because we want integration by parts. So if

$$L = \sum_{|\alpha| \leq m} a_\alpha \partial_x^\alpha,$$

then

$$L_1 = \sum_{|\alpha| \leq m} (-1)^{|\alpha|} a_\alpha \partial_x^\alpha.$$

Using this we can write $(LE)\varphi = E(L_1\varphi)$, so we want this E to send $L_1\varphi$ to $\varphi(0)$.

Now it suffices to show that $L_1\varphi \mapsto \varphi(0)$ defined on $L_1\mathcal{S}(\mathbb{R}^n)$ is bounded. Then we have Hahn–Banach and so extend to $\mathcal{S}(\mathbb{R}^n)$ with the same bound.

Note that on $\mathcal{S}(\mathbb{R}^n)$, the Fourier transform is a isomorphism of topological vector spaces. Let Q be the symbol of the characteristic polynomial of L_1. Then we want a bound for

$$Q(\xi)\hat{\varphi}(\xi) \mapsto \varphi(0).$$

That is, we want to show that

$$|\varphi(0)| \le C_\ell \|Q(\xi)\hat{\varphi}(\xi)\|_{\ell,\mathcal{S}}.$$

We apply Malgrange–Ehrenpreis to $\eta \mapsto Q(\eta+\xi)\hat{\varphi}(\eta+\xi)$ for fixed ξ. Then we will get

$$|\hat{\varphi}(\xi)| \le C\frac{1}{(2\pi)^n}\int_{\theta_j=0}^{2\pi} |(Q\hat{\varphi})(\xi+e^{i\theta_1},\dots,\xi+e^{i\theta_n})|d\theta_1\cdots d\theta_n.$$

Then

$$\begin{aligned}
|\varphi(0)| &\le \int_{\xi\in\mathbb{R}^n} |\hat{\varphi}(\xi)| \\
&\le \frac{C}{(2\pi)^n}\int_{\theta_j=0}^{2\pi}\int_{\xi\in\mathbb{R}^n} |(Q\hat{\varphi})(\xi+e^{i\theta_1},\dots,\xi+e^{i\theta_n})|d\xi d\theta_1\cdots d\theta_n \\
&= C\int_{\xi\in\mathbb{R}^n} \|(Q\hat{\varphi})(\xi)\| \\
&= C\int_{\xi\in\mathbb{R}^n} \frac{1}{1+|\xi|^{n+1}}(1+|\xi|^{n+1})|Q\hat{\varphi}(\xi)| \le C'\|\varphi\|_{\ell,\mathcal{S}},
\end{aligned}$$

because we have rapid decay. Here you need to be a bit careful because we are shifting in the complex direction, but this can be done.

There is another way of handling this problem. If we have $LE = \delta_0$, then we have $E(L_1\varphi) = \varphi(0)$. Then

$$\varphi(0) = \int \hat{\varphi}(\xi) = \int \frac{Q\hat{\varphi}}{Q}.$$

If Q has no zero, we are fine, but if Q has a zero, we can't just do this. So what we can do is to shift the domain. Instead of ξ_n, we are going to use $\xi_n + i\tilde{\eta}(\xi_1,\dots,\xi_{n-1})$, where $\tilde{\eta}$ depends on $\xi_1,\dots,\xi_{n-1}$. Then define

$$E(\varphi) = \int_{\xi_1,\dots,\xi_{n-1}}\int_{\xi_n\in\mathbb{R}}\Big(\frac{\hat{\varphi}}{Q}\Big)(\xi_1,\dots,\xi_{n-1},\xi_n+i\tilde{\eta}(\xi_1,\dots,\xi_{n-1})).$$

If we can choose $\tilde{\eta}$ so that $Q \ge c > 0$, we would have

$$\begin{aligned}
E(L_1\varphi) &= \int_{\xi_1,\dots,\xi_n} \hat{\varphi}(\xi_1,\dots,\xi_{n-1},\xi_n+i\tilde{\eta}(\xi_1,\dots,\xi_{n-1})) \\
&= \int_{\xi_1,\dots,\xi_n} \hat{\varphi}(\xi_1,\dots,\xi_n) = \varphi(0)
\end{aligned}$$

because the Fourier transform is holomorphic. We still need to be talk about how to choose $\tilde{\eta}$, but this is the most elegant and explicit way to solve the problem.

After this, we want to know whether if the fundamental solution is smooth, outside the origin. This is mostly in done in the case of elliptic operators, e.g., the Laplacian.

21 November 9, 2017

People were looking at simple linear PDEs with constant coefficients. Then where L is is a differential equation, we want to solve $LE = \delta_0$. Here $\mathcal{D}(\mathbb{R}^n)$ is the direct limit of the Fréchet space $\mathcal{D}_K(\mathbb{R})$. This is not nice, so we look at the Schwartz space $\mathcal{S}(\mathbb{R}^n)$, which is a Fréchet space. Another nice thing is that $\hat{}$ maps $\mathcal{S}(\mathbb{R}^n)$ to itself, and $\mathcal{S}(\mathbb{R}^n)'$ is a tempered distribution.

The fundamental solution was obtained first by Malgrange–Ehrenpreis by reflection with the unit circle, and next by extension of a linear functional. We wanted to take the Fourier transform and say $Q\hat{E} = 1$, but then we had this problem of dividing by zero. What we want is

$$E(L_1\varphi) \to \varphi(0)$$

and then extend it to $E : \mathcal{D}(\mathbb{R}^n) \to \mathbb{C}$. So we check that E on $L_1(\mathcal{D}(\mathbb{R}^n))$ is already continuous. Then we want to check that for fixed K,

$$|\varphi(0)| \leq C_K \|L_1\varphi\|_{\ell_K,K}.$$

The Malgrange–Ehrenpreis trick is

$$|\hat{\varphi}(\xi)| \leq \frac{C}{(2\pi)^n} \int_{0\leq\theta_j\leq 2\pi} |(Q\hat{\varphi})(\xi_1 + e^{i\theta_1}, \dots, \xi_n + e^{i\theta_n}| d\theta_1 \cdots d\theta_n.$$

Then

$$\begin{aligned} |\varphi(0)| &= \left| \int_{\xi\in\mathbb{R}^n} \hat{\varphi}(\xi) \right| \leq \int_{\xi\in\mathbb{R}^n} |\hat{\varphi}(\xi)| \\ &\leq \frac{C}{(2\pi)^n} \int_{0\leq\theta_j\leq 2\pi} \int_{\xi\in\mathbb{R}^n} |(Q\hat{\varphi})(\xi_1 + e^{i\theta_1}, \dots, \xi_n + e^{i\theta_n})| d\theta_1 \cdots d\theta_n. \end{aligned}$$

Here

$$(Q\hat{\varphi})(\xi_j + e^{i\theta_j}) = \int_{x\in K} (L_1\varphi)(x) e^{-2\pi i \sum_\nu x_\nu(\xi_\nu + \cos\theta_n + i\sin\theta_n)}$$

and so

$$|(Q\hat{\theta})(\xi_j + e^{i\theta_j})| \leq \int_{x\in K} |(L_1\varphi)(x)| e^{2\pi \sum_\nu |\xi_\nu|} \leq C_K'' \sup_{x\in K} (L_1\varphi)(x).$$

But this is not good enough because we have to integrate over $\xi \in \mathbb{R}^n$. So we need another bound. Consider L_2 the linear PDE with constant coefficients the symbol $(1+\xi^2)^\ell$. Then the same bound gives us

$$|(1 + \textstyle\sum_{\nu=1}^n (\xi_\nu + e^{i\theta_\nu})^2)^\ell (Q\hat{\varphi})(\xi_j + e^{i\theta_j})| \leq C_K'' \sup_{x\in K} |(L_2L_1\varphi)(x)|.$$

Then some technical inequality gives

$$|(Q\hat{\varphi})(\xi_j + e^{i\theta_j})| \leq C_K'' \frac{8\ell}{(1+\xi^2)^\ell} \sup_{x\in K} |(L_2L_1\varphi)(x)|.$$

Now we can integrate over ξ and this gives

$$|\varphi(0)| \leq C_K''' \sup_{x\in K} |(L_2L_1\varphi)(x)|.$$

21.1 Division problem for distributions

The thing about tempered distributions is that $\hat{-}$ maps $\mathcal{S}(\mathbb{R}^n)$ to $\mathcal{S}(\mathbb{R}^n)$. So $\hat{-}$ also maps $\mathcal{S}(\mathbb{R}^n)'$ to $\mathcal{S}(\mathbb{R}^n)'$.

If we want to solve $LE = \delta_0$, the question is whether you can divide a tempered distribution by a polynomial. In particular, is multiplication by a polynomial Q surjective in $\mathcal{S}(\mathbb{R}^n)'$? If you go to the dual, we are asking if multiplication by Q is injective with closed image in $\mathcal{S}(\mathbb{R}^n)$.

Our first goal is to show that multiplication by Q on $\mathcal{S}(\mathbb{R}^n)$ is injective, and the second goal is to show that this implies that multiplication by Q in $\mathcal{S}(\mathbb{R}^n)'$ is surjective.

Proposition 21.1. *For a nonzero polynomial Q, multiplication by Q on $\mathcal{S}(\mathbb{R}^n)$ is injective and has closed image.*

Injectivity is no trouble, because Q vanishes on a very small set. The problem is to check that the image is closed. This amounts to bounding the norm of ψ/Q by the norm of ψ.

Let's look at the simple case of a single variable. How can you bound

$$(1+x^2)^\alpha \partial_x^\beta \left(\frac{\psi}{(x-a_1)\cdots(x-a_n)} \right)$$

in terms of $(1+x^2)^\alpha \partial_x^\beta \psi(x)$. This is going to be L'Hôpital's rule. But what about in higher dimensions? It took people something like 4 years to deal with this. We are not going to do this, but there is a way of doing this, due to Whitney. Basically you assign to each point some jet with some proximity condition.

21.2 Changing the domain of integration

In Malgrange–Ehrenpreis, we replace $\mathbb{R} \times \mathbb{R}^{n-1}$ by $S^1 \times \mathbb{R} \times \mathbb{R}^{n-1}$. But here we replace it by $(\mathbb{R} + i\tilde{\eta}) \times \mathbb{R}^{n-1}$ in $\mathbb{C} \times \mathbb{R}^{n-1}$. If we move it well, we get an explicit formula for E.

$$LE(\varphi) = E(L_1\varphi) = \varphi(0).$$

If Q is bounded from below, we would get

$$E(\varphi) = \int_{\mathbb{R}^n} \frac{\hat{\varphi}}{Q}(\xi)$$

and this would be an answer. This is good already good for tempered distributions. This is because we would have

$$E(L_1\varphi) = \int_{\mathbb{R}^n} \frac{\widehat{L_1\varphi}}{Q} = \int_{\mathbb{R}^n} \hat{\varphi} = \varphi(0).$$

For instance, if $L = (1-\Delta)^N$ then $Q = (1+4\pi|x|^2)^N$ and so we can just do this.

But what if Q has a zero? We are going to choose $\tilde{\eta}_1$ such that fixed $\xi_2, \dots, \xi_n$, we have $\{\eta_1 = \tilde{\eta}_1\}$ disjoint from the zeros of $Q(\xi_1 + i\eta_1, \xi_2, \dots, \xi_n)$. We can do this so that $|Q|$ is even bounded from below.

If $\varphi \in \mathcal{D}_K(\mathbb{R}^n)$ then $\hat{\varphi}$ is holomorphic on $\mathbb{C}^n$. Then we can write

$$E(\varphi) = \int \Big(\frac{\hat{\varphi}}{Q}\Big)(\xi_1 + i\tilde{\eta}_1(\xi_2, \dots, \xi_n), \xi_2, \dots, \xi_n) d\xi_1 \cdots d\xi_n.$$

We need to check that this is really a fundamental solution. We have

$$E(L_1\varphi) = \int_{\mathbb{R}^n} \hat{\varphi}(\xi_1 + i\tilde{\eta}_1(\xi_2, \dots, \xi_n), \xi_2, \dots, \xi_n),$$

and then we want to justify that this is equal to $\int_{\mathbb{R}^n} \hat{\varphi}(\xi_1, \dots, \xi_n) = \varphi(0)$.

To justify this, we introduce L_2 with symbol $(1 + \xi_1^2)^\ell$. Then

$$\begin{aligned}(1 + (\xi_1 + i\eta_1)^2)^\ell \hat{\varphi}(\xi_1 + i\eta_1, \xi_2, \dots, \xi_n) &= \widehat{(L_2\varphi)}(\xi_1 + i\tilde{\eta}_1, \xi_2, \dots, \xi_n) \\ &= \int_{\mathbb{R}^n} (L_2\varphi)(x_1, \dots, x_n) e^{-2\pi i((\xi_1 + i\eta_1)x_1 + \xi_2 x_2 + \cdots + \xi_n x_n)}.\end{aligned}$$

Then todo

22 November 14, 2017

Last times we were solving $LE = \delta_0$ to get the fundamental solution E, in the context of distributions. One reason we're not happy with $E \in \mathcal{D}'(\mathbb{R}^n)$ is that we can't take the Fourier transform. To take the Fourier transform, we want our distribution to be in $\mathcal{S}(\mathbb{R}^n)'$. So there is another question, of whether E can be chosen to be a tempered distribution. This is a division problem which was solved by Hörmander.

We're not going to talk about this in great detail, but the main idea is to generalize L'Hôpital's rule. We want to show that

$$\|f\|_{m',\mathcal{S}(\mathbb{R}^n)} \leq C\|Pf\|_{m,\mathcal{S}(\mathbb{R}^n)}.$$

You can just try to expand the partial derivatives of $D^\alpha(Pf)$ and do something, which is actually what people did initially, but it quickly becomes complicated as n becomes large.

22.1 Elliptic regularity

There is the analytic Laplacian

$$\Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}$$

on $\mathbb{R}^n$. In the context of Hodge theory, there is the geometric Laplacian

$$(dd^* + d^*d)\omega = -\Delta\omega.$$

In $\mathbb{R}^n$, Δ is just applying the analytic Laplacian to each coefficient.

For elliptic operators, we have interior regularity.

Definition 22.1. Let L be a single linear partial differential equation with constant coefficients and order m. Then L is said to be **elliptic** if

$$|P(\xi)| \geq c|\xi|^m$$

for some $c > 0$ and all sufficiently large $|\xi|$, where P is the symbol for L.

This just means that the if P_m is the principal (degree-m homogeneous) part of P then $|P_m(\xi)| \geq c|\xi|^m$.

Definition 22.2. L is **hypo-elliptic** if Lu being C^∞ on an open set Ω implies that u is C^∞.

We want to prove that L is hypo-elliptic. Let $Lu = f$ be C^∞. We introduce a (almost) fundamental solution by ignoring the C^∞ functions. This is also called a parametrix.

Definition 22.3. Let L be a differential operator constant coefficient on $\mathbb{R}^n$. Let Q be a distribution on $\mathbb{R}^n$. The distribution Q is said to be a **parametrix** for L if $LQ = \delta_0 + r$ with $r \in C^\infty(\mathbb{R}^n)$.

Theorem 22.4. *If L is elliptic then there exists a parametrix Q which is regular in the sense that Q is C^∞ on $\mathbb{R}^n - \{0\}$.*

Proof. The proof uses properties of the symbol P. The original method of Fourier analysis was to take the inverse Fourier transform of

$$\frac{1}{P(\xi)}$$

to get E, if it is possible. There are two difficulties: one is when $P(\xi) = 0$, and the other is at $\xi = \infty$, because it might not decay as fast as we want. To solve the second problem, we multiply by a large power of $|\xi|$. To solve the first problem, we use a C^∞ cut-off function $\gamma(\xi)$ on $\mathbb{R}$ such that $0 \le \gamma(\xi) \le 1$, and $\gamma(\xi) \equiv 0$ near the $\xi = 0$, where $P(\xi)$ could be small. We are going to take the inverse Fourier transform of γ/P.

So define Q as the inverse Fourier transform of γ/P. Let β be an arbitrary number and consider

$$\Delta^N\Big((2\pi i\xi)^\beta \frac{\gamma(\xi)}{P(\xi)}\Big).$$

Its growth at ∞ is $|\xi|^{m+2N-|\beta|}$, and if $2N + m - |\beta| > n$, then it is L^2. So its Fourier transform

$$(-4\pi|x|^2)^N \partial_x^\beta Q$$

is L^2 on $\mathbb{R}^n$. This shows that Q is C^{m-1} on $\mathbb{R}^n - \{0\}$ for all m. Because we have chosen m to be any number, we see that Q is actually C^∞ on $\mathbb{R}^n - \{0\}$.

Now we have

$$\widehat{LQ} = P \cdot \frac{\gamma}{P} = \gamma = 1 + (\gamma - 1).$$

The function $\gamma - 1$ is C^∞ on $\mathbb{R}^n$, with compact support, so its inverse Fourier transform is C^∞. Therefore $LQ - \delta$ is the inverse Fourier transform of $\gamma - 1$, which is smooth. $\square$

Generally, when $LE = \delta_0$ we can solve $Lu = f$ by taking the convolution $u = E * f$. But we need something like E is of compact support in order for this to make sense.

The idea is to take a cut-off function χ_ϵ such that $\chi_\epsilon \equiv 1$ on $\overline{B_{\epsilon/2}(0)}$ such that $\operatorname{supp}\chi_\epsilon \subseteq B_\epsilon$. We replace Q by $\chi_\epsilon Q$. Then

$$L(\chi_\epsilon Q) = \chi_\epsilon(LQ) + (L(\chi_\epsilon Q) - \chi_\epsilon LQ) = \delta_0 + \chi_\epsilon r + (L(\chi_\epsilon Q) - \chi_\epsilon LQ)$$

and $\chi_\epsilon r$ and $L(\chi_\epsilon Q) - \chi_\epsilon LQ$ are smooth.

Theorem 22.5 (Elliptic implies hypo-elliptic)**.** *Let L be an elliptic operator on $\mathbb{R}^n$. Let U be a distribution on an open subset Ω of $\mathbb{R}^n$ and assume that LU is C^∞ on Ω. Then U is C^∞ on Ω.*

Proof. Consider an open set $\Omega' \subseteq \Omega$ such that $\overline{\Omega'} + B_\epsilon(0) \subseteq \Omega$. Then there exists a parametrix Q for L with support in $B_\epsilon(0)$.

Let $f = LU$. Then

$$Q * f = (LQ) * U = (\delta_0 + r) * U = U + r * U$$

which is smooth. This shows that U is smooth on Ω', and we can let Ω' go to Ω. $\square$

23 November 16, 2017

For a distribution it is allowed to divide by a polynomial. This is all about the fundamental solution. Then people tried to do this for tempered distributions. To do this it suffices to show

$$\|g\|_{\ell',\mathcal{S}} \le \tilde{c}\|Pg\|_{\ell,\mathcal{S}}.$$

Let me give you how to solve this, and then other dimensions will be similar.

23.1 Division problem of tempered distributions by polynomials

Let me first tell you L'Hôpital's rule from Taylor expansion. We have

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \int_{t=a}^{x} \frac{(t-a)^n}{n!} f^{(n+1)}(t)dt$$

and then

$$R_n(x) = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(a + O(x-a)).$$

If there is a zero at a with order $\ge n$, then

$$\frac{f(x)}{(x-a)^n} = \frac{f^{(n)}(a)}{n!} + \frac{x-a}{n+1} f^{(n+1)}(a + O(x-a)).$$

Now consider the polynomial

$$P(x) = \prod_j (x-a_j)^{k_j}.$$

Suppose $f, g \in \mathcal{S}(\mathbb{R})$ satisfy $f = Pg$. We want to bound the sup norm of $g(x) = f(x)P(x)$ where $\operatorname{ord}_{a_j} f \ge k_j$. We need to worry about the cases

(i) x is near some a_j,

(ii) x is bounded away from all zeros of P.

For the second case, we are dividing by a smooth function so we don't have to worry. For the first case, we are going to have

$$\frac{f(x)}{P(x)} = \frac{f(x)/(x-a_j)^{k_j}}{P(x)/(x-a_j)^{k_j}}.$$

Then

$$\left|\frac{f(x)}{P(x)}\right| \le \frac{1}{c} \sup_{|y-a|\le\eta} \left|\frac{1}{h!} f^{(h)}(y) + \cdots\right| \le C_\eta \sup_{\ell\le h+1, |y-x|\le\eta} |f^{(\ell)}(y)|.$$

For the points outside $\bigcup_j B_\eta(a_j)$, we have have $|g| \le \tilde{C}|f|$ so

$$\sup_{x\in\mathbb{R}}(1+|x|)^\ell |g(x)| \le C^* \sup_{y\in\mathbb{R},\nu\le\max_j h_j+1}(1+|y|^\ell)|f^{(\nu)}(y)|.$$

For derivatives, we have

$$D^m f \sum_{p=0}^m \binom{m}{p}(D^pP)(D^{m-p}g) = P(D^m g) + \sum_{p=1}^m (D^pP)(D^{m-p}g).$$

Then

$$P(D^m g) = D^m f - \sum_{p=1}^m \binom{m}{p}(D^pP)(D^{m-p}g)$$

and use induction to get the bound.

You can read Hörmander, Arkiv. für Math. **53** (1958), 555–568 or M. Atiyah, Comm. Pure Appl. Math **23** (1970), 145–150. Here is what Atiyah did. In the case when the polynomial is locally nice, like

$$P = \prod_{\nu=1}^{p_j}(x_\nu - a_\nu(j))^{\ell_{\nu,j}},$$

then we are good. So what you do is to change coordinates and blow up. For instance, if you have $P(x,y) = x^3 - y^2$ then we can use $y = xu$ and $x = v$ to get $u^2 = v$. Then you still get a manifold, and you can do analysis locally. After you do the resolution of singularity and get a manifold, and also a "tempered distribution" on the manifold, you push forward.

23.2 Sobolev embedding

When we have a distribution, we test against compact support. Then the next best thing you can get is a tempered distribution, unless you have something special like elliptic regularity. But we want to say how bad a distribution is.

Let T be a compactly supported distribution. We define the **order** as the minimal ℓ such that we can bound

$$|T(g)| \le \|g\|_{\ell,\mathcal{D}_K} = \sup_{\nu\le\ell}|D^\nu g|$$

for a K that contains a neighborhood of the support of T. For instance, $\partial_x^\alpha \cdot \delta_0$ has order $|\alpha|$ and functions have order 0.

To test the function, we need more spaces, like $L_k^2 = W^{2,k}$. We want an embedding $W_0^{p,k} \subseteq C_0^m$. Here, we can do stuff like

$$\|Df\|_{L^2} = \int |x\hat{f}(x)|^2 = \int x^2 f(x)\cdot f(x) \le \|D^2 f\|_{L^2}\|f\|_{L^2}.$$

This suggests that we can use Hölder to bound things. The following is a theorem of Nirenberg, Ann. d. Scnola Normale Superiore d. Pisa **13** (1959), 115–162.

Theorem 23.1 (Nirenberg)**.** *Let $r \in L^q(\mathbb{R}^n)$ and $D^m u \in L^r$ for $1 \leq q, r \leq \infty$. Then*

$$\|D^j u\|_{L^p} \leq C \|D^m u\|_{L^p}^a \|u\|_{L^q}^{1-a}$$

for $j/m \leq a \leq 1$. Actually there are two exceptions:

- *if $j = 0$ and $rm < n$, $q = \infty$, then we need the additional assumption $u \in L^{\tilde{q}}$ for some $\tilde{q} > 0$.*
- *if $1 < r < \infty$ and $m - j - n/r \in \mathbb{N} \cup \{0\}$ then the inequality holds for a satisfying only $j/m \leq a < 1$.*

The main techniques will be:

- Applying the fundamental theorem of calculus applied to the radial direction and then averaging over the angular directions. This is the simplest one, but cannot handle the critical "equal" condition.
- Fundamental theorem of calculus applied to each rectangular direction. This is Nirenberg's trick.
- Moser's trick of iterating.
- Integration by parts to control more direction, which is similar to Gårding's inequality techniques.

Let me tell you the first simple trick, which shows

$$W_0^{1,p}(\Omega) \subseteq L^\infty(\Omega).$$

Assume that Ω is bounded for simplicity. Take a point $x \in \Omega$, and then integrate along the radial direction to get something about x. Then average over the paths around the sphere and use Hölder's inequality.

24 November 21, 2017

We had these four techniques in proving the Sobolev inequalities. The first technique of applying the fundamental theorem of calculus in the radius works in the following way. To avoid the boundary term, we are going to assume that the function has compact support. We want to show that

$$W_0^{p,1} \subseteq L^\infty_{\mathrm{loc}}(\mathbb{R}^n)$$

if $p > n$. Let $u \in C_c^\infty$. We want to show that

$$\|u\|_{L^\infty} \leq C\|u\|_{L_1^p}.$$

Let $\Omega \subset \mathbb{R}^n$ be a bounded set and we are assuming $\operatorname{supp} u \subset \Omega$. Because u vanishes outside a large ball, we have

$$u(x) = -\frac{1}{\mathrm{Vol}(S^{n-1})} \int_{\hat{e} \in S^{n-1}} \int_{t=0}^{T} \frac{\partial}{\partial t} u(x + t\hat{e}) dt d\hat{e}.$$

Then

$$\begin{aligned} |u(x)| &\leq \frac{1}{\mathrm{Vol}(S^{n-1})} \int_{B^n(x,T)} |\operatorname{grad} u| \frac{1}{r_x^{n-1}} \\ &\leq \left(\int_{B^n(x,T)} |\operatorname{grad} u|^p \right)^{1/p} \left(\int \left(\frac{1}{r_x^{n-1}} \right)^{p/(p-1)} \right)^{1-\frac{1}{p}} \leq C\|u\|_{L_1^p} \end{aligned}$$

where r_x is the distance to x. Because $p > n$, the integral is finite.

We can generalize this technique. For $m \geq 1$, we have the Taylor expansion

$$f(x) = \sum_{h=0}^{m} \frac{1}{k!} f^{(h)}(a)(x-a)^h + \int_{t=a}^{x} \frac{1}{m!} f^{(m+1)}(t)(b-t)^m dt.$$

The result coming from this inequality is

$$\|u\|_{L^\infty} \leq C\|u\|_{W_0^{k,p}}.$$

24.1 Nirenberg's trick of integrating in a rectangular direction

This is quite impressive. You want to fix the axis direction and then integrate in these directions. But to be fair, you need to do it from the other end too. Consider the jth coordinate. We have

$$|2u(x)| \leq \int_{-\infty}^{\infty} |\partial_j u(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_n)| dt.$$

We take the product for all the j's and get

$$|2u(x)|^n \leq \prod_{j=1}^{n} \int_{-\infty}^{\infty} |\partial_j u(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_n)| dt.$$

But this is only for one coordinate. So we integrate with respect to x_1. The first factor is independent of x_1, so we can take it out and have $n-1$ factors. Then by Hölder,

$$\int |f_1 \cdots f_{n-1}| \le \left(\int |f_1|^{n-1}\right)^{\frac{1}{n-1}} \cdots \left(\int |f_{n-1}|^{n-1}\right)^{\frac{1}{n-1}}.$$

Using this, we get

$$\int_{x_1=-\infty}^{\infty} |2u(x)|^{\frac{n}{n-1}} \le \left(\int_{t=-\infty}^{\infty} \partial_j u(t, x_2, \dots, x_n)\right)^{\frac{1}{n-1}} \prod_{j=2}^{n} \left(\int_{x_1=-\infty}^{\infty} \int_{t=-\infty}^{\infty} |\partial_j u(x-1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n)| dt dx_1\right)^{\frac{1}{n-1}}.$$

But we have integrate for x_2 too. Again, there is precisely one factor that does not depend on x_2, so we can do the same thing. After all this, we get

$$\int_{\mathbb{R}^n} |2u(x)|^{\frac{n}{n-1}} \le \left(\prod_{j=1}^{n} \int_{\mathbb{R}^n} |\partial_j u|\right)^{\frac{1}{n-1}}.$$

Then

$$\left(\int_{\mathbb{R}^n} |2u(x)|^{\frac{n}{n-1}}\right)^{\frac{n-1}{n}} \le \frac{1}{n} \sum_{j=1}^{n} \int_{\mathbb{R}^n} |\partial_j u|$$

by the arithmetic-geometric mean inequality. This inequality is precise, even with constants. Therefore

$$\|u\|_{L^{n/(n-1)}} \le C\|u\|_{L_1^1}.$$

This is the $p = 1$ case. In general, we can generalize to $1 \le p < n$. The inequality here is going to be

$$\|u\|_{L^{p^*}(\mathbb{R}^n)} \le C\|u\|_{W_0^{1,p}(\mathbb{R}^n)},$$

where the Sobolve conjugate exponent is

$$p^* = \frac{np}{n-p} \text{ for } 1 \le p < n.$$

The usual way of getting this inequality is to replace u by $|u|^\gamma$. Then the case $p = 1$ implies

$$\||u|^\gamma\|_{L^{\frac{n}{n-1}}(\mathbb{R}^n)} \le C\|D(|u|^\gamma)\|_{L^1(\mathbb{R}^n)},$$

and then you can use Hölder on the right hand side and choose an appropriate γ.

24.2 Moser's iteration trick

This is an way to get the sup norm estimate. I'll illustrate to recover the sup estimate from the spherical coordinate method. You can do this for the rectangular integration, but I am trying to just tell you how this works. We had the inequality

$$\||u|^\gamma\|_{L^{\frac{n}{n-1}}} \leq C\gamma \||u|^{\gamma-1}\|_{L^{\frac{p}{p-1}}} \|Du\|_{L^p}$$

for $p, \gamma > 1$, by Hölder. We are going to iterate this to get the estimate on infinity.

First, note that we can scale the function so that $C\|Du\|_{L^p(\mathbb{R}^n)} = 1$. Then

$$\||u|^\gamma\|_{L^{\frac{n}{n-1}}} \leq \gamma \||u|^{\gamma-1}\|_{L^{\frac{p}{p-1}}},$$

which means that

$$\|u\|_{L^{\gamma n'}} \leq \gamma^{\frac{1}{\gamma}} \|u\|^{\frac{\gamma-1}{\gamma}}_{L^{\frac{(\gamma-1)p}{p-1}}} = \gamma^{\frac{1}{\gamma}} \|u\|^{\frac{\gamma-1}{\gamma}}_{L^{(\gamma-1)p'}}.$$

Let us assume that $p > n$, so that we have $n' > p'$. Then we increase the norm number by a factor of n'/p' every time we apply the inequality. Here, we can be a bit generous and replace the $L^{(\gamma-1)p'}$ norm by the $L^{\gamma p'}$ norm, using Hölder. There will be a constant coming from the volume of Ω, but we can again scale Ω so that its volume is 1. Then

$$\|u\|_{L^{\gamma n'}} \leq \gamma^{1/\gamma} \|u\|^{1-\frac{1}{\gamma}}_{L^{\gamma p'}}.$$

Write $\delta = \frac{n'}{p'}$. If you iterate this, you get

$$\|u\|_{L^{\delta^\nu n'}} \leq \prod_{\mu=1}^{\nu} (\delta^\nu)^{\frac{1}{\delta^\mu}}.$$

This converges and therefore

$$\|u\|_{L^\infty(\mathbb{R}^n)} \leq C\|u\|_{W_0^{1,p}(\mathbb{R}^n)}.$$

25 November 28, 2017

We want to look at measures in general. We want something like the Lebesgue decomposition, into an absolutely continuous part and a jump part.

We are also going to look at ergodicity. Suppose $T : X \to X$ is a measure-preserving map. We want to say something about the convergence of

$$\frac{1}{m}\sum_{k=0}^{m-1} T^k(f) \to \text{const.}$$

Here, T is assumed to be ergodic in the sense that $T(S) = S$ up to measure zero implies $S = 0$ or $S = X$. Birkhoff proved this, and we are going to do this in three parts. First we will prove convergence without assuming ergodicity. This was motivated by statistical mechanics.

25.1 Radon–Nikodym derivative

This is the second half of the fundamental theorem of calculus in Lebesgue theory. We have introduced the maximal function

$$f^*(x) = \sup_{x\in B} \frac{1}{m(B)} \int_B |f|$$

If $f \in L^1$, then it is not true that $f^* \in L^1$. But it is true that $f \in L^2$ implies $f^* \in L^2$. This was assigned in the homework. Then we can show that

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

almost everywhere.

Now we want to do the other part. This is differentiating and then integrating.

Definition 25.1. A **measure space** (X, μ) contains the data of a σ-algebra, a collection $\mathcal{E}$ of X containing $\emptyset, X$ that is closed under countable unions and complements, and a measure $\mu : \mathcal{E} \to [-\infty, \infty]$ with countable additivity.

Example 25.2. Let $X = \mathbb{R}^k$ with the standard metric. Let $\mathcal{E}$ be the σ-algebra generated by the open set, the Borel σ-algebra. Then we can define μ as the Lebesgue measure.

We can do the same thing, and we get monotone convergence, dominated convergence, and Fatou. I'm not going to do this because we've done it and everything works the same way.

There still is the question of how to define a measure from σ-algebra. The way to do it is to define an exterior measure first and then use the inf of the exterior measures containing it. The usual way to do this is Carathéodory's criterion. This says that A is measurable if and only if $m^*(A\cap E)+m^*(E-A) = m^*(E)$ for all E.

Theorem 25.3. *Assume that μ, ν are two (signed) σ-finite ($|\mu|$ and $|\nu|$) measures. Then we can decompose $\nu = \nu_a + \nu_s$ where ν_a is absolutely continuous with respect to μ (i.e., $\mu(E) = 0$ implies $\nu_a(E) = 0$) and ν_s is (mutually) singular with respect to μ (i.e., $X = A \cup B$ for disjoint A, B and $\operatorname{supp} \nu_s \subseteq B$ and $\operatorname{supp} \mu \subseteq A$). Moreover,*

$$\nu_a(E) = \int_E f d\mu$$

for some f measurable on X.

Proof. Von Neumann's proof is the most elegant. This uses Riesz representation. Let us consider the special case $\mu, \nu \geq 0$ and $\mu(X), \nu(X) < \infty$. Let $\rho = \mu + \nu$ and consider the Hilbert space $L^2(X, \varphi)$. Consider the linear functional

$$\Phi(f) = \int f d\nu.$$

Then this is continuous because X has finite measure, and so there exists a $g \in L^2(X, \rho)$ such that

$$\Phi(f) = \int_X f d\nu = \int_X f g d\rho.$$

His main idea is that all the information between μ and ν is encoded in g. First note that $0 \leq g \leq 1$ because testing with $f = \chi_E$ gives a contradiction. Now consider the sets

$$B = \{g = 1\}, \quad A = \{g < 1\}.$$

Then $\mu(B) = 0$ by testing with $f = \chi_B$.

Now what happens in $E \subseteq A$? Set $f = \chi_E(1 + g + \cdots + g^{n-1})$. Then we get

$$\int_E (1 - g^n) d\nu = \int_E g(1 + g + \cdots + gn - 1) d\mu.$$

As $n \to \infty$, we get, by monotone convergence,

$$\nu(E) = \int_E \frac{g}{1 - g} d\mu.$$

If it is not σ-finite, you take the union, and if it is signed, you can just separate into positive and negative parts. $\square$

25.2 Ergodic measure theory

Let (X, μ) be a nonnegative measure with $\mu(X) = 1$. Let $T : X \to X$ be a measurable map and assume that it is an isometry, i.e., measure-preserving. (Usually it is not assumed to be bijective.) Our goal is to to look at the stable situation. We want to look at the limit $\lim_{n\to\infty} T^n$, but this is too complicated. So we use the Césaro limit. In that case, we are looking at $\frac{1}{m}(1 + T + T^2 + \cdots + T^{m-1})$.

Of course, you can't add measure spaces. So you take a function f defined on X, and let $\tau : X \to X$ be an isometry. Then we define

$$(Tf)(x) = f(\tau(x)).$$

So the goal now is to look at the average

$$A_m(f) = \frac{1}{m}\sum_{h=0}^{m-1} T^h(f).$$

For functions, there are different kinds of convergences. There is strong convergence, convergence in norms. There is convergence in measure, and there is almost everywhere convergence. People were interested in almost everywhere convergence, because they wanted to get back to the space. We'll discuss only the case (X, μ) with $\mu \geq 0$ and $\mu(X) = 1$.

Theorem 25.4 (Mean ergodic theorem). *Let $\tau : X \to X$ be measure-preserving and let $(Tf)(x) = f(\tau(x))$. Then for $f \in L^2(X)$,*

$$A_m(f) = \frac{1}{m}\sum_{h=0}^{m-1} T^h f \to P(f),$$

as $m \to \infty$, where $P(f)$ is the L^2-projection to $S = \{f : Tf = f\} = \ker(I - T)$.

Proof. The key is to decompose $\mathcal{H} = S \oplus S^\perp$. Clearly S goes to S. Now we need to know the behavior of T on $S^\perp$. Note that $\ker(I - T) = S$ is the orthogonal complement of $\operatorname{im}(1 - T^*)$.

But note that T is an isometry, so $T^*T = I$. (We have $(T^*Tf, g) = (f, g)$.) We want to show that $Tf = f$ if and only if $T^*f = f$. We now hat $Tf = f$ implies $T^*f = f$. For the other direction, we use the equality condition on Cauchy–Schwartz. If $T^*f = f$, then

$$\|f\|^2 = (f, f) = (f, T^*f) = (Tf, f) \leq \|Tf\|\|f\|.$$

So $Tf = cf$ and then $c = 1$ and so $Tf = f$.

What this shows is that $S = \ker(1 - T^*) = \ker(1 - T)$. So the orthogonal complement of S is the closure of $\operatorname{im}(1 - T)$.

So every $f \in \mathcal{H}$ can be decomposed as

$$f = F_0 + (1 - T)F_1 + F_2$$

where $\|F_2\| < \epsilon$. Then the $(1 - T)F_1$ telescopes and contributes nothing. So it converges to F_0 in L^2. $\square$

Theorem 25.5 (Maximal ergodic theorem). *Let $f \in L^1(X, \mu)$ and define a maximal function*

$$f^*(x) = \sup_{m \geq 1} \frac{1}{m}\sum_{k=0}^{m-1} |f(\tau^k(x))|.$$

Then there exists a universal constant $A > 0$ such that

$$\mu(\{x : f^*(x) > \alpha\}) \leq \frac{A}{\alpha} \|f\|_{L^1}.$$

Proof. Let us first prove the special case $X = \mathbb{Z}$, μ the cardinal measure, and $\tau : n \mapsto n + 1$. We do this by replacing $\mathbb{Z}$ by $\mathbb{R}$ and replacing points with half-closed intervals of length 1. Then this just follows from the Hardy–Littlewood bound.

The general case can be reduced to this. Consider $X \times \mathbb{Z}$ with the product measure, with the map $(x, n) \mapsto (\tau x, n+1)$. Then you can do this carefully. □

Theorem 25.6 (Pointwise ergodic theorem)**.** *Suppose $f \in L^1(X)$ where $\mu(X) = 1$. Let $\tau : X \to X$ be measure preserving. Then for almost all $x \in X$,*

$$A_m(f) = \frac{1}{m} \sum_{h=0}^{m-1} f(\tau^h(x)) \to P'(f)$$

for almost everywhere, and by Fatou, $\int_X |P'f| d\mu \leq \int_X |f| d\mu$.

Proof. First note that L^2 is dense in L^1 by truncating the range. Here, we have the decomposition $\mathcal{H} = S \oplus S^\perp$ and so $f = F_0 + (1 - T)F_1 + F_2$. We don't need anything for F_0, and $(1-T)F_1$ is also telescoped well. To deal with F_2, we use the maximal bound. Then the places it does not converge is going to have measure zero. □

26 November 30, 2017

There is the Radon transform, which used in tomography. This is similar to Fourier transform, where are trying to recover f from integration of f on lines. Here, we can actually recover $\hat{f}$ from this information, because every Fourier transform is an integral of these.

26.1 Hausdorff measure

These are used for fractals, or self-similar objects. What is the main difference between Hausdorff measures and Euclidean/Lebesgue measures? For Lebesgue measures, we started out with building blocks, intervals or cubes. So these are volume-based measures. But then you need to know the dimension to start out with.

The Hausdorff-measure is diameter-based. This is much more powerful, because it is defined for any metric space, regardless of the dimension. Then you can use it to test and define the dimension.

Definition 26.1. Let X be a metric space, and let E be a subset. Given $\delta > 0$, choose a d for testing the dimension. We define

$$\mathscr{H}_d^\delta(E) = \inf\left\{\sum_j (\operatorname{diam} F_j)^d : E \subseteq \bigcup_j F_j \text{ and } \operatorname{diam} F_j < \delta\right\}.$$

Then the exterior measure is defined as

$$m_d^*(E) = \lim_{\delta\to 0} \mathscr{H}_d^\delta(E).$$

Note that if $m_\alpha^*(E) < \infty$ for some $\alpha \geq 0$, then $m_b^*(E) = 0$ for $\beta > \alpha$. This is because there is an extra factor of $\delta^{\beta-\alpha}$. Likewise, if $m_\alpha^*(E) > 0$ for some $\alpha > 0$ then $m_\beta^*(E) = \infty$ for $\beta < \alpha$.

Definition 26.2. The **Hausdorff dimension** is defined as $\sup\{\alpha : m_\alpha^*(E) = \infty\} = \inf\{\alpha : m_\alpha^*(E) = 0\}$.

If F_1 and F_2 are positively separated sets, then $m_\alpha^*(F_1) + m_\alpha^*(F_2) = m_\alpha^*(F_1 \cup F_2)$. This is because we can set δ sufficiently small and they don't interfere.

Example 26.3. Consider the Cantor set $\mathcal{C} \subseteq [0,1]$. Its Hausdorff dimension is $\alpha = \log 2/\log 3$. We can cover $\mathcal{C}$ by some 2^k intervals of form $[3^{-k}l, 3^{-k}(l+1)]$ and then $m_\alpha^*(\mathcal{C}) \leq 1$. The other direction uses the Cantor–Lebesgue function. This function is α-Hölder continuous. So we get $m_\alpha^\alpha(\mathcal{C}) > 0$.

Lemma 26.4. *Let $f : E \to F$, where E is compact and $F = f(E)$. If f is γ-Hölder continuous with* $\operatorname{dist}(f(x), f(y)) \leq M \operatorname{dist}(x,y)$*, then*

$$m_\beta^*(f(E)) \leq M^\beta m_\alpha^*(E)$$

where $\beta = \alpha/\gamma$.

Example 26.5. There is also the Sierpinski triangle. This set is a triangle, minus the middle triangle of side length $\frac{1}{2}$, and from each three remaining triangle remove the half-size triangles and so on. This has dimension $\log 3/\log 4$.

Example 26.6. There is also the Koch curve. There is also the Peano curve, or the space-filling curve.

There are also self-similar constructions, with rescaling by a factor of κ.

Theorem 26.7. *Let $S_1, \ldots, S_m$ be m separated similarities by a factor of κ. That is, there exists an open $\mathcal{O}$ such that $S_j(\mathcal{O})$ are disjoint in $\mathcal{O}$. Then there exists a unique F such that F is the disjoint union of $S_j(F)$. Moreover, its Hausdorff dimension is $\log m/\log(1/\kappa)$.*

26.2 Radon transform

The emphasis here is on the smoothness, or the Hölder continuity of the result. Let me tell you the answer first.

Definition 26.8. The **Radon transform** is, for a function f on $\mathbb{R}^d$, $t \in \mathbb{R}$, and γ is a unit direction in $\mathbb{R}^d$,

$$R(f)(t,\gamma) = \int_{P_{t,\gamma}} f.$$

Here $P_{t,\gamma} = \{x : x \cdot \gamma = t\}$.

The result concerns "well-posedness". In the real situation, everything is measured approximately. So it would be bad if a small error in measurement gives a large change in the inverse function. Consider

$$R^*(f)(\gamma) = \sup_{t,\mathbb{R}} |R(f)(t,\gamma)|$$

and then we are going to use $\int_{S^{d-1}} R^*(f)(\gamma) d\sigma(\gamma)$ as the measure of the error.

Theorem 26.9. *We have*

$$\int_{S^{d-1}} R^*(f)(\gamma) d\sigma(\gamma) \le c(\|f\|_{L^1} + \|f\|_{L^2})$$

for $d \ge 3$ and some $c > 0$ only depending on d.

Theorem 26.10. *If f is continuous on $\mathbb{R}^d$ with compact support, with $d \ge 3$, then*

$$\int_{\gamma \in S^{d-1}} \sup_{t_1 \neq t_2} \frac{|R(f)(t_1,\gamma) - R(f)(t_2,\gamma)|}{|t_1 - t_2|^\alpha} d\sigma(\gamma) \le c(\|f\|_{L^1} + \|f\|_{L^2})$$

for $0 < \alpha < \frac{1}{2}$.

If we want to invert the Radon transform, we perform the Fourier transform in the remaining variable. So if we do Fourier transform only on t, we get

$$\hat{R}(f)(\tau,\gamma) = \hat{f}(\tau\gamma).$$

This is why we need the condition $d \geq 3$. The main tool we were using were invariance of L^2 under the Fourier transform. Here, we are using the Euclidean version and the spherical version. So if f is continuous and compactly supported, then

$$\int_{S^{d-1}} \int_{-\infty}^{\infty} |\hat{R}(f)(\lambda,\gamma)|^2 |\lambda|^{d-1} d\lambda d\sigma(\gamma) = 2\int_{\mathbb{R}^d} |f(x)|^2 dx.$$

Lemma 26.11. *Let F be a function on $\mathbb{R}$, and let $d \geq 3$. If $\sup|\hat{F}| \leq A$ and $\int_{-\infty}^{\infty} |\hat{F}(\lambda)|^2 |\lambda|^{d-1} d\lambda \leq B^2$, then $\sup_{t\in\mathbb{R}} |F(t)| \leq c(A+B)$. Moreover, if $0 < \alpha < \frac{1}{2}$ then*

$$|F(t_1) - F(t_2)| \leq C_\alpha |t_1 - t_2|^\alpha (A+B).$$

Index