Math 55b - Honors Real and Complex Analysis

Taught by Yum-Tong Siu
Notes by Dongryul Kim

Spring 2016

This course, which was a continuation of Math 55a, was taught by Yum-Tong Siu. We met on Tuesdays and Thursdays from 2:30 to 4:00 pm in Science Center 222. There were 12 students taking the course. There was one in-class midterm and also a take-home final exam. The course assistants were Calvin Deng and Vikram Sundar.

Contents

1 January 26, 2016

1.1 Outline

This is course is about real and complex analysis. We will cover the following:

1) Rigorous foundation
   We will first do differentiation and integration in one dimension. This will include the fundamental theorem of calculus. Then we generalize the concepts into higher dimension, and talk about commutation of two differentiations and integration, differential forms, curvature, and Stokes's theorem.
2) Complex analysis
   Actually we had a taste of this when we proved the fundamental theorem of algebra. We will do Cauchy's theory.
3) Solving differential equations
   We solve differential equations using Fourier analysis. We will also go into partial differential equations with constant coefficients and fundamental solutions.

1.2 Convergence of a sequence

Now we get down to business. The important thing about analysis is of course taking limits. At the beginning of last semester, we constructed the real numbers by looking at Dedekind cuts. This gives us the least upper bound property. In other words, if $A \subset \mathbb{R}$ admits an upper bound, then A admits a least upper bound.

The next main technique is limits.

Definition 1.1. A real sequence $\{x_n\}$ has **limit** a if and only if given any $\epsilon > 0$ there exists a positive integer $N \in \mathbb{Z}$ such that $|x_n - a| < \epsilon$ for any $n \geq N$.

We can formulate the limit property without having the actual limit.

Definition 1.2. A real sequence $\{x_n\}$ has the **Cauchy property** if for any $\epsilon > 0$ there exists a positive integer $N \in \mathbb{Z}$ such that $|x_n - x_m| < \epsilon$ for any $n, m \geq N$.

And in fact, we have the following theorem.

Theorem 1.3. *A real sequence $\{x_n\}$ converges to some a if and only if $\{x_n\}$ is a Cauchy sequence.*

We will first use the least upper bound property to prove the following lemma:

Lemma 1.4. *Every infinite sequence $\{x_n\}$ in a bounded subset A of $\mathbb{R}$ admits a convergence subsequence, i.e., there exists an infinite increasing sequence $\{n_k\} \subset \mathbb{Z}_+$ such that $\lim_{k\to\infty} x_{n_k} = a$.*

Proof. We invoke the existence of the supremum and the infimum. The idea of the proof is to construct a new sequence $\{a_m\}$ defined by

$$a_m = \sup_{n \geq m} \{x_n\}.$$

Then clearly $a_1 \geq a_2 \geq \cdots$, and then we can define $a = \inf_n a_n = \lim_{n\to\infty} a_n$. For future reference, we will make a definition.

Definition 1.5. For a bounded real sequence $\{x_n\}$, we define the **limit supremum** as

$$\limsup_{n\to\infty} x_n = \inf_m \sup_{n \geq m} x_n.$$

So let $a = \limsup_{n\to\infty} x_n$. Given any integer $k > 0$, we see that $a_m - \frac{1}{k}$ is not an upper bound. This means that $a_m - \frac{1}{k} < x_n \leq a_m$ for some $n \geq m$. We use the fact that $a + \frac{1}{k}$ is not a lower bound of a_n. Something like this will prove the theorem. □

Proof of theorem 1.3. It can be seen from the triangle inequality that if a sequence converges, then it is Cauchy.

We now need to show that if a sequence Cauchy, it converges. Using the lemma, we see that given any $\epsilon > 0$ there exists N_1, N_2 such that $|x_n - x_m| < \epsilon$ for any $n, m \geq N_1$, and $|x_{n_k} - a| < \epsilon$ for any $k \geq N_2$. Then we get for any $k \geq \max\{N_1, N_2\}$,

$$|x_k - a| \leq |x_k - x_{n_k}| + |x_{n_k} - a| < \epsilon + \epsilon = 2\epsilon.$$

This ends the proof. □

1.3 Uniform convergence

Later on, we will start doing integration over higher dimensions, and we will need to look at sequences that are multi-indexed. We can look at a doubly indexed sequence $\{x_n^{(m)}\}$. Then there are two ways to get a limit. We can first fix n and get $\lim_{m\to\infty}$ and then get the whole limit, or first fix m. But will these two limits be the same, or in other words,

$$\lim_{n\to\infty} \lim_{m\to\infty} x_n^{(m)} = \lim_{m\to\infty} \lim_{n\to\infty} x_n^{(m)}?$$

It is not true in general, but it is under a certain condition.

Definition 1.6. A sequence of real sequence $x_n^{(m)} \to a^{(m)}$ **converges uniformly** in m as $n \to \infty$ if and only if given any $\epsilon > 0$, there exists a $N = N_\epsilon$ such that $|x_n^{(m)} - a^{(m)}| < \epsilon$ for any $n \geq N$ and m.

Theorem 1.7. *Assume that $x_n^{(m)} \to a^{(m)}$ converges uniformly, and that for any fixed m, we have $x_n^{(m)} \to a_n$. Then*

$$\lim_{n\to\infty} \lim_{m\to\infty} x_n^{(m)} = \lim_{m\to\infty} \lim_{n\to\infty} x_n^{(m)}.$$

(This includes the fact that the limit exists.)

Proof. We first show that $\lim_{n\to\infty} a_n$ exists. Interpreting the uniform convergence condition in terms of Cauchy sequences, we see that for any $\epsilon < 0$, there is a $N = N_\epsilon$ such that

$$|x_n^{(m)} - x_k^{(m)}| < \epsilon$$

for any $n, k \geq N$ and m. Taking the limit when $m \to \infty$, we get $|a_n - a_k| \leq \epsilon$ and thus $\{a_n\}$ is a convergence sequence. Let its limit be a.

We now show that $\lim_{m\to\infty} a^{(m)} = a$. Take any $\epsilon > 0$. Then there is a N_1 such that for any $n \geq N_1$, we have

$$|a - a_n| < \epsilon \tag{1}$$

for any $n \geq N_1$. Also, by uniform convergence, we see that there is an N_2 such that

$$|x_n^(m) - a^{(m)}| < \epsilon \tag{2}$$

for any $n \geq N_2$ and m. Now we fix a sufficiently large n. Then there is a M such that

$$|a_n - x_n^{(m)}| < \epsilon \tag{3}$$

for any $m > M$ and the fixed n. Adding up, we get

$$|a - a^{(m)}| \leq |a - a_n| + |x_n^(m) - a^{(m)}| + |a_n - x_n^{(m)}| < 3\epsilon.$$

This shows that $\lim_{m\to\infty} a^{(m)} = a$. ☐

2 January 28, 2016

Last time we looked at sequences with double indices. These things are important, because we want to interchange two limiting processes sometimes. For instance, whether $\frac{\partial}{\partial x}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}\frac{\partial}{\partial x}$ is true, or whether $\lim \sum = \sum \lim$ is true, etc. These became important in solving differential equations.

The next step is to replace the discrete variable n in x_n by a continuous variable x in $f(x)$.

2.1 Limit of a function

Definition 2.1. Let f be a function defined in (a, b), where $a, b \in \mathbb{R}$. For a point $c \in (a, b)$, the equation $\lim_{x\to c} f(x) = L$ means that given any $\epsilon > 0$, there exists a $\delta > 0$ such that $|f(x) - L| < \epsilon$ for any $0 < |x - c| < \delta$.

There is an equivalent formulation in terms of sequences.

Proposition 2.2. *Let $f : (a, b) \to \mathbb{R}$ be a function. Then $f(x) \to L$ as $x \to c$ if and only if for any sequence $x_n \to c$ we have $f(x_n) \to L$.*

Proof. We first prove the only if direction. Given a sequence $x_n \to c$ and a given $\epsilon > 0$, there is a $\delta > 0$ such that $0 < |x - c| < \delta$ implies $|f(x) - L| < \delta$. Then the sequence x_n eventually enters that intervals and thus $f(x_n)$ converges to L. (Actually there might be a problem if $x_n = c$ but it is a technical detail.)

The other direction can be checked. □

Definition 2.3. A function $f : (a, b) \to \mathbb{R}$ is **continuous** at $c \in (a, b)$ if and only if $f(x) \to f(c)$ as $x \to c$.

There are three important properties of continuous functions.

1) The sup and inf of a continuous function $f : [a, b] \to \mathbb{R}$ is achieved by f at some point of $[a, b]$.

   Proof. Let $A = \sup_{[a,b]} f$, where for now we even allow $A = \infty$. Then for some sequence x_n in $[a, b]$, such that $f(x_n) \to A$. Then since $[a, b]$ is bounded, there is a convergent subsequence x_{n_k} converging to $x^* \in [a, b]$. This means that $f(x_{n_k}) \to f(x^*)$ and this has to be A. That is, $f(x^*) = A$. □

2) Any function $f : [a, b] \to \mathbb{R}$ is **uniformly continuous**, i.e., for an arbitrary $\epsilon > 0$, there is a δ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$.

   Proof. Suppose it is false. This means that there is an $\epsilon > 0$ such that the implication is false for any $\delta > 0$. In particular, let $\delta = 1/n$. Then there exists two points $x_n, y_n \in [a, b]$ for which $|x_n - y_n| < 1/n$ but $|f(x_n) - f(y_n)| \geq \epsilon$. Because x_n is an infinite bounded sequence, there is a subsequence x_{n_k} that converges to some x^* as $k \to \infty$. Then there is another subsequence of the subsequence $y_{n_{k_l}}$ converging to some y^* as $l \to \infty$. Then using the continuity of f, we get $|f(x^*) - f(y^*)| \geq \epsilon$. But since $|x_n - y_n| < 1/n$, we have $x^* = y^*$. This contradicts that $\epsilon > 0$. □

3) Let $f : [a, b] \to \infty$ be a continuous function. Then for any ξ between $f(a)$ and $f(b)$, there is a $c \in [a, b]$ such that $f(c) = \xi$.

Proof. Suppose that ξ is not in the image of f. Then there exists an $\epsilon > 0$ such the $[\xi - \epsilon, \xi + \epsilon]$ is not in the image of f. This is because if for every n there is some x_n such that $f(x_n) \in [\xi - \frac{1}{n}, \xi + \frac{1}{n}]$, then there is a convergent subsequence, and we can use the continuity of f to show find that the limit of the subsequence is a solution.

So there is an $\epsilon > 0$ such that for any x, either $f(x) < \xi - \epsilon$ or $f(x) > \xi + \epsilon$. Let

$$A = \{x \in [a, b] : f(x) < \xi - \epsilon\}, \quad B = \{x \in [a, b] : f(x) > \xi + \epsilon\}.$$

Then clearly $A \cup B = [a, b]$ and $a \in A$ and $b \in B$. (We assume without loss of generality that $f(a) < f(b)$.) Let $c = \sup A$. We observe that c cannot be in A, because if $c \in A$, then $f(c) < \xi - \epsilon$ and then there is a $\delta > 0$ such that again $f(c + \delta < \xi - \epsilon$ and then $c + \delta \in A$ and then it contradicts the assumption that c is the supremum of A. On the other hand, if $c \in B$ this means that for any n there is an $x_n \in A$ and $c - 1/n < x_n < c$ and this sequence converges to c as $n \to \infty$ and thus $f(x_n) \to f(c)$ but because $c \in B$ we have $f(c) > \xi + \epsilon$ but $f(x_n) < \xi - \epsilon$ for any n. This is clearly a contradiction. □

2.2 Differentiation

Definition 2.4. Let $f : (a, b) \to \mathbb{R}$ be a real function. For an $c \in (a, b)$, we define the **derivative** of f at c as

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

if it exists.

The most important property of the derivative is the following.

Theorem 2.5 (Mean value theorem)**.** *Let $f(x) : [a, b] \to \mathbb{R}$ be a continuous function at suppose that $f'(x)$ exists on (a, b). Then there is a $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Geometrically, this means that for any continuous graph, there is a line tangent to the graph which is parallel to the line connecting $(a, f(a))$ and $(b, f(b))$.

Proposition 2.6 (Critical point property)**.** *Let $f : (a, b) \to \mathbb{R}$ be a continuous function, and assume that either $f(c) = \sup f$ or $f(c) = \inf c$ for some $c \in (a, b)$. If $f'(c)$ exists, then $f'(c) = 0$.*

Proof. The proof is simple. For instance, suppose that $f(c) = \sup f$. Then we have $(f(x) - f(c))/(x - c) \leq 0$ if $x > c$ and $(f(x) - f(c))/(x - c) \geq 0$ if $x < c$. Then as x approaches c, we see that $f'(c) = 0$. □

Proof of the mean value theorem. We first consider the simpler case $f(a) = f(b) = 0$. (This is Rolle's theorem.) By the extremal property, $\sup f$ and $\inf f$ are both achieved somewhere. If both are zero, then $f \equiv 0$ and thus we can set any c. If either sup or inf are nonzero, the derivative becomes zero at that point.

Now we consider the general case. We do an interpolation to get the general case from the special case. Let

$$F(x) = f(a)\frac{x - b}{a - b} + f(b)\frac{x - a}{b - a}.$$

Then we have F is a linear function, and $F(a) = f(a)$ and $F(b) = f(b)$. Apply Rolle's theorem to $f(x) - F(x)$ and we get some $c \in (a, b)$ such that

$$f'(c) = F'(c) = \frac{f(b) - f(a)}{b - a}.$$

□

This is needed to prove the fundamental theorem of calculus. As a corollary, we get the following.

Corollary 2.7. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If $f' \equiv 0$ on (a, b), then f is a constant function on $[a, b]$.*

2.3 Riemann integration

A function $f : [a, b] \to \mathbb{R}$ is given. For simplicity, let us assume that f is continuous on $[a, b]$.

Definition 2.8. A **partition** P of $[a, b]$ means a finite sequence

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b.$$

Definition 2.9. Given a partition P, let us choose $\xi_j \in [x_{j-1}, x_j]$ for each $j = 1, \ldots, n$. Then then **Riemann sum** is defined as

$$\sum_{j=1}^{n} f(\xi_j)(x_j - x_{j-1}).$$

The **upper sum** and the lower sum are defined as

$$U(P, f) = \textstyle\sum_{j=1}^{f} \Big(\sup_{[x_{j-1}, x_j]}\Big)(x_j - x_{j-1}),$$
$$L(P, f) = \textstyle\sum_{j=1}^{f} \Big(\inf_{[x_{j-1}, x_j]}\Big)(x_j - x_{j-1}).$$

Proposition 2.10. *Assume that $f : [a, b] \to \mathbb{R}$ is continuous. Then*

$$\inf_P U(p, f) = \sup_P L(p, f),$$

and we define this value as

$$\int_a^b f(x)dx = \inf_P U(p, f).$$

Proof. This follows from the uniform continuity of f. Given any $\epsilon > 0$ there exists a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. Then if $\max_{1 \le j \le n}(x_j - x_{j-1}) < \delta$, then we have

$$0 \le U(P, f) - L(P, f) < \epsilon(b - a).$$

Then the infimum of $U(P, f)$ and supremum of $L(P, f)$ cannot be different. □

Theorem 2.11. *Let $a < b < c$ be real numbers and $f : [a, c] \to \mathbb{R}$ be a continuous function. then*

$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx.$$

The key to proving this theorem is the refinement of partitions. Let P be the partition

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

and Q be

$$a = \xi_0 < \xi_1 < \cdots < \xi_{\nu-1} < \xi_\nu = b.$$

We say that Q is a **refinement** of P if $\{\xi_0, \dots, \xi_\nu\} \supset \{x_0, \dots, x_n\}$. We observe that if Q is a refinement of P, then

$$L(P, f) \le L(Q, f) \le U(Q, f) \le U(P, f).$$

We will prove this theorem next time.

3 February 2, 2016

Let us recall some things from last class. If f is continuous on $[a,b]$, then we define

$$\int f = \inf_P U(P,f) = \sup_P L(P,f).$$

Theorem 3.1. *Let $f : [a,b] \to \mathbb{R}$ be a continuous function and let $a < c < b$. Then*

$$\int_a^b f(x)dx = \int_a^c f(x)dx = \int_c^b f(x)dx.$$

Proof. For any partition P_1 of $[a,c]$ and P_2 of $[c,b]$, we can always put them together to form a partition P on $[a,b]$. Conversely, given any partition Q of $[a,b]$ we can refine the partition by sticking in c at some point and dividing it into a partition Q_1 of $[a,c]$ and Q_2 of $[c,b]$.

This shows that each side gives a better estimation than the other, and thus that both values are within ϵ error for each $\epsilon > 0$. Therefore the two values are equal. □

3.1 Fundamental theorem of calculus

Theorem 3.2 (Fundamental theorem of calculus)**.** *(a) Let $f : [a,b] \to \mathbb{R}$ be a continuous function, and let*

$$F(x) = \int_a^x f(t)dt.$$

Then F is differentiable on (a,b) and $F'(x) = f(x)$ for each $x \in (a,b)$.
(b) Let $F : (\tilde{a}, \tilde{b}) \to \mathbb{R}$ be a function that is differentiable, and whose derivative is continuous. Then

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

Proof of (a). The difference quotient is

$$F'(x) = \lim_{h\to 0} \frac{\int_a^{x+h} f(t)dt - \int_a^x f(t)dt}{h} = \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)dt.$$

Because we want to show that this is $f(x)$, we subtract $f(x)$ and get

$$\left|\frac{1}{h}\int_x^{x+h} f(t)dt - f(x)\right| = \left|\frac{1}{h}\int_x^{x+h} (f(t) - f(x))dt\right| \le \sup_{x\le t\le x+h} |f(t) - f(x)|.$$

But this clearly goes to zero as $h \to 0$. □

Proof of (b). We let the upper limit b vary. That is, we check instead that

$$g(x) = \int_a^x F'(t)dt - (F(x) - F(a))$$

is zero for all $x \in [a,b]$. We can check that g is continuous on $[a,b]$, and is differentiable on some open interval containing $[a,b]$. In fact, we have

$$g'(x) = F'(x) - F'(x) = 0$$

after applying part (a). Thus by the mean value theorem, g is a constant function on $[a,b]$. Because $g(a) = 0$, we see that $g(b) = 0$. □

We have thus finished the logical foundation of real analysis, starting from the Peano axioms in 55a.

3.2 Partial and total derivatives

Let $f(x,y)$ be a continuous function with two variables x and y.

Definition 3.3. We define the **partial derivatives** as

$$D_1 f(x,y) = \frac{\partial}{\partial x} f(x,y) = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}$$
$$D_2 f(x,y) = \frac{\partial}{\partial y} f(x,y) = \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h}.$$

We want to prove something like $D_2 D_1 f = D_1 D_2 f$.

There is also a notion of total differentiation. But let us first define the limit in higher dimension.

Definition 3.4. Let $f : (a,b) \times (c,d) \to \mathbb{R}$ be a function, and choose a point $(\xi,\eta) \in (a,b) \times (c,d)$. We say

$$\lim_{(x,y)\to(\xi,\eta)} f(x,y) = L$$

if given any $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x,y) - L| < \epsilon$ whenever $0 < \|(x,y) - (\xi,\eta)\| < \delta$.

Definition 3.5. We say that $f(x,y)$ is continuous at (ξ,η) if and only if

$$\lim_{(x,y)\to(\xi,\eta)} f(x,y) = f(\xi,\eta).$$

Differentiation can be viewed as an approximation to order higher than 1. By definition, we have $f'(a) = \lim_{x\to a}(f(x) - f(a))/(x-a)$. If we write

$$f(x) = f(a) + f'(x)(x-a) + E(x)(x-a)$$

then this just means that $\lim_{x\to a} E(x) = 0$. That is, the derivative $f'(x)$ is just an approximation of f as a linear function.

This interpretation can be also applied in higher dimensions. the total differentiation of f at (a,b) is a linear polynomial in x and y of degree at most 1 that approximates f to order higher than 1. That is, it is the tangent plane.

Definition 3.6. A function $f(x, y)$ is (totally) **differentiable** if and only if there exists a polynomial $C + A(x - a) + B(y - b)$ such that

$$\lim_{(x,y)\to(a,b)} \frac{f(x, y) - (C + A(x - a) + B(y, b))}{\sqrt{(x - a)^2 + (y - b)^2}} = 0.$$

If you specialize to $y = b$, then we get the partial derivative. That is,

$$\frac{\partial f}{\partial x}(a, b) = A, \quad \frac{\partial f}{\partial y}(a, b) = B.$$

That is, if we have the total derivative, then we automatically get the partial derivatives. The next question is when will we be able to do the converse. We will need additional assumptions.

But I want to say something about complex functions.

Definition 3.7. Let $f : \mathbb{C} \to \mathbb{C}$ be a complex function. Then

$$f'(c) = \lim_{z\to c} \frac{f(z) - f(c)}{z - c}.$$

The complex version is far more restrictive than the $\mathbb{R}^2 \to \mathbb{C}$ one, because

$$(D + iE) + (P + iQ)(z - c) = (D + iE) + (P + iQ)(x - a) + (iP - Q)(y - b)$$

where $c = a + ib$ and $z = x + iy$. This links to the J operator we were looking at in the previous semester.

Let $f(x+yi) = u(x, y)+iv(x, y)$ be a differentiable function in the complex sense. Then we can look at the 2×2 matrix

$$T = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}.$$

This should satisfy

$$JT = TJ, \quad \text{where } J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

4 February 4, 2016

We introduced differentiation and integration. We now go into the case of high dimension. We recall that a function $f(x, y)$ is differentiable at (a, b) if there are A and B such that

$$\lim_{(x,y)\to(a,b)} \frac{f(x,y) - f(a,b) - A(x-a) - B(y-b)}{\sqrt{(x-a)^2 + (y-b)^2}} = 0.$$

This linearizes the function, and we can compose them as

$$L(f \circ g) = (Lf) \circ (Lg).$$

This is left as an homework.

4.1 Total differentiability from partial differentiability

Now we look at the relation between total and partial differentiation. If f is totally differentiable, then we can restrict the domain to one line to get partial differentiation. But the existence of partial differentiation does not always imply total differentiability. We need additional assumptions to show that the error

$$E(x,y) = f(x,y) - \Big(f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b)\Big)$$

is small enough. We estimate this error using the "TV-antenna" method. Using the mean value theorem, we have

$$\begin{aligned} f(x,y) - f(a,b) &= \big(f(x,y) - f(x,b)\big) + \big(f(x,b) - f(a,b)\big) \\ &= (D_2 f)(x,\eta)(y-b) + (D1f)(\xi,b)(x-a) \end{aligned}$$

for some $b < \eta < y$ and $a < \xi < x$, and thus

$$E(x,y) = ((D_1 f)(x,\eta) - (D_1 f)(a,b))(x-a) + ((D_2 f)(\xi,b) - (D_2 f)(a,b))(y-b).$$

If we assume that both $D_1 f$ and $D_2 f$ are continuous at (a, b), there is an δ such that

$$|(D_j f)(\sigma,\tau) - (D_j f)(a,b)| < \epsilon$$

for any $\sqrt{(\sigma-a)^2 + (\tau-b)^2} < \delta$. Then we see that

$$\Big|\frac{E(x,y)}{\sqrt{(x-a)^2 + (y-b)^2}}\Big| < 2\epsilon$$

if $\sqrt{(x-a)^2 + (y-b)^2} < \delta$. Thus we have the following theorem.

Theorem 4.1. *Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is partially differentiable, and assume that $D_1 f$ and $D_2 f$ are continuous. Then f is totally differentiable.*

4.2 Commutativity of differentiation

Theorem 4.2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose that D_2D_1f is continuous at (a,b) and that D_2f exists in a neighborhood of (a,b). Then $D_2D_1f = D_1D_2f$ at (a,b).*

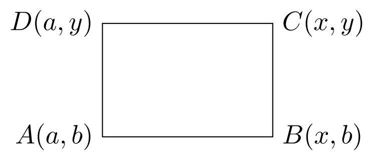
Let me diverge first. The key point is orientation and boundary. Let V be a vector space of dimension n over $\mathbb{R}$. An orientation is an ordered basis $e_1, \dots, e_n$, or an element $e_1 \wedge \cdots \wedge e_n \in \bigwedge^n V = \mathbb{R}$. Consider a triangle $P_0P_1P_2$. Once we write it as $P_0P_1P_2$, we have given it an orientation. The boundary of it can be given as

$$\partial(P_0P_1P_2) = P_1P_2 - P_0P_2 + P_0P_1.$$

More generally, we have

$$\partial(P_0 \cdots P_n) = \sum_{j=0}^{n} (-1)^j (P_0 \cdots \hat{P}_j \cdots P_n).$$

Now the key is that the boundary of the boundary is always empty.



Example 4.3. Consider a rectangle $\square ABCD$. Its boundary is

$$\partial(\square ABCD) = AB + BC + CD + DA,$$

and its boundary

$$\partial^2(\square ABCD) = (B - A) + (C - B) + (D - C) + (A - D) = 0$$

disappears.

Now what has it to do with the commutation of derivatives? The reason it works is because

$$\partial(BC + DA) = \partial(BA + DC) = A + C - B - D.$$

Theorem 4.4 (Weaker version)**.** *Let $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose that both D_2D_1f and D_1D_2f are continuous at (a,b). Then $D_2D_1f = D_1D_2f$ at (a,b).*

Proof. We look at

$$f(x,y) - f(a,y) + f(a,b) - f(x,b) = (f(x,y) - f(a,y)) - (f(x,b) - f(a,b)).$$

Looking $f(x,-) - f(a,-)$ as a function, we can apply the mean value theorem twice an get

$$\begin{aligned}(f(x,y)-f(a,y))-(f(x,b)-f(a,b)) &= (D_2(f(x,\eta)-f(a,\eta))(y-b) \\ &= (D_1D_2f)(\xi,\eta)(x-a)(y-b)\end{aligned}$$

for some $a<\xi<x$ and $b<\eta<y$. Likewise, we get

$$\begin{aligned}(f(x,y)-f(x,b))-(f(a,y)-f(a,b)) &= (D_1(f(\sigma,y)-f(\sigma,b)))(x-a) \\ &= (D_2D_1f)(\sigma,\tau)(x-a)(y-b)\end{aligned}$$

for some $a<\sigma<x$ and $b<\tau<y$. Now sending $x\to a$ and $y\to b$, we get the result. □

How do we get the stronger version? When we move from the different quotient to the differentiation, there are two choices. First we can, just as we did, use the mean value theorem and then use the continuity, or we can simply take the limit and directly get $f'(a)$. The idea is to use the mean value theorem twice on one side, and to use the limit twice for the other side.

Proof of the stronger version. Using the σ and τ line of the previous proof, we have

$$\frac{1}{x-a}\Big(\frac{f(x,y)-f(x,b)}{y-b} - \frac{f(a,y)-f(a,b)}{y-b}\Big) = (D_2D_1f)(\sigma,\tau).$$

If we send $y\to b$, we immediate get

$$\frac{1}{x-a}\big((D_2f)(x,b)-(D_2f)(a,b)\big) = (D_2D_1f)(\sigma,b).$$

We now send $x\to a$. Because D_2D_1f is continuous, we see that the limit exists. Thus we both get the existence of $D_1D_2f(a,b)$ and that

$$(D_1D_2f)(a,b) = (D_2D_1f)(a,b).$$

□

4.3 Double integration

First let us prove Fubini's theorem from $D_1D_2f = D_2D_1f$.

Theorem 4.5 (Fubini's theroem)**.** *Suppose that $f:[a,b]\times[c,d]\to\mathbb{R}$ is continuous. Then*

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

Before we prove this, we must make sense out of the statement by showing that $\int_{y=c}^{d} f(x,y)dy$ is continuous in x. This is, in other words, the commutativity of the limit and the integral. Because f is continuous, f is uniformly continuous, and we see that

$$\lim_{x\to\xi} f(x,y) = f(\xi,y)$$

converges uniformly. Then given $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x,y) - f(\xi,y)| < \epsilon$$

for any $|x-\xi| < \delta$ and any $c \le y \le d$. Then we have

$$\Big| \int_{y=c}^{d} f(x,y) - \int_{y=c}^{d} f(\xi,y) \Big| < (d-c)\delta$$

and thus we see that the limit and the integral commutes.

Proof. We look at the function

$$\int_{s=a}^{x} \Big(\int_{t=c}^{y} f(s,t)dt \Big) ds, \quad \int_{t=c}^{y} \Big(\int_{s=a}^{x} f(s,t)ds \Big) dt.$$

If we take the D_2D_1 of the left hand side, we have from the fundamental theorem of calculus,

$$D_2D_1 \int_{s=a}^{x} \Big(\int_{t=c}^{y} f(s,t)dt \Big) ds = f(x,y).$$

Nest, we take the D_2D_1 of the right hand side. By using some kind of uniformity argument again, we see that

$$D_2D_1 \int_{t=c}^{y} \Big(\int_{s=a}^{x} f(s,t)ds \Big) dt = D_2 \int_{t=c}^{y} f(x,t)dt = f(x,y).$$

Then we get something like $D_2D_1F = D_2D_1G$. We can write this as $D_2D_1(F - G) = 0$, and checking it at $y = c$, we get $D_1(F-G) = 0$. Then again checking at $x = a$, we get $F - G = 0$. $\square$

5 February 9, 2016

Today I am going to prove Fubini's theorem without using the fundamental theorem of calculus.

5.1 Double integration

Theorem 5.1 (Fubini). *Let $a < b$ and $c < d$ be real numbers, and $f : [a, b] \times [c, d] \to \mathbb{R}$ be a continuous function. Then*

$$\iint_{[a,b]\times[c,d]} f = \int_a^b \Big(\int_c^d f(x, y)dy \Big) dx.$$

But we have to first define the double integral.

Definition 5.2. Let E be a bounded subset of $\mathbb{R}^2$; in particular, let $E \subset (\tilde{a}, \tilde{b}) \times (\tilde{c}, \tilde{d})$. Consider a double partition P of $(\tilde{a}, \tilde{b}) \times (\tilde{c}, \tilde{d})$, which is some x_is and y_is with

$$\tilde{a} \le x_0 < x_1 < \cdots < x_m \le \tilde{b}, \quad \tilde{c} \le y_0 < y_1 < \cdots < y_n \le \tilde{d}.$$

Denote $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$. For a *nonnegative* function $f : E \to \mathbb{R}_{\ge 0}$, we define the lower sum as

$$L(P, f, E) = \sum_{R_{ij} \subset E} \big(\inf_{R_{ij}} f \big)(\text{area of } R_{ij}),$$

and the upper sum as

$$U(P, f, E) = \sum_{R_{ij} \cap E \neq \emptyset} \big(\sup_{R_{ij} \cap E} f \big)(\text{area of } R_{ij}).$$

If the supremum of the lower sum and the infimum of the upper sum are equal, we say that f is **double integrable**, and let its value be the **double integration**. If f is not nonnegative, we represent f as a difference of two nonnegative functions and then calculate each integral.

Outline of proof of Fubini's theorem. We first note that f is uniformly continuous because its is continuous. So we can make a very fine partition so that the difference in any two values in a single rectangle is at most ϵ.

Because I don't want to write down many indices, I will just write the Riemann sums as $\sum_v$, $\sum_h$, and $\sum_{\text{rect}}$. First for a given x, we have

$$\Big| \int_{y=c}^d f(x, y)dy - \sum_v f(x, \bullet) \Big| < \epsilon(d - c).$$

Now when we add these inequalities up as the Riemann sum again, we will get,

$$\Big| \sum_h \int_{y=c}^d f(x, \bullet) - \sum_h \sum_v f(x, \bullet) \Big| < \epsilon(d - c)(b - a). \tag{1}$$

Also, looking $\int_{y=c}^{d} f(x,y)dy$ as a function of x, we will have

$$\Big| \int_{x=a}^{b} \int_{y=c}^{d} f(x,y)dydx - \sum_{h} \int_{y=c}^{d} f \Big| < \epsilon(d-c)(b-a). \qquad (2)$$

Combining (1) and (2), we have

$$\Big| \sum_{h} \sum_{v} f - \sum_{h} \int_{y=c}^{d} f \Big| < 2\epsilon(d-c)(b-a)$$

and thus making ϵ small enough we get the desired result. □

5.2 Idea of Stieltjes integration and differential forms

The motivation for this is to interpret point mass or point charge. Suppose you have point charges on a line and want to calculate the potential. Then would want to calculate $\int_{x=a}^{b} \log(\gamma(x))\delta(x)dx$ where $\delta(x)$ is the charge density. But in the point charge case, δ is not actually a function. So instead of looking at $\int_{x=a}^{b} \log(\gamma(x))\delta(x)dx$, we let $\int \delta(x)dx = g(x)$ and look at something like $\int_{x=a}^{b} \log(\gamma(x))dg(x)$.

The derivative can be interpreted as an approximation by a polynomial of degree at most 1 to order greater than 1.

Definition 5.3. A **form** is a homogeneous polynomial. If $f(x,y)$ is a differentiable function, then $df(x,y)$ is the homogeneous part of the polynomial of degree at most 1 which approximates f at (x,y) to order greater than 1. This is referred to as a **differential** **1-form**.

Note that given a point (x,y), the derivative $df : \mathbb{R}^2 \to \mathbb{R}$ is an $\mathbb{R}$-linear map. For the function $f \equiv x$, we can look at its derivative dx. Also, we can look at dy. The chain rule then states that

$$(df)(x,y) = \Big(\frac{\partial f}{\partial x}\Big)_{(x,y)} (dx)_{(x,y)} + \Big(\frac{\partial f}{\partial y}\Big)_{(x,y)} (dy)_{(x,y)}.$$

The Leibniz notation is $e_1 = \frac{\partial}{\partial x}$ and $e_2 = \frac{\partial}{\partial y}$. This notation identifies the directional derivative and the vector. Using the notation, we can write things like

$$\Big(\frac{\partial f}{\partial x}\Big)_{(x,y)} = (df)_{(x,y)} \Big(\frac{\partial}{\partial x}\Big)_{(x,y)}, \quad (dx)_{(x,y)} \Big(\frac{\partial}{\partial x}\Big)_{(x,y)} = 1.$$

We can have things more complicated. If we have a function $f : U \to G$ and $g : G \to \Omega$ defined on open sets of U, G, Ω of $\mathbb{R}^m, \mathbb{R}^n, \mathbb{R}^p$, the chain rule states that

$$d(g \circ f) = (dg) \circ (df).$$

6 February 11, 2016

6.1 Stieltjes integration

There are three motivations for introducing the Stieltjes integral. The first is to unify the approach of point charges and continuous charge density. The second is to introduce Lebesgue integral. The third is to look at differential forms.

Definition 6.1. Let f and g be continuous functions on $[a, b]$. Given a partition P

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b,$$

we define the **upper sum**

$$U(P, f, g) = \sum_{j=1}^{n} \big(\sup_{[x_{j-1}, x_j]} f \big)(g(x_j) - g(x_{j-1}))$$

and likewise the **lower sum** $L(P, f, g)$ for inf. If $\sup L(P, f, g) = \inf U(P, f, g)$, then we define the **Stieltjes integration** as $\int f dg = \sup L(P, f, g)$.

Theorem 6.2. *Let f be continuous on $[a, b]$ and let g be a nondecreasing function on $[a, b]$. Then $\int f dg$ exists.*

But this nondecreasing condition is too strong. So we relax this condition in the following way.

Definition 6.3. A continuous function $g : [a, b] \to \mathbb{R}$ is called to have **bounded variation** if $g = g_1 - g_2$ for some nondecreasing functions g_1 and g_2, or equivalently,

$$\sup_P \sum_{j=1}^{n} |g(x_j) - g(x_{j-1})| < \infty.$$

The equivalence part will be left as a homework problem.

If g is differentiable and g' is continuous, we have

$$\int f dg = \int f g'.$$

Also in the special case $f = h \circ g$, we have

$$\int_a^b f dg = \int_a^b (h \circ g) g' = \int_{g(a)}^{g(b)} h,$$

that is, if g' is continuous and positive. In this integration, the role for g is simply a change of variables.

That property gives rise to the notion of differential forms. We look at integration on a curve. Let $\varphi : [a, b] \to \mathbb{R}^n$ be a continuous function, and let C be its image. Consider any continuous functions $F_1, \dots, F_N, G_1, \dots, G_N : \mathbb{R}^n \to \mathbb{R}$. We define

$$\int_C \sum_{j=1}^N F_j dG_j = \int_{t=a}^b \sum_{j=1}^N (F_j \circ \varphi) d(G_j \circ \varphi).$$

Then this is invariant under a reparametrizaion, and we define $\sum_{j=1}^N F_j dG_j$ as a differential form on an open set $U \subset \mathbb{R}^n$. We will get back to this later.

6.2 Higher-dimensional differentiation

A differentiable function on an n dimensional space is a function that is approximated by a polynomial of degree at most 1 to an order greater than 1. When given m functions $f_1, \dots, f_m$ on an open set U of $\mathbb{R}^n$, we can construct a function $f : U \to \mathbb{R}^m$ defined as $f = (f_1, \dots, f_n)$.

Definition 6.4. The **derivative** of f at a, denoted by $(df)(a)$, is a $\mathbb{R}$-linear map $\mathbb{R}^n \to \mathbb{R}^m$, which is the homogeneous part of the approximate polynomial. That is, if $f(x) = f(a) + ((df)(a))(x - a) + E(x)$, then

$$\lim_{x \to a} \frac{\|E(x)\|_{\mathbb{R}^m}}{\|x - a\|_{\mathbb{R}^n}} = 0.$$

The chain rule can be stated like this.

Theorem 6.5 (Chain rule). *Let $a \in U \subset \mathbb{R}^n$ be an open set and let $W \subset \mathbb{R}^m$ also be open. Consider functions $f : U \to W$ and $g : W \to \mathbb{R}^l$, where $f(a) = b$. If f is differentiable at a and g is differentiable at b, then $g \circ f$ is differentiable at a, and moreover,*

$$(d(g \circ f))(a) = ((dg)(b)) \circ ((df)(a)).$$

We introduce a notational convention of regrind a vector as a differential operator. Let $v \in \mathbb{R}^n$ where $v = (v_1, \dots, v_n)$, and U be an open set in $\mathbb{R}^n$ with $a \in U$. Consider a line L in $\mathbb{R}^n$ through a in the direction of v. Now given a function φ, we can associate a real number as

$$\varphi \mapsto \frac{d}{dt} \varphi(a_1 + v_1 t, \dots, a_n + v_n t)|_{t=0} = \sum_{j=1}^n v_j \frac{\partial \varphi}{\partial x_j}\Big|_a.$$

Thus v can be identified with the differential operator $\xi : \varphi \mapsto \sum_{j=1}^n v_j \frac{\partial \varphi}{\partial x_j}|_a = \xi(\varphi)$. The inverse process is also possible. Given a differential operator ξ, we can let $v_j = \xi(\varphi)$ for $\varphi \equiv x_j$. Thus, v can be identified with the operator ξ.

Now note that this gives a coordinate free description of a vector. Given a map $f : U \to \mathbb{R}^m$, we can describe the differential $(df)(a)$ in a coordinate free manner. How do we do this? Let $f(U) \subset W$ be an open set in $\mathbb{R}^m$. For a function ψ on W, we can simply pull back the map as

$$\big(((df)(a))(\xi)\big)(\psi) = \xi(\psi \circ f).$$

6.3 Differential forms

Now let us get back to differential forms. If F is a real function on an open set $U \subset \mathbb{R}^n$, then the derivative

$$(dF)(a) : \mathbb{R}^n \to \mathbb{R}$$

is a $\mathbb{R}$-linear map. That is, it is an element of $(\mathbb{R}^n)^* = \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$. We can write

$$(dF)(a) = \sum_{j=1}^{n} \Big(\frac{\partial F}{\partial x_j}\Big)_a (dx_j)(a).$$

Because a vector is identified with a differential operator, we can say that

$$\Big(\frac{\partial}{\partial x_1}\Big)_a, \dots, \Big(\frac{\partial}{\partial x_n}\Big)_a \in \mathbb{R}^n$$

is an $\mathbb{R}$-basis for $\mathbb{R}^n$. On the other hand,

$$(dx_1)_a, \dots, (dx_n)_a \in (\mathbb{R}^n)^*$$

is the dual basis.

We define the tangent space $T_{\mathbb{R}^n,a}$ as the the space of differential operators. That is,

$$T_{\mathbb{R}^n,a} = \mathbb{R}\Big(\frac{\partial}{\partial x_1}\Big)_a \oplus \cdots \oplus \Big(\frac{\partial}{\partial x_n}\Big)_a.$$

Its dual space can be described as

$$\big(T_{\mathbb{R}^n,a}\big)^* = \mathbb{R}(dx_1)_a \oplus \cdots \oplus \mathbb{R}(dx_n)_a.$$

As I have explained, a form is a homogeneous polynomial. A 1-form is a homogeneous polynomial of degree 1, but for general k, In this context, a k-form is a homogeneous polynomial of degree k that is alternating. So for general k, a differential k-form looks like

$$\sum_{j_1,\dots,j_k} a_{j_1,\dots,j_k} (dx_{j_1})_a \wedge \cdots \wedge (dx_{j_k})_a \in \textstyle\bigwedge^k (T_{\mathbb{R}^n,a})^*.$$

Then we can do all the things we did last semester.

Now what do this buy us? We can now do the extra computation and generalize the fundamental theorem of calculus to higher dimensions. We first look at the fundamental theorem of calculus in the context of Stieltjes integration. The theorem can be stated as

$$\int_{[a,b]} df = f(b) - f(a),$$

where df is a differential 1-form at each point. The function f, by notational convention, is a differential 0-form.

Intuitively this means that the evaluation of df on $[a,b]$ is the evaluation of f on the 2 points a,b which form the boundary of $[a,b]$. If we generalize this an evaluation of a differential k-form on an k-dimensional object will have to be the evaluation of a $(k-1)$-form on the boundary of the object. We know how the boundary is defined, and now the problem is how to get the k-form from the $(k-1)$-form.

Let us get back to the rectangle case. Let $R = [a,b] \times [c,d]$. Then we see that

$$\begin{aligned} \partial R &= [(a,c),(b,c)] + [(b,c),(b,d)] + [(b,d),(a,d)] + [(a,d),(a,c)] \\ &= (\partial[a,b]) \times [c,d] - [a,b] \times (\partial[c,d]). \end{aligned}$$

(This is somewhat analogous to the Lepnitz formula $d(fg) = (df)g + f(dg)$, except for the sign.) This motivates the definition $d(\varphi \wedge \psi) = (d\varphi) \wedge \psi + (-1)^k \varphi \wedge (d\psi)$, where φ is a k-form and ψ is a l-form.

Definition 6.6. Let ω be a be a k-form on an open set U of $\mathbb{R}^n$. Then we can write

$$\omega = \sum_{1 \le j_1 < \cdots < j_k \le n} f_{j_1,\ldots,j_k} dx_{j_1} \wedge \cdots \wedge dx_{j_k}$$

where $f_{j_1,\ldots,j_k}$ are functions on U. Then we define its **derivative** as

$$d\omega = \sum_{1 \le j_1 < \cdots < j_k \le n} \left(df_{j_1,\ldots,j_k}\right) \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_k}.$$

Actually, we have to justify that this is independent of basis. That is, given any $\xi_1, \ldots, \xi_{k+1} \in T_{\mathbb{R},a}$, we need to check that $(d\omega)(\xi_1, \ldots, \xi_{k+1})$ is well-defined. We only have ω, which can evaluate only k vectors at once. So we need a way to reduce $k+1$ vectors to k vectors.

Let us look at the case $k = 1$. We have ω, which evaluates a single vector, and we are asked to make a 2-form that evaluates two vectors ξ, η. We note that ξ is a differential operator, sos if we put in a function f, it gives another function $\xi(f)$. Also, we have

$$\xi(fg)(a) = f(a)\xi(g) + \xi(f)g(a) \in \mathbb{R}.$$

Then we can look at the Lie operator

$$[\xi, \eta]f = \xi(\eta f) - \eta(\xi f)$$

and define everything. This is the idea of Cartan's formula, and we will do it next time.

7 February 16, 2016

We know that

$$\int_{[a,b]} df = f(b) - f(a)$$

in the sense of Stieltjes integration. We are trying to generalize this as

$$\int_{\text{config.}} \text{derivative of an obj.} = \int_{\text{boundary of config.}} \text{obj.}$$

Let ω be a 1-form on $U \subset \mathbb{R}^n$, and consider a curve Γ given by the map $\varphi : [a,b] \to U$. If we assume that φ is continuously differentiable. Let $\omega = \sum_j f_j dg_j$. Then we can define

$$\int_\Gamma \omega = \int_{[a,b]} \sum_j (f_j \circ \varphi) d(g_j \circ \varphi) = \sum_{[a,b]} \varphi^* \omega,$$

where φ^* is simply the pull-back map. (That is, the substitution map.)

7.1 Cartan's formula

We continue from last time. This Cartan's formula define the differential of a (differential) form in a coordinate free manner. Let us trace the development of the formula. Forms are dual to vectors. We want to get a 2-form from a 1-form. In the dual version, this is the same as getting a single vector from two vectors.

Now we note that vectors are actually directional differentiations. Sophus Lie defined the Lie bracket as

$$[\xi, \eta] f = \xi(\eta f) - \eta(\xi f).$$

It follows that

$$(df)([\xi, \eta]) = \xi((df)(\eta)) - \eta((df)(\xi))$$

and if we assume $d^2 f = 0$, then

$$0 = d(df) = \xi((df)(\eta)) - \eta((df)(\xi)) - (df)([\xi, \eta]).$$

This motivates us to define

$$2(d\omega)(\xi, \eta) = \xi(\omega(\eta)) - \eta(\omega(\xi)) - \omega([\xi, \eta]).$$

Definition 7.1 (Cartan's formula)**.** Let ω be a k-form. Then we define

$$\begin{aligned}(k+1) d\omega(\xi_1, \dots, \xi_{k+1}) = &\sum_{j=1}^{k+1} (-1)^{j+1} \xi_j (\omega(\xi_1, \dots, \widehat{\xi_j}, \dots, \xi_{k+1})) \\ &+ \sum_{1 \le j < l \le k+1} (-1)^{j+l} \omega([\xi_j, \xi_l], \xi_1, \dots, \widehat{\xi_j}, \dots, \widehat{\xi_l}, \dots, \xi_{k+1}).\end{aligned}$$

We can now state Stokes's theorem.

7.2 Stokes's theorem

As we have discussed, we can define the integration of a 1-form over a curve C. If it is given by $\varphi : [a, b] \to U \subset \mathbb{R}^n$, we define

$$\int_C \omega = \int_{[a,b]} \varphi^* \omega.$$

Now let G be an open set in $\mathbb{R}^k$, and let D be the image of φ, which is a continuous map $\varphi : G \to U \subset \mathbb{R}^n$. For a k-form ω on U, we define

$$\int_D \omega = \int_G \varphi^* \omega.$$

Theorem 7.2 (Stokes's theorem)**.** *Let D be an open domain in $\mathbb{R}^k$ and let ω be a $(k-1)$-form. Then*

$$\int_D d\omega = \int_{\partial D} \omega.$$

Example 7.3. Let us look at the example

$$D = \{a \le x \le b, g(x) \le y \le h(x)\} \subset \mathbb{R}^2.$$

Let $\omega = P(x, y)dx$. Then

$$d\omega = \frac{\partial P}{\partial y} dy \wedge dx$$

and thus

$$\int_D d\omega = \int_{x=a}^{b} \int_{y=g(x)}^{h(x)} \frac{\partial P}{\partial y} dy dx = \int_{x=a}^{b} (P(x, h(x)) - P(x, g(x)))dx = \int_{\partial D} \omega.$$

7.3 Other formulations

Let us look at the 2-dimensional version of Stokes's theorem, when it was not formulated in terms of differential forms. Let $D \subset \mathbb{R}^2$ and let

$$\omega = P(x, y)dx + Q(x, y)dy = \begin{pmatrix} P(x, y) \\ Q(x, y) \end{pmatrix} = \vec{v}.$$

Then

$$d\omega = \Big(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\Big) dx \wedge dy$$

and then

$$\int_{\partial D} \omega = \int_{x=0}^{\ell} (P \tfrac{dx}{ds} + Q \tfrac{dy}{ds}) ds = \int_{\partial D} (\vec{v} \cdot \vec{T}) d.$$

Thus we can formulate the Stokes's theorem as

$$\int_{\partial D} (\vec{v} \cdot \vec{T}) dx = \int_D \operatorname{curl} \vec{v}$$

where

$$\text{curl} = \begin{vmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{vmatrix}.$$

We can formulate the Stokes's theorem in terms of the Hodge star operator.[1] For simplicity, let us look at the case $k = 2$. Then we have

$$\vec{e}_1 = \frac{\partial}{\partial x}, \quad \vec{e}_2 = \frac{\partial}{\partial y}, \quad V = \mathbb{R}^2 = T_{\mathbb{R}^2}.$$

Let us choose the orientation as $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$. Then we have

$$\begin{cases} *\vec{e}_1 = \vec{e}_2, \\ *\vec{e}_2 = -\vec{e}_1, \end{cases} \qquad \begin{cases} *(dx) = -dy, \\ *(dy) = dx. \end{cases}$$

Then the statement

$$\int_D d(*\omega) = \int_{\partial D} *\omega$$

just becomes

$$\int_{\partial D} (\vec{v} \cdot \vec{n}) ds = \int_D \operatorname{div} \vec{v}.$$

[1] We defined this in the first semester.

8 February 18, 2016

We look at special cases first. Suppose that we are looking at the region

$$\{a \le x \le b,\ f(x) \le y \le g(x)\}.$$

If $\omega = P(x,y)dx$ then $d\omega = -\frac{\partial P}{\partial y} dx \wedge dy$. Then we have

$$\int_{x=a}^{b} \int_{y=f(x)}^{g(x)} -\frac{\partial P}{\partial y} dx \wedge dy = -\int_{x=a}^{b} \big(P(x,g(x)) - P(x,f(x))\big)dx = \int_{\partial D} P dx.$$

Likewise, if the region looks like

$$\{c \le y \le d,\ \sigma(x) \le y \le \tau(x)\}$$

then the theorem is proved easily.

The general case can be proved by cutting up the manifold into many pieces. Given a region, we can cut it up into many rectangular pieces, and they can be added up. This requires the use of the implicit function theorem, which we haven't covered yet. It basically says that an implicit function can be locally reparameterized as a function of form $f(x) = y$.

8.1 Stokes's theorem in a 2-dimensional space

Let us look at the 2-dimensional case, and let $\omega = Pdx + Qdy$. We consider an inner product on $V = \mathbb{R}^2$, and then V becomes self-dual. We recall that the tangent vector space is identified with its dual, i.e.,

$$\mathbb{R}\frac{\partial}{\partial x} \oplus \mathbb{R}\frac{\partial}{\partial y} = \mathbb{R}dx \oplus \mathbb{R}dy.$$

Then ω can be identified with

$$\vec{v} = P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y}.$$

There are two different ways we can apply the Stokes's theorem in two dimensions.

$$\int_D d\omega = \int_{\partial D} \omega, \quad \int_D d(*\omega) = \int_{\partial D} (*\omega).$$

Let us first look at the left one. Using the identification, we can write it as

$$\iint_D \Big(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\Big) dxdy = \int_{\partial D} Pdx + Qdy = \int_{s=0}^{\ell} \Big(P\frac{dx}{ds} + Q\frac{dy}{ds}\Big) ds = \int_{\partial D} (\vec{v} \cdot \vec{T}) ds,$$

where ∂D is parametrized by arc-length with the variable s, and

$$\vec{T} = \Big(\frac{dx}{ds}, \frac{dy}{dx}\Big)$$

is the unit tangent vector. If we define

$$\mathrm{curl} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$$

then we can write the whole thing simply as

$$\int_D \mathrm{curl}\,\vec{v} = \int_{\partial D} (\vec{v}\cdot\vec{T})ds.$$

Let us now look at the next application. We set the orientation to be

$$*(dx) = -(dy), \quad *(dy) = (dx).$$

Then we see that

$$*\omega = -Pdy + Qdx, \quad d(*\omega) = \Big(-\frac{\partial Q}{\partial y} - \frac{\partial P}{\partial x}\Big)dx\wedge dy$$

and thus Stokes's theorem tells us

$$\int_D \Big(-\frac{\partial Q}{\partial y} - \frac{\partial P}{\partial x}\Big)dxdy = \int_{\partial D}(-Pdy + Qdx) = \int_{\partial D}\Big(-P\frac{dy}{ds} + Q\frac{dx}{ds}\Big)ds.$$

If we let

$$\vec{n} = -*\vec{T} = \Big(\frac{dy}{ds}, -\frac{dx}{ds}\Big) \text{ and } \mathrm{div} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y},$$

then we see

$$\int_D \mathrm{div}\,\vec{v} = \int_{\partial D}(\vec{v}\cdot\vec{n})ds.$$

8.2 Stokes's theorem in a 3-dimensional space

We can apply the Stokes's theorem to

(1) a surface D inside $\mathbb{R}^3$ with piecewise smooth boundary, or

(2) a domain $D \subset \mathbb{R}^3$ with piecewise smooth boundary.

For each case, we can apply to the original form ω or $*\omega$. For a 1-form ω, the theorem becomes

$$\int_D \big((\mathrm{curl}\,\vec{v})\cdot\vec{n}\big)dA = \int_{\partial D}(\vec{v}\cdot\vec{T})ds.$$

This is called the curl theorem. Here,

$$dA = \sqrt{(dy\wedge dz)^2 + (dz\wedge dx)^2 + (dx\wedge dy)^2}$$

is a nonnegative function defined on $\bigwedge^2 T_{\mathbb{R}^3}$.

Let me explain more about this. Consider a map $\varphi : \mathbb{R}^2 \supset G \to D$ where G has coordinates u and v. Then φ induces a map $\bigwedge^2 T_{\mathbb{R}^2} \to \bigwedge^2 T_{\mathbb{R}^3}$. And then the area of $\varphi(G)$ is given by

$$\int \big((d\varphi)(du\wedge dv)\big)dA.$$

8.3 Stokes's theorem in an n-dimensional space

Let ω be any 1-form. Then we see that $*\omega$ is an $(n-1)$-form, and $d(*\omega)$ will be an n-form.

Definition 8.1. Consider a vector field $\vec{v}$. This corresponds to a 1-form ω. Then we define

$$d(*\omega) = (\operatorname{div} \vec{v}) dx_1 \wedge \cdots \wedge dx_n.$$

If we let $\omega = f_1 dx_1 + \cdots + f_n dx_n$, then we have

$$*\omega = \sum_{j=1}^{n} (-1)^{j-1} f_j (dx_1 \wedge \cdots \wedge dx_{j-1} \wedge dx_{j+1} \wedge \cdots \wedge dx_n)$$

and then

$$d(*\omega) = \sum_{j=1}^{n} \frac{\partial f_j}{\partial x_j} x_1 \wedge \cdots \wedge x_n.$$

Now consider a domain $D \subset \mathbb{R}^3$. When we apply the Stokes's theorem to ω and D, we get

$$\int_D \operatorname{div} \vec{v} = \int_D d(*\omega) = \int_{\partial D} *\omega = \int_{\partial D} (\vec{v} \cdot \vec{n}) dV_{n-1},$$

where

$$dV_{n-1} = \sqrt{\sum_{j=1}^{n} (dx_1 \wedge \cdots \wedge dx_{j-1} \wedge dx_{j+1} \wedge \cdots \wedge dx_n)^2}.$$

8.4 Stokes's theorem and the Maxwell equations

We note that

$$\operatorname{curl} = *d, \quad \operatorname{div} = d * .$$

This actually implies that

$$\operatorname{div} \operatorname{curl} \omega = (d*)(*d)\omega = -d^2\omega = 0.$$

Anyways, let us look at the Maxwell equations.

$$\begin{cases} \operatorname{div} E = 4\pi\rho & \operatorname{curl} E = -\frac{1}{c}\frac{\partial B}{\partial t} \\ \operatorname{div} B = 0 & \operatorname{curl} B = \frac{1}{c}\frac{\partial E}{\partial t} + \frac{4\pi}{c} J \end{cases}$$

There was originally no term $\frac{1}{c}\frac{\partial E}{\partial t}$ in the equation, but Maxwell observed that $\operatorname{div} \operatorname{curl} B$ has to be zero and added the correction term. When in vacuum, there have $\operatorname{curl} \operatorname{curl} \vec{v} = -\Delta \vec{v}$ for $\vec{v} = E, B$, and see that the speed of light is c.

9 February 23, 2016

9.1 Implicit function theorem

Let us take the 2-dimensional case first. Let f be a function with $(0,0)$. We want to ask whether the solution set of $f(x,y)=0$, which will probably look like a curve, can be parametrized. We parametrize it by one of the coordinates, so that it will look like $(x,g(x))$. The main tools we will use are the intermediate value theorem and the mean value theorem.

Theorem 9.1 (Implicit function theorem)**.** *Let $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ be a real function defined on a neighborhood of $(0,0)$. Suppose that first-order derivatives of f are continuous and $(\partial f/\partial y)(0,0) \neq 0$. Then there exists a rectangle $0 \in [a,b]\times[c,d] \subset U$ such that there exists continuously differentiable function $g : (a,b) \to (c,d)$ such that*

$$f(x,g(x)) \equiv 0$$

on $a < x < b$.

Proof. Without loss of generality let $\kappa = (\partial f/\partial y)(0,0) > 0$. Since the derivatives are continuous, we can restrict ourselves to a smaller open subset U such that

$$\frac{\partial f}{\partial y}(x,y) > \frac{\kappa}{2}, \quad \left|\frac{\partial f}{\partial x}(x,y)\right| \leq A$$

for any $(x,y) \in U$.

Consider some small τ such that both $(0,\tau),(0,-\tau) \in U$. Since $\partial f/\partial y$ is always positive, we see that $f(0,\tau) > 0$ and $f(0,-\tau) < 0$. There there exists a $\sigma > 0$ such that

$$f(x,\tau) > 0, \qquad f(x,-\tau) < 0$$

for all $-\sigma \leq x \leq \sigma$. Then by the mean value theorem, there exists a $g(x)$ such that $f(x,g(x)) = 0$, and since it is increasing, it is unique.

We now try to say something about the differentiability of g. We first show that $|g(x)/x|$ is bounded only in terms of κ and A. By the mean value theorem, there exists a $0 < \xi < 1$ such that

$$f(x,g(x)) = f(0,0) + (D_1 f)(\xi x, \xi g(x))x + (D_2 f)(\xi x, \xi g(x))g(x).$$

Since $f(0,0) = f(x,g(x)) = 0$, it follows that

$$\left|\frac{g(x)}{x}\right| < \frac{A}{\kappa/2} < \frac{2A}{\kappa}.$$

We now show that $g'(0)$ exists, and is in fact, $-(D_1 f)(0,0)/(D_2 f)(0,0)$. This is because from the previous equality,

$$\frac{g(x)}{x} = -\frac{(D_1 f)(\xi x, \xi g(x))}{(D_2 f)(\xi x, \xi g(x))},$$

and when we send $x \to 0$, the right hand side goes to

$$\lim_{x\to 0} \frac{g(x)}{x} = -\frac{(D_1 f)(0,0)}{(D_2 f)(0,0)}. \qquad \square$$

Let us now go to more variables. For example, let us just take one more and consider $f(x, y, z) = 0$. Suppose that x and y are independent and z are dependent. Then if $(\partial f/\partial z)(0, 0, 0) \neq 0$, then the zero locus is $z = g(x, y)$ for some differentiable g. I will not repeat the argument.

We now consider the case where there are many dependent variables. Consider the system of linear equations

$$\begin{cases} f(x,y,z) = 0 \\ g(x,y,z) = 0. \end{cases}$$

Let x be the independent variable and y, z be the dependent variables. In this case, the condition becomes $d\Phi$ being nonsingular, where

$$\Phi : (y, z) \mapsto (f, g).$$

How do we prove this? From the nonsingular condition, it follows that one of $\partial f/\partial z$ or $\partial g/\partial z$ at $(0, 0, 0)$ is nonzero. Without loss of generality let $\partial f/\partial z \neq 0$. We first look at f, and there must be a $\varphi(x, y)$ such that $f(x, y, \varphi(x, y)) \equiv 0$. Then we apply the theorem to $g(x, y, \varphi(x, y))$ viewed as a function with variables x and y. A computation shows that the condition is satisfied, and thus everything works out.

9.2 Inverse mapping theorem

Let

$$\begin{cases} u = f(x,y) \\ v = g(x,y) \end{cases}$$

be a function $(x, y) \mapsto (u, v)$. We want to find the inverse of this function. We apply the implicit function theorem to

$$\Phi : (x, y) \mapsto (f(x, y) - u, g(x, y) - v)$$

and we see that the inverse exists if the determinant of the Jacobian is nonzero.

9.3 Cauchy's theory

Cauchy's theory is simply viewing $\mathbb{R}^2$ as an 1-dimensional vector space over $\mathbb{C}$. We recall that if V is an $\mathbb{R}$-vector space of dimension $\dim_{\mathbb{R}} V = m$ and there is a $\mathbb{R}$-linear map $J : V \to V$ such that $J^2 = -\mathrm{id}_V$, then V can be given an complex structure.

We have

$$T_{\mathbb{R}^2} = \mathbb{R}\frac{\partial}{\partial x} \oplus \mathbb{R}\frac{\partial}{\partial y}.$$

We define $J : T_{\mathbb{R}^2} \to T_{\mathbb{R}^2}$ as

$$J\Big(\frac{\partial}{\partial x}\Big) = \frac{\partial}{\partial y}, \quad J\Big(\frac{\partial}{\partial y}\Big) = -\frac{\partial}{\partial x}.$$

If we have an inner product space, we can say that $T_{\mathbb{R}^2} = \mathbb{R}dx \oplus \mathbb{R}dy$ and then

$$J(dx) = -dy, \quad J(dy) = dx.$$

Now one can check that

$$dz = dx + idy$$

is a $\mathbb{C}$-linear map $T_{\mathbb{R}^2} = T_{\mathbb{C}} \to \mathbb{C}$.

We now recall that a complex derivative a function f at c is defined as something like

$$f(z) = f(c) + f'(c)(z-c) + E(z)(z-c)$$

where $\lim_{z\to c} E(z) = 0$. We see that $df : T_{\mathbb{C},c} \to \mathbb{C}$ is then a $\mathbb{R}$-linear function, if f is simply a differentiable function in terms of $\mathbb{R}$. Since there is a decomposition $T_{\mathbb{C},c} = \mathbb{C}dz \oplus \mathbb{C}d\bar{z}$, we see that there always are two complex numbers $a, b \in \mathbb{C}$ such that

$$df(c) = adz + bd\bar{z}.$$

That is, we can always write something like

$$df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar{z}}d\bar{z}.$$

We note that z is complex differentiable at c if and only if $\partial f/\partial \bar{z}$ is zero at c.

Theorem 9.2. *Let D be a bounded domain in $\mathbb{C}$ with piecewise continuously differentiable boundary. Let f be a function on a neighborhood U of $\overline{D}$ such that $f'(z)$ exists and is continuous on U. Then*

$$\int_{\partial D} f(z)dz = 0.$$

Proof. This is because if we let $\omega = f(z)dz$ then $d\omega \equiv 0$. □

So the next thing was that people was unhappy with the condition that f' is continuous. We needed this to apply the fundamental theorem of calculus change it to some integral of a continuous function. But this looks unnecessary because they just cancel out and become zero. Goursat solved this problem by bypassing the real numbers. What does this mean? We approximate f by a polynomial of degree ≤ 1 to an order > 1. Then we cut the domain into small pieces and add them up.

10 February 25, 2016

Suppose that $\omega = f(z)dz$ for a complex differentiable function f. Then since both df and dz are $\mathbb{C}$-linear, we get $df \wedge dz = 0$ and thus $d\omega = 0$. Then directly applying Stokes's theorem, we get the following theorem.

Theorem 10.1 (Cauchy)**.** *If f is complex differentiable at points on $\overline{D}$ and its derivative is continuous, then*

$$\int_{\partial D} f(z)dz = 0.$$

Goursat's theorem generalizes the theorem by dropping the continuously differentiable condition. In practice this is not that useful since if something is complex differentiable then it is infinitely differentiable. But its idea is useful.

10.1 Goursat's theorem

Goursat first proved it for rectangles. Let R_0 be an rectangle and let f be a complex differentiable function on $\overline{R_0}$. We break the rectangle down to smaller rectangles and let $R_0 = \bigcup_\nu R_\nu$. Now we can approximate f on R_ν as

$$f(z) = f(c) + f'(c)(z-c) + E(z)(z-c)$$

for some c and $\lim_{z\to c}|E(z)| = 0$. Now we can calculate the integral of f on the boundary of R_0 as

$$\int_{\partial R_0} f(z)dz = \sum_\nu \int_{\partial R_i} f(z)dz + \sum(\text{error}).$$

Then the first term automatically vanishes, and thus what we want is a sum of a lot of error terms.

Theorem 10.2 (Goursat)**.** *If f is complex differentiable at points on $\overline{D}$, then*

$$\int_{\partial D} f(z)dz = 0.$$

Proof. Assume that $|\int_{\partial R_0} f(z)dz| = c > 0$. Then we can divide the rectangle up into four pieces, and $|\int_{\partial R_1} f(z)dz| \geq c/4$. Then we can divide up into things, and ad infinitum. We thus get a nest of rectangles

$$R_0 \supset R_1 \supset R_2 \supset \cdots \supset R_\nu \supset \cdots \to a.$$

with $|\int_{\partial R_\nu} f(z)dz| \geq c/4^\nu$.

We now look at the approximation

$$f(z) = f(a) + f'(z)(z-a) + E(z)(z-a)$$

and for each $\epsilon > 0$, there is an ν_0 such that $|E(z)| < \epsilon$ on ∂R_ν for any $\nu \geq \nu_0$. Then we can integrate f on the boundary and get

$$\left| \int_{\partial R_\nu} f(z)dz \right| = \left| \int_{\partial R_\nu} E(z)(z-a) \right| < \frac{c \cdot \epsilon}{4^\nu},$$

because the length of ∂R_ν is a constant times $2^{-\nu}$ and the absolute value of $z-a$ is also less than a constant times $2^{-\nu}$. Thus we get a contradiction. □

10.2 Cauchy's integral formula

Definition 10.3. A function f is **holomorphic** on an open set U if and only if f' exists at every point on U.

Proposition 10.4 (Mean value property). *If f is holomorphic on a neighborhood of a disk in $\mathbb{C}$, then the value of f at the center is the average of the value of f on the circle.*

Theorem 10.5 (Cauchy integral formula). *Let f be holomorphic on an neighborhood U of $\overline{D}$, where D is a bounded domain in $\mathbb{C}$ with piecewise continuous differentiable boundary. Assume further that f' is continuous. If $a \in D$, then*

$$f(a) = \frac{1}{2\pi i} \oint_{z \in D} \frac{f(z)dz}{z-a}.$$

Proof. We first note that $f(z)/(z-a)$ is holomorphic on $U \setminus \{a\}$. Consider a small disk

$$\Delta_\epsilon(a) = \{z \in \mathbb{C} : |z-a| < \epsilon\}$$

entirely contained in D. Then

$$\oint_{\partial D} \frac{f(z)dz}{z-a} = \int_{\partial(D \setminus \overline{\Delta_\epsilon(a)})} \frac{f(z)dz}{z-a} dz + \oint_{\partial \Delta_\epsilon(a)} \frac{f(z)}{z-a} dz = \oint_{\partial \Delta_\epsilon(a)} \frac{f(z)}{z-a} dz.$$

Now we can approximate

$$\left| \int_{\partial \Delta_\epsilon(a)} \frac{f(z)dz}{z-a} - \int_{\partial \Delta_\epsilon(a)} \frac{f(a)dz}{z-a} \right| = \left| \int_{\partial \Delta_\epsilon(a)} \frac{f(z)-f(a)}{z-a} dz \right| \to 0$$

as $\epsilon \to 0$. Also, we can calculate

$$\int_{\partial \Delta_\epsilon(a)} \frac{f(a)dz}{z-a} = 2\pi i f(a).$$

Therefore sending $\epsilon \to 0$, we get

$$\oint_{\partial D} \frac{f(z)dz}{z-a} = \oint_{\partial \Delta_\epsilon(a)} \frac{f(z)}{z-a} dz = 2\pi i f(a).$$

□

Proof of the mean value property. Let D be the disk, and assume that the center of D is 0. Then applying the Cauchy integral formula, we have

$$f(0) = \frac{1}{2\pi i}\int_{\partial D}\frac{f(z)dz}{z}.$$

If we parametrize it by $\theta \mapsto (R\cos\theta, R\sin\theta)$, we immediately get

$$f(0) = \frac{1}{2\pi i}\int_{\theta=0}^{2\pi i}\frac{f(Re^{i\theta})Rie^{i\theta}d\theta}{Re^{i\theta}} = \frac{1}{2\pi}\int_{\theta=0}^{2\pi i} f(Re^{i\theta})d\theta. \qquad \square$$

We can now prove the fundamental theorem of algebra.

Theorem 10.6 (Fundamental theorem of algebra)**.** *Let $P(z) = a_nz^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0$ for $a_n \neq 0$ and $n \geq 1$. There there exists some $z_0 \in \mathbb{C}$ such that $P(z_0) = 0$.*

Proof. Suppose not. Then $1/P(z)$ is holomorphic on $\mathbb{C}$. We have

$$\begin{aligned}\Big|\frac{1}{P(0)}\Big| &= \Big|\frac{1}{2\pi i}\int_{\theta=0}^{1}\frac{d\theta}{a_nR^ne^{i\theta n} + \cdots + a_1Re^{i\theta} + a_0}\Big| \\ &\leq \sup_{0\leq\theta\leq 2\pi}\frac{1}{|a_nR^ne^{i\theta n} + \cdots + a_1Re^{i\theta} + a_0|}.\end{aligned}$$

The right hand side goes to zero as $R \to \infty$, and thus we get a contradiction. $\square$

We look at Cauchy's integral formula again.

$$f(z) = \frac{1}{2\pi i}\int_{\zeta\in\partial D}\frac{f(\zeta)d\zeta}{\zeta - z}$$

for all $z \in D$. This suggests that any holomorphic function on a bounded domain can be written as the limit of a linear combination of translates of $1/z$. These are called **kernels**.

Theorem 10.7. *Let $D \subset \mathbb{C}$, and consider a holomorphic function f on D. Let $a \in D$ and r be the distance from a to ∂D. Then*

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)$$

on $\{|z-a| < r\}$ for some $c_0, c_1, \dots$.

Proof. From the Cauchy integral formula, we get

$$f(z) = \frac{1}{2\pi i}\int_{\zeta\in\partial D}\frac{f(\zeta)d\zeta}{\zeta - z} = \int_{\zeta\in\partial D} f(\zeta)\sum_{n=0}^{\infty}\frac{(z-a)^n}{(\zeta-a)^{n+1}}d\zeta.$$

Then we can let

$$c_n = \frac{1}{2\pi i}\int_{\zeta\in\partial D}\frac{f(\zeta)d\zeta}{(\zeta-a)^{n+1}}.$$

There are might be some convergence issues, but it is fine because we have uniform convergence. $\square$

11 March 1, 2016

11.1 Bergman kernel

In the Cauchy kernel, we integrate some function over the boundary of some domain D. It comes from the Laplacian.

The Bergman kernel comes from the idea of a orthogonal projection. Let $\mathcal{H}$ be the $\mathbb{C}$-vector space of square integrable holomorphic functions on D. We can introduce an inner product

$$(f,g) = \int_D f(z)\overline{g(z)}.$$

If we choose an orthonormal basis $f_1, f_2, \dots$ then any function F can be expanded as

$$\begin{aligned} F(z) &= \sum_j (F, f_j) f_j(z) = \sum_j \Big(\int_{\zeta\in D} F(\zeta)\overline{f_j(\zeta)} \Big) f_j(z) \\ &= \int_{\zeta \in D} \Big(\sum_j f_j(z)\overline{f_j(\zeta)} \Big) F(\zeta). \end{aligned}$$

This is simply a speculation since we do not know even whether it converges. Bergman proved that this actually converges, and the kernel is defined as $K_B(z,\zeta) = \sum_j f_j(z)\overline{f_j(\zeta)}$.

11.2 Curvature of a surface

If we take two partial differentiations of a function, they clearly commute. Gauss was the first person to ask the question of taking the differentiation of a vector field of a curved space. Let $S \subset \mathbb{R}^3$ be a surface, and let C be a curve on S given by $P : (-1,1) \to S$. For each t, let $\vec{v}(t)$ be a vector tangential to S. How do we differentiate this vector with respect to t? Actually we can always differentiate its with respect to t because it is simply a vector in $\mathbb{R}^3$. But we want it to still be tangential to S. So we project it to the plane tangent to S. This is called the covariant differentiation ∇_t.

Generally, given a vector field on an open subset of S, we can parametrize the open set with the variables u and v, and let $\vec{r}(u,v)$ be the vector field. Then we can look at the differentiation ∇_u and ∇_v. Does the two differentiation commute, i.e., $\nabla_u\nabla_v\vec{r} = \nabla_v\nabla_u\vec{r}$? In general they don't, because the projection messes everything up.

Now Gauss wanted to define the curvature. On a plane curve, we define the curvature as how much the direction changes. On a surface, the direction can be considered as a normal vector. The Gauss map is the map that sends point to the normal vector at that point in S^2. Let D be a domain near P, and D' be the image of D under the Gaussian map. The Gaussian curvature at P is defined as

$$\lim_{D\to P} \frac{\text{area of } D'}{\text{area of } D}.$$

12 March 3, 2016

12.1 The celebrated theorem of Gauss

Let $X \subset \mathbb{R}^3$ be a surface embedded in 3-space. We want to define the curvature of this. But how do we define the curvature of a plane curve? We define it as

$$\kappa = \frac{d\vec{T}}{ds} \cdot \vec{N},$$

where everything is parameterized by the length of the curve, $\vec{T}$ is the tangent vector, and $\vec{N}$ is the normal vector.

Motivated by this, we define the Gauss map $X \to S^2$ that maps a point to the normal vector at that point. Then we define the **Gaussian curvature** at a point P as

$$\lim_{D \to P} \frac{\Delta A_{S^2}}{\Delta A_X}.$$

More generally, for any hypersurface $X^{(n)}$[2] inside $\mathbb{R}^{n+1}$, we define the scalar curvature as

$$\lim_{D \to P} \frac{\Delta V_{S^n}}{\Delta V_X}.$$

If we want to define the curvature for things of larger codimension, we have to look at the Grassmannians, but we won't do this.

The second definition of the Gaussian curvature is reducing to the plane curve case. Let Π be a plane containing the normal vector $\vec{n}$ at P, and let $C = \Pi \cap X$. We then define the **Gaussian curvature** as

$$\Big(\min_{\Pi} \kappa(C_\Pi)\Big)\Big(\max_{\Pi} \kappa(C_\Pi)\Big).$$

At first glance, they look quite unrelated, but we shall prove that they are equivalent.

The third definition is using the covariant differentiation. We first parametrize the surface by two variables u, and v, so that $\vec{r}_u$ and $\vec{r}_v$ are unit vectors and are orthogonal to each other. Then we define the curvature using the commutator of the covariance differentiations

$$(\nabla_v \nabla_u \vec{r}_u - \nabla_u \nabla_v \vec{r}_u) \cdot \vec{r}_v.$$

Consider a curve on a surface $t \mapsto (u(t), v(t))$. Then we can calculate the length of the curve as

$$\int_{t=a}^{b} \Big\| \frac{d\vec{r}}{dt} \Big\| dt = \int_{t=a}^{b} \Big\| \frac{d\vec{r} \cdot d\vec{r}}{dt \cdot dt} dt.$$

If we let

$$E = \vec{r}_u \cdot \vec{r}_u, \quad F = \vec{r}_u \cdot \vec{r}_v, \quad G = \vec{r}_v \cdot \vec{r}_v,$$

[2] The superscript n just indicates that the manifold is of dimension n.

we see that $d\vec{r}\cdot d\vec{r} = Edu^2 + 2Fdudv + Gdv^2$, which is actually a homogeneous quadratic polynomial defined on $T_P X$. So the arc length only depends on E, F, and G. In fact, we the celebrated theorem of Gauss says the following.

Theorem 12.1 (Theorema Egregium)**.** *The Gaussian curvature is computed by E, F, G up to their*

As a remark, for an n-dimensional manifold, the fundamental form is also an quadratic form on $T_P(X)$.

12.2 Proof of the Theorema Egregium

Consider a plane Π passing through the normal vector $\vec{n}$ at a point P. The tangential vector $\vec{T}(s)$ is given as

$$\vec{T}(s) = \frac{d\vec{r}}{ds}$$

and thus the curvature of the intersection of Π and X is given by

$$\kappa = \frac{d^2\vec{r}}{ds^2}\cdot\vec{n}.$$

Now applying the chain rule multiple times, we see that

$$\kappa = \frac{Ddu^2 + 2D'dudv + D''dv^2}{Edu^2 + 2Fdudv + Gdv^2}$$

for

$$D = \vec{r}_{uu}\cdot\vec{u}, \quad D' = \vec{r}_{uv}\cdot\vec{u}, \quad D'' = \vec{r}_{vv}\cdot\vec{u}.$$

We want to look at the maximum and minimum of this thing, where the ratio $du : dv$ varies. If we take the derivative, we see that they are attained at the points of

$$\begin{cases} (D-\kappa E)du + (D'-\kappa F)dv = 0 \\ (D'-\kappa E)du + (D''-\kappa G)dv = 0. \end{cases}$$

Then we can plug things in and see that

$$\kappa_{\min}\kappa_{\max} = \frac{DD'' - D'^2}{EG - F^2}.$$

Now let us look at the first definition of the Gaussian curvature. Since the area element is $\|\vec{r}_u \times \vec{r}_v\| dudv$, we see that the Gaussian curvature is λ in the equation

$$\vec{n}_u \times \vec{n}_v = \lambda(\vec{r}_u \times \vec{r}_v).$$

Now we have

$$(\vec{n}_u \times \vec{n}_v)\cdot(\vec{r}_u \times \vec{r}_v) = (\vec{n}_u\cdot\vec{r}_u)(\vec{n}_v\cdot\vec{r}_v) - (\vec{n}_u\cdot\vec{r}_v)(\vec{n}_v\cdot\vec{r}_u).$$

How do we compute this? Since $\vec{n} \cdot \vec{r}_u \equiv 0$, we have $\vec{n}_u \cdot \vec{r}_u + \vec{n} \cdot \vec{r}_{uu} = 0$. Using these kind of identities, we get that that what we want to compute is actually $DD'' - D'^2$, and if we do some more calculation, we get

$$\lambda = \frac{DD'' - D'^2}{EG - F^2}.$$

Finally, we need to check that $DD'' - D'^2$ can be expressed in terms of E, F, G, and their first order derivatives. We do this by checking that the covariance differentiation can be expressed by these three things. We note that many vector field can be locally written as $f\vec{r}_u + g\vec{r}_v$. Then its covariant derivative is

$$\nabla(f\vec{r}_u + g\vec{r}_v) = df\vec{r}_u + f\nabla\vec{r}_u + dg\vec{r}_v + g\nabla\vec{r}_v,$$

where the vectors applied the vector fields are interpreted as directional derivatives.

Thus we can imitate this definition and intrinsically defining

$$\nabla_u\vec{r}_u = \vec{r}_{uu} - (\vec{r}_{uu} \cdot \vec{n})\vec{n}.$$

The result is a linear combination of $\vec{r}_u$ and $\vec{r}_v$, and thus we can simply write

$$\nabla_u\vec{r}_u = \Gamma^u_{uu}\vec{r}_u + \Gamma^v_{uu}\vec{r}_v,$$

where the **Christoffel symbols** Γ are defined as

$$\nabla_u\frac{\partial}{\partial u} = \Gamma^u_{uu}\frac{\partial}{\partial u} + \Gamma^v_{uu}\frac{\partial}{\partial v}$$

and blah blah and so forth. Then we get in total 8 Christoffel symbols, and we note that the two lower indices are symmetric, i.e., $\Gamma^\bullet_{uv} = \Gamma^\bullet_{vu}$. Now how do we compute them? We have

$$\vec{r}_{uu} - (\vec{r}_{uu} \cdot \vec{n})\vec{n} = \Gamma^u_{uu}\vec{r}_u + \Gamma^v_{uu}\vec{r}_v$$

and when we take the inner product with $\vec{r}_u$, we get

$$\frac{1}{2}E_u = \frac{1}{2}(\vec{r}_u \cdot \vec{r}_u)_u = \vec{r}_u u \cdot \vec{r}_u = \Gamma^u_{uu}E + \Gamma^v_{uu}F.$$

Now we have a lot of these equations, and when we solve the equation we get the formulas for the Γ's.

Finally, we check that the first and second definitions of the curvature are equivalent to the third definition. That is, we claim that the Gaussian curvature is

$$\pm([\nabla_u, \nabla_v]\vec{r}_u, \vec{r}_v) = \pm\Big([\nabla_u, \nabla_v]\frac{\partial}{\partial u}, \frac{\partial}{\partial v}\Big).$$

We denote $u = x^1$ and $v = x^2$. Then we have

$$\vec{r}_{x^j x^k} - (\vec{r}_{x^j x^k} \cdot \vec{n})\vec{n} = \sum_i \Gamma^i_{jk}\vec{r}_{x^i}.$$

When we take the derivative, we have

$$\frac{\partial}{\partial x^l}(\vec{r}_{x^j x^k} - (\vec{r}_{x^j x^k} \cdot \vec{n})\vec{n}) = \vec{r}_{x^l x^j x^k} - (\vec{r}_{x_j x^k} \cdot \vec{n})\vec{n}_{x^l} \pmod{\vec{n}}.$$

Thus the commutator of the covariant derivatives is

$$\nabla_l \nabla_j \Big(\frac{\partial}{\partial x^k}\Big) - \nabla_j \nabla_l \Big(\frac{\partial}{\partial x^k}\Big) = -(\vec{r}_{x^j x^k}\vec{n})\vec{n}_l - \dots$$

and after some more computations we get the result.

13 March 8, 2016

In the problem set, we saw that there is a unique covariant differentiation ∇_j such that

$$\nabla_j \frac{\partial}{\partial x_k} = \sum_l \Gamma^l_{jk} \frac{\partial}{\partial x_k}$$

with symmetric Christoffel symbols. This is called the **Levi-Cevita connection**. Why suddenly talk about connections? This is because we want to compare vector fields at difference points. To compare two tangent vectors at different points, we see a means to transport a vector at a point to another point in a parallel manner.

13.1 Gauss curvature and parallel transport

Now I want to start talking about the relation between Gauss curvature and parallel transport. This is important because this marked the start of global analysis, such as Morse theory.

Suppose that you have a tangent vector $\vec{v}$ at some point P. Consider a small closed loop C containing P, and let us parallel transport $\vec{v}$ along the curve C. After $\vec{v}$ come back to its original position, it might form a nonzero angle with the original $\vec{v}$. The curvature is then defined as

$$\kappa = \lim_{C \to P} \frac{\text{angle}}{\text{area inside } C}.$$

We can compute the angle by using the first definition of the Gaussian curvature. Under the Gauss map, the parallel transport is conserved, because the normal vector are the same. Then we can use the camera shutter technique and approximate the curve by a polygon, and we easily see the the angle of the parallel transport is the same as the area. Therefore this definition agrees with the original definitions of the curvature.

13.2 Green's formula

We observe

$$\operatorname{div}(f \operatorname{grad} g) = \operatorname{grad} f \cdot \operatorname{grad} g + f \Delta g.$$

Then when we integrate this over Ω and apply the divergence theorem, we have

$$\int_{\partial\Omega} f((\operatorname{grad} g) \cdot \vec{n}) = \int_\Omega \operatorname{grad} f \cdot \operatorname{grad} g + \int_\Omega f \Delta g.$$

Then

$$\int_{\partial\Omega} f(\nabla_{\vec{n}} g) - \int_{\partial\Omega} (\nabla_{\vec{n}} f) g = \int_\Omega f(\Delta g) - \int_\Omega (\Delta f) g.$$

One application of this is the Green's kernel. We want to a Green function G such that $\Delta G = \delta_P$ and $G = 0$ on $\partial\Omega$. This is analogous to the electrostatic setting of a cavity shaped like Ω in a metal conductor that is grounded. Then G becomes the potential function, and plugging this in, we get

$$\int_{\partial\Omega} f(\nabla_{\vec{n}} G) = f(0) - \int_{\Omega} (\Delta f) g.$$

I think I bore people. Let us move on to the next topic.

13.3 Fourier series

The differential equations we are interested in are linear, and has constant coefficients. We use the functions e^{inx} for $n \in \mathbb{Z}$ as building blocks. This is good, because

$$\frac{d}{dx} e^{inx} = ine^{inx}.$$

So Fourier's method is the algebraization of problems in analysis. If we have a differential operator acting on a building block, we get

$$\Big(\sum_{j=0}^{m} a_j \frac{d^j}{dx^j}\Big) e^{inx} = \Big(\sum_{j=0}^{m} a_j (in)^j\Big) e^{inx}.$$

The problem is that it is not rigorous. We just let $f(x) = \sum c_n e^{inx}$ and suppose that

$$\Big(\sum_{j=0}^{n} \frac{d^j}{dx^j}\Big) f(x) = \sum_{n\in\mathbb{Z}} \Big(c_n \sum_{j=0}^{m} a_j (in)^j\Big) e^{inx}.$$

Now Lebesgue comes into the picture.

The keys is orthonormality of e^{inx}. We easily see that

$$\int_{x=-\pi}^{\pi} e^{imx} \overline{e^{inx}} dx = 2\pi \delta_{mn}.$$

Hence we see that $(1/\sqrt{2\pi}) e^{inx}$ are orthonormal. Then we can write

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

Because we don't want to look at 2π, we write

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

and consider $\sum c_n e^{inx}$ (and hope that it is equal to f). This c_n is called the **nth coefficient of the Fourier series of f.**

Assume that f is continuous on $[-\pi, \pi]$ with $f(-\pi) = f(\pi)$. We have

$$\begin{aligned}\sum_{k=-n}^{n} c_k e^{ikx} &= \sum_{k=-n}^{n} \Big(\frac{1}{2\pi}\int_{y=-\pi}^{\pi} f(y)e^{-iky}dy\Big)e^{ikx} \\ &= \frac{1}{2\pi}\int_{y=-\pi}^{\pi} f(y)\Big(\sum_{k=-n}^{n} e^{ik(x-y)}\Big)dy \\ &= \frac{1}{2\pi}\int_{y=-\pi}^{\pi} f(y)D_n(x-y)dy = f * D_n,\end{aligned}$$

where the **Dirichlet kernel** D_n is defined as

$$D_n(x) = \sum_{k=-n}^{n} e^{ikx},$$

and the **convolution** $*$ is defined as

$$f * g(x) = \int_{z=-\pi}^{\pi} f(z)g(x-z)dz = \int_{z=-\pi}^{\pi} g(z)f(x-z)dz.$$

Let us compute the Dirichlet kernel first.

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

We want to show that

$$S_n(x) = \sum_{k=-n}^{n} c_k e^{ikx} = \frac{1}{2\pi}\int_{x=-\pi}^{\pi} f(y)\frac{\sin(n+\frac{1}{2})(x-y)}{\sin\frac{x-y}{2}}dy$$

goes to $f(x)$ as $n \to \infty$. In other words, we want

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

to converge to 0. If we know that f'' is continuous, then $(f(x)-f(y))/(\sin(x-y)/2)$ is continuously differentiable in y. Then we can do an integration by parts and conclude that S_n actually converges to f.

14 March 10, 2016

Let us continue our discussion on Fourier analysis. We are looking at a periodic function on $\mathbb{R}$ with period 2π, or alternatively a function on $[-\pi, \pi]$. We want to express f as $\sum c_n e^{inx}$ where

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

Because we are working with Riemann integrability, we assume that f is continuous up to a finite number of jumps.

Now we look at

$$s_n = \sum_{k=-n}^{n} c_k e^{ikx} = f * D_n.$$

We want to prove that $s_n \to f$ as $n \to \infty$, or alternatively $D_n \to \delta$. Here, the δ is the Dirac delta. But this does not behave nicely enough. So we instead look at the Féjer kernel defined as

$$F_n = \frac{D_1 + \cdots + D_n}{n}.$$

It turns out that $F_n * f \to f$ uniformly for f continuous. This is introduced in the homework exercise. This fact implies that the completeness of the space we are working in.

14.1 Riemann-Lebesgue lemma

Before I go in to the Lebesgue theory, I want to discuss two things. Let us go back to the Dirichlet-Dini kernel and look at the identity

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

Because $(f(x-y) - f(x))/\sin\frac{y}{2}$ does not depend on n, we can let this function be g. But then, g is kind of a continuous function, and $\sin(n + \frac{1}{2})y$ is a high-frequency sine wave. So things next to each other almost cancel out.

Lemma 14.1 (Riemann-Lebesgue)**.** *If g is integrable, then*

$$\int_{x=-\pi}^{\pi} g(x) \sin nx dx \to 0$$

as $n \to \infty$.

Proof. We use

$$\left\| f - \sum_{k=-n}^{n} c_k e^{ikx} \right\| \geq 0.$$

Expanding everything out, we see that this is in fact equivalent to

$$\|f\|^2 - 2\pi \sum_{k=-n}^{n} |c_k|^2 \geq 0.$$

Therefore we get **Bessel's inequality** that says

$$\sum_{k=-\infty}^{\infty} |c_k|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f|^2.$$

When does the equality hold? Because $\sum_{k=-n}^{n} c_k e^{ikx}$ is the projection, it better approximates f than any other sum, i.e.,

$$\left\| f - \sum_{k=-n}^{n} c_k e^{ikx} \right\| \leq \left\| f - \sum_{k=-n}^{n} \gamma_k e^{ikx} \right\|$$

for any other γ. If f can be uniformly approximated by a sum of e^{ikx}, then we see that the identity holds. □

14.2 Lebesgue measure

Lebesgue turned the table around and asked what set satisfies a given property. The problem is then to measure the size of a set E in $\mathbb{R}$. Everyone knows the size of $[a, b]$ or $(a, b]$ or $[a, b)$ or so forth. They are all $b - a$. To measure other sets, we need to approximate it by what we know.

Theorem 14.2 (Structure theorem for an open set in $\mathbb{R}$)**.** *Any open set $\mathcal{O} \subset \mathbb{R}$ is an at-most countable union of disjoint open intervals.*

Proof. For every $x \in \mathcal{O}$, let I_x be the maximum open interval $(a_x, b_x) \subset \mathcal{O}$ containing x. To attain this, we can simply set

$$a_x = \inf\{a : x \in (a, b) \subset \mathcal{O} \text{ for some } b\},$$
$$b_x = \sup\{b : x \in (a, b) \subset \mathcal{O} \text{ for some } a\}.$$

Then the set $\mathcal{O}$ is the disjoint union of these open intervals I_x. □

For any open set $\mathcal{O}$, we define its measure to be

$$m(\mathcal{O}) = \sum_{j \in J} m(I_j).$$

Generally, for any set $E \subset \mathbb{R}$, we define its **exterior measure** as

$$m_*(E) = \inf_{\mathcal{O} \supset E} m(\mathcal{O}).$$

We can only work with open sets, because we only know how to measure open sets. This is the approximation from the outside.

In order to approximate reasonably, we also need to approximate from the inside. One possibility is (in fact this is Lebesgue's original formulation) to define the interior measure of E as

$$\sup_{\text{closed } F \subset E} m(F),$$

where we define $m(F)$ by measuring a sufficiently large interval and taking out the measure of its complement. Then we can define E to be measurable if and only if the interior measure and the exterior measure agree. But nowadays, we simply define it using only the outer measure.

Definition 14.3. A set $E \subset \mathbb{R}$ is **Lebesgue measurable** if and only if given any $\epsilon > 0$ there exists an open set $\mathcal{O} \supset E$ such that

$$m_*(\mathcal{O} - E) < \epsilon.$$

Now one thing you want to know is what happens when you take the union, intersection, complete of measurable sets. Let us backstop a little bit and look at the exterior measure.

Proposition 14.4 (Countable subadditivity)**.**

$$m_*\left(\bigcup_{j=1}^{\infty} E_j\right) \le \sum_{j=1}^{\infty} m_*(E_j).$$

Note that there is an inequality because of the possible overlaps.

Proof. Given $\epsilon > 0$, there exists an open $\mathcal{O}_j \supset E_j$ such that $m_*(E_j) \ge m(\mathcal{O}_j) - \epsilon/2^j$. Then $\bigcup_j E_j \subset \bigcup_j \mathcal{O}_j$ and hence

$$\sum_j m_*(E_j) \ge \sum_j m(\mathcal{O}_j) - \epsilon \ge m(\bigcup_j \mathcal{O}_j) - \epsilon$$

and hence we get the desired result. □

Proposition 14.5 (Additivity for positively-distanced sets)**.** *If the distance between E_1 and E_2 is greater than 0, then $m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2)$.*

Proof. There exist open sets $\tilde{\mathcal{O}}_j \supset E_j$ that are disjoint. Then the result becomes trivial. □

Which sets are measurable?

- Open sets are measurable.
- Sets with zero exterior measure are measurable. (This is clear from the modern definition.)
- Countable union of measurable sets is measurable. (This you can break up ϵ into $\sum \epsilon/2^j$ and then approximate sets with error $\epsilon/2^j$.)

- Closed sets are measurable. (For any closed set F and $\epsilon > 0$ there is an open $\mathcal{O} \supset F$ such that $m(\mathcal{O}) \leq m_*(F) + \epsilon$. Then $\mathcal{O} - F$ is open, and thus is a countable union of open intervals. We can find finitely many intervals so that they approximate $\mathcal{O} \setminus F$ well.)

15 March 22, 2016

15.1 More properties of the measure

We have shown the countable subadditivity of the exterior measure.

Proposition 15.1. *For any sets E_j,*

$$m_*(\bigcup_j E_j) \leq \sum_j m_*(E_j).$$

Conversely, we have:

Proposition 15.2. *If $\operatorname{dist}(E_1, E_2) > 0$, then $m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2)$.*

Proposition 15.3. *Any countable union of measurable set is measurable.*

Proof. Let E_j be the sets and for each j, consider an open set $\mathcal{O}_j$ containing E_j such that $\mathcal{O}_j \supset E_j$ and $m_*(\mathcal{O}_j - E_j) < \epsilon 2^{-j}$. Then we see that

$$m_*\Big(\Big(\bigcup_j \mathcal{O}_j\Big) \setminus \Big(\bigcup_j E_j\Big)\Big) \leq \sum_j m_*(\mathcal{O}_j \setminus E_j) < \epsilon.$$

Thus the union is measurable. □

Proposition 15.4. *Any closed set is measurable.*

Proof. First assume that the closed set F is bounded, and let $F \subset (a, b)$. Then $(a, b) - F = \bigcup_j (a_j, b_j)$ is an open set. For each j, look at the slightly smaller closed interval $[c_j, d_j] \subset (a_j, b_j)$. Then it turns out that $(a, b) \setminus \bigcup_j [c_j, d_j]$, which is an open set containing F, is an approximation of F.

Now for an unbounded F, we observe that $F = \bigcup_N (F \cap [-N, N])$ is a countable union of measurable sets. □

Proposition 15.5. *The complement of a measurable set is measurable.*

Proof. Let E^c be a measurable set, and let $\mathcal{O}_n$ be the open set containing E^c such that $m_*(\mathcal{O}_n \setminus E^c) < 1/n$. Then we see that $\mathcal{O}_n^c \subset E$. Let $S = \bigcup_n \mathcal{O}_n^c$. Then for any n,

$$m_*(E \setminus S) \leq m_*(\mathcal{O}_n \setminus E^c) < \frac{1}{n}.$$

This implies that $E \setminus S$ has exterior measure zero, and thus is measurable. Also S is measurable, because it is the countable union of closed sets. Therefore $E = S \cup (E \setminus S)$ is also measurable. □

Now we have all the tools to construct measurable sets. Open or closed sets are measurable, and countable unions, intersections, and complements are also measurable.

Theorem 15.6. *If E_j are countably many disjoint measurable sets and $E = \bigcup_j E_j$, then*

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

Proof. We already know that $m(E) \le \sum_j m(E_j)$. Now we prove the other direction.

Assume first that each E_j is bounded. Then for each j there is a closed $F_j \subset E_j$ such that $m(E_j \setminus F_j) < \epsilon 2^{-j}$. Because F_j is compact, and disjoint, we see that the have pairwise positive distance it it follows that

$$m\Big(\bigcup_j F_j\Big) = \sum_j m(F_j).$$

Now we use the bounded case to get back. Let $T_n = \{x \in \mathbb{R} : N \le |x| < N+1\}$. Then

$$E = \bigcup_N (E \cap T_N) = \bigcup_N \Big(\bigcup_j (E_j \cap T_N)\Big) = \bigcup_j \Big(\bigcap_N (E_j \cap T_N)\Big) = \bigcup_j E_j.$$

Because each T_N is bounded, we follow this guideline and say

$$\begin{aligned} m(E) &= \sum_N m(E \cap T_N) = \sum_N \sum_j m(E_j \cap T_N) \\ &= \sum_j \sum_N m(E_j \cap T_N) = \sum_j m(E_j). \end{aligned}$$

□

We see that there is an analogue between series and a sequence of sets. In fact, we have the following.

Proposition 15.7. *Let $E_1 \subset E_2 \subset \cdots$ be a nested sequence of measurable sets, and let $E = \bigcup_j E_j$. Then $m(E) = \lim_{j\to\infty} m(E_j)$.*

Proof. You just break it put into

$$E = E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \cdots.$$

□

But more importantly, we have the limit property for the other direction.

Proposition 15.8. *Let $E_1 \supset E_2 \supset \cdots$ be a nested sequence of measurable sets, and let $E = \bigcap_i E_j$. Then $m(E) = \lim_{j\to\infty} m(E_j)$.*

Proof. In this case, E is the collapsed telescope and E_1 is the full telescope. We likewise have

$$E_1 = E \cup (E_1 \setminus E_2) \cup (E_2 \setminus E_3) \cup \cdots$$

and thus

$$m(E_1) = m(E) + \sum_j (m(E_j) - m(E_{j+1})) = m(E) + m(E_1) - \lim_{j\to\infty} m(E_j).$$

The result follows. □

Let me justify why I said this was more important. When we define the Lebesgue sum, we are going to approximate everything from the inside. That is, we will define $E_j = \{x : y_j < f(x) \le y_{j+1}\}$ and add up the "rectangles" $y_j m(E_j)$.

15.2 Egorov's theorem and the measurability of a function

Theorem 15.9 (Egorov). *Let $E \subset \mathbb{R}$ be a measurable set with $m(E) < \infty$. Let $f_j(x)$ be a function on j, with $f_j \to f$ as $j \to \infty$ at every point of E. Assume that each f_j is measurable. Then given any $\epsilon > 0$, there is a closed $F_\epsilon \subset E$ such that $m(E - F_\epsilon) < \epsilon$ and $f_j \to f$ uniformly on F_ϵ.*

What does it mean for a function to be measurable? We are trying to define the measure of the set of points under the graph of f. We want to partition the target space into $0 = y_0 < \cdots < y_m = M$ and we need all sets $E_j = \{x : y_{j-1} \le f(x) < y_j\}$ to be measurable.

Definition 15.10. f is **measurable** if $\{x : f(x) < c\}$ is measurable for all $c \in \mathbb{R}$.

Proof of Egorov's theorem. We define

$$E_{n,N} = \{x \in E : |f_j(x) - f(x)| < \frac{1}{n} \text{ for all } j > N\}.$$

This is what we need in order to get uniform convergence of f_j. Because f_j pointwise converges to f, we see that $E_{n,N} \nearrow E$ as $N \to \infty$. Then by the elescoping,

$$\lim_{N\to\infty} m(E_{n,N}) = m(E).$$

That is, there is a N_n such that $m(E \setminus E_{n,N_n}) < 2^{-n}$. Given $\epsilon > 0$, there exists an ℓ_ϵ such that $\sum_{n \ge \ell_\epsilon} \frac{1}{2^n} < \frac{\epsilon}{2}$.

Let $A_\epsilon = \bigcap_{n \ge \ell_\epsilon} E_{n,N_n}$. Then

$$m(E \setminus A_\epsilon) \le \sum_{n \ge \ell_\epsilon} m(E - E_{n,N_n}) < \frac{\epsilon}{2}.$$

Moreover, f_j uniformly converges on A_ϵ. This is because for any $\delta > 0$ we can set $n \ge \max(\ell_\epsilon, \delta^{-1})$. Then for nay $k \ge N_n$, we have $|f_k(x) - f(x)| < \delta < \frac{1}{n}$ for $x \in A_\epsilon \subset E_{n,N_n}$. □

15.3 Lebesgue integration

Let us first look at the special case when f is defined on a bounded interval $[a, b]$, is measurable, and $0 \le f < M$. One way to do it is looking at the partition $0 \le y_0 < y_1 < \cdots < y_{n-1} < y_n = M$, defining $E_j = \{y_{j-1} \le f(x) < y_j\}$, and looking at the limit of

$$\sum_{j=1}^{n} y_{j-1} m(E_j).$$

But does the limit exist?

We reformulate the problem by writing the partial sum as the integral of a "**simple function**." That is, we define

$$\chi_{E_j}(x) = \begin{cases} 1 & \text{if } x \in E_j \\ 0 & \text{if } x \notin E_j \end{cases}$$

and let

$$\varphi(x) = \sum_{j=1}^{n} y_{j-1} \chi_{E_j}.$$

Then we define the **Lebesgue integral** as

$$\int_a^b f = \sup \int \varphi,$$

where φ varies over simple functions at most f.

In fact, we can pick out a particular set of φ. For the partition, $y_j = \frac{j}{n} M$, let us denote the resulting simple function by φ_j. Then we claim that

$$\int_a^b f = \lim_{n \to \infty} \varphi_n.$$

More generally, if $\varphi_n \nearrow f$ at every point, and $\{\varphi_n\}$ is almost uniformly Cauchy, then the conclusion holds.

Applying Egorov's theorem, we see that there is a closed set $F_\epsilon \subset [a, b]$ such that $m([a, b] \setminus F_\epsilon) < \epsilon$. If we denote $\varphi_n = \sum_j c_{n,j} \chi_{E_{n,j}}$, and φ_n is uniformly Cauchy on F_ϵ, we have

$$\left| \int \varphi_n - \sum_j c_{n,j} m(E_{n,j} \cap F_\epsilon) \right| \le m([a, b] - F_\epsilon) \cdot 2M < 2\epsilon M.$$

16 March 24, 2016

We will look at three convergence theorems; the bounded convergence theorem, the dominated convergence theorem, and the monotone convergence theorem. Then there is a tool called Fatou's lemma. Next we will look at the fundamental theorem in calculus.

There are two important examples.

Example 16.1 (A non-measurable set)**.** Let us consider the interval $(0,1)$ in $\mathbb{R}$ identify it with S^1. Make an equivalence relation so that $x \sim y$ if and only if $y - x \in \mathbb{Q}$. For each equivalence class, choose a representative using the Axiom of Choice, and let it be E. Then the whole circle is the disjoint union

$$S^1 = \coprod_{q \in \mathbb{Q}} (E + q).$$

If E is measurable, then $m(E) = m(E+q)$. But if $m(E) = 0$, then $m(S^1) = 0$, and if $m(E) > 0$, then $m(S^1) = \infty$. Both leads to a contradiction, and hence E cannot be measurable.

Example 16.2 (The Cantor set)**.** Let $E_0 = [0,1]$. Then take away the middle third and let $E_1 = [0,1/3] \cup [2/3,1]$. Then for each interval, take away the middle thirds again and let $E_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$ and so forth. Then

$$C = \bigcap_i E_i$$

is a measure zero set, and is called the **Cantor set**.

16.1 Lebesgue integration

There are two tools to use: approximation by simple functions through equally spaced ever-finer partition of the target space, and almost convergence techniques.

Let f be a measurable function on $[a,b]$ such that $0 \le f < M$. Look at the partition $0 = y_0 < \cdots < y_n = M$ such that $y_j = \frac{j}{n}M$, and let

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

Then E_j is measurable since f is measurable. We define

$$\varphi_n = \sum_{j=1}^n y_{j-1}\chi_{E_j}.$$

This is a **simple function**, i.e., a finite $\mathbb{R}$-linear combination of characteristic functions of measurable sets of finite measure. Define the integral of φ_n as

$$\int \varphi_n = \sum_{j=1}^n y_{j-1} m(E_j).$$

More generally, we can consider the f to be any bounded function with finite measure support. This is because any such function can be represented as $\varphi_j \nearrow f$.

So we have $0 \leq \varphi_k \leq M$ such that they have a common fine measurable support and $\varphi_k \to f$ pointwise. By Egorov's theorem, for any $\epsilon > 0$ there exists a $A_\epsilon \subset E$ such that $m(E - A_\epsilon) < \epsilon$ and $\varphi_k \to f$ uniformly in A_ϵ. Then

$$\Big| \int \varphi_k - \int \varphi_l \Big| \leq \Big| \int_{A_\epsilon} (\varphi_k - \varphi_l) \Big| + \Big| \int_{E - A_\epsilon} (\varphi_k - \varphi_l) \Big|$$

and thus we see that the sequence $\int \varphi_k$ is a Cauchy sequence. This implies that the sequence has a limit. Also, this limit is independent of the choice of $\{\varphi_k\}$, because if we take another sequence $\{\psi_k\}$, the sequence $\varphi_1, \psi_1, \varphi_2, \psi_2, \cdots$ also has a limit. Thus we can define the integral of f as

$$\int f = \lim_{k \to \infty} \int \varphi_k.$$

Then we can extend the definition.

Definition 16.3. Let f be any measurable function on $\mathbb{R}$ such that $f \geq 0$. Then we define

$$\int f = \sup \int g$$

over all bounded measurable g with $0 \leq g \leq f$ that is supported on a set with finite measure.

The convention is that we allow $f(x) = +\infty$ for some numbers x. Also, $\int f$ is always defined, either as $+\infty$ or some number smaller than ∞.

There are some standard properties of the integral.

Proposition 16.4 (Linearity of integration)**.** *For any $a, b \geq 0$ and measurable functions $f, g \geq 0$,*

$$\int (af + bg) = a \int f + b \int g.$$

Proof. Multiplying by a scalar is not a big deal. The hard part is to get addition. All we need to do is show that for any $0 \leq h \leq f + g$, we can split h up into $0 \leq \varphi \leq f$ and $0 \leq \psi \leq g$. This is true because we can define $\varphi = \min(h, f)$ and $\psi = h - \varphi$. □

For a measurable set $E \subset \mathbb{R}$ and a measurable function $f \geq 0$, we define

$$\int_E f = \int f \chi_E.$$

Proposition 16.5 (Additivity of integration)**.** *If $E_1 \cap E_2 = \emptyset$, then*

$$\int_{E_1 \cup E_2} f = \int_{E_1} f + \int_{E_2} f.$$

Proposition 16.6 (Truncation)**.** *If $f \geq 0$ on $\mathbb{R}$ is integrable, i.e., $\int f < \infty$, then for any $\epsilon > 0$ there exists a bounded set $B_\epsilon \subset \mathbb{R}$ such that $\int_{\mathbb{R}-B_\epsilon} f < \epsilon$.*

This result follows from the monotone convergence theorem, which will be proved soon.

16.2 Fatou's lemma

Theorem 16.7 (Bounded convergence theorem)**.** *Suppose $f_n \to f$ almost everywhere, and $|f_n| \leq M$ for every n. Suppose that there is a measurable set E such that $m(E) < \infty$ and $\operatorname{supp} f_n \subset E$ for every n. Then $\int f_n \to \int f$.*

Proof. By Egorov's theorem, for any $\epsilon > 0$ we can find a subset $A_\epsilon \subset E$ with $m(E - A_\epsilon) < \epsilon$ such that $f_n \to f$ uniformly on A_ϵ. For any $\delta > 0$ there exists a N_δ such that $|f_k(x) - f(x)| \leq \delta$ for any $k > N_\delta$ and $x \in A_\epsilon$. Then

$$\left| \int_E f - \int_E f_k \right| \leq \left| \int_{A_\epsilon} f - f_k \right| + \left| \int_{E-A_\epsilon} f - f_k \right| \leq \delta m(E) + 2M\epsilon.$$

it follows that $\int f_n \to \int f$. □

Lemma 16.8 (Fatou's lemma)**.** *Let $f_n \geq 0$ be measurable on $\mathbb{R}$. Then*

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n.$$

Proof. We first reduce to $g_n \nearrow f$ by using the definition of lim inf. Let

$$g_n = \inf_{k\geq n} f_k, \quad f = \liminf_{k\to\infty} f_k.$$

Then we have $g_k \nearrow f$. Then it suffices to prove

$$\int f \leq \lim_{n\to\infty} \int g_n.$$

Let us dominate $\int f$ as $\int f = \sup \int \varphi$ where $0 \leq \varphi \leq f$ and φ is bounded and finite-measure supported. If we let $\varphi_n = \min(\varphi, g_n$, then $\varphi_n \nearrow \varphi$. Then by the bounded convergence theorem, we see that

$$\lim_{n\to\infty} \int g_n \geq \lim_{n\to\infty} \varphi_n = \int \varphi.$$

Then we can take the supremum over all φ and get

$$\int f = \sup \int \varphi \leq \lim_{n\to\infty} \int g_n.$$ □

Theorem 16.9. *Let $f_n \geq 0$ be measurable on $\mathbb{R}$ and suppose that $|f_n| \leq g$ almost everywhere and $f_n \to f$ almost everywhere. Then $\lim_{n\to\infty} \int f_n = \int f$.*

Proof. Fatou's lemma says that $\int f \leq \liminf_{n\to\infty} \int f_n$. Also, $\int f_n \leq \int f$ and thus $\limsup \int f_n \leq \int f$. It follows that $\lim_{n\to\infty} \int f_n = \int f$. □

An immediate consequence is:

Theorem 16.10 (Nonnegative monotone convergence theorem)**.** *Suppose that $f_n \geq 0$ on $\mathbb{R}$ is measurable and $f_n \leq f_{n+1}$. Then*

$$\int \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int f_n.$$

17 March 29, 2016

We have Egorov's theorem, Fatou's lemma, and the convergence theorems. These are pretty much all the tools we currently have.

We recall the definition of the Lebesgue integral.

Definition 17.1. If $f \geq 0$ is a bounded, finite measure supported measurable function, we define

$$\int f = \lim_{n\to\infty} \int \varphi_n$$

where $\varphi_n \to f$ almost everywhere.

If $f \geq 0$ is any measurable function, we define

$$\int f = \sup \int \varphi$$

where φ ranges over measurable functions $0 \leq \varphi \leq f$ that is finite measure supported and is bounded.

If f is any measurable, function, we define

$$\int f = \int f_+ - \in f_-$$

where $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$.

Theorem 17.2 (Dominated convergence theorem)**.** *Let $f_n \to f$ almost everywhere, and assume that there exists an integrable g such that $|f_n| \leq g$ on $\mathbb{R}$. Then*

$$\lim_{n\to\infty} \int |f - f_n| = 0.$$

Proof. We can truncate with respect to the domain, and we can do it with respect to the target. First we have that for any $\epsilon > 0$ there exist a B_ϵ with finite measure such that $\int_{\mathbb{R}-B_\epsilon} |f| < \epsilon$. This follows from the monotone convergence theorem. So we can only look at things in a finite measure domain.

Now let $E_n = \{x \in \mathbb{R} : |f(x)| \leq n \text{ and } |x| \leq n\}$. Then for a sufficiently large N, we have $\int \chi_{E_N} f_n \to \int \chi_{E_N} f$. So we use the two kinds of truncation. The details are posted in the lecture notes. □

Theorem 17.3 (Absolute continuity)**.** *Let f be a integrable function defined on $\mathbb{R}$. Then given ay $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that $\int_E |f| < \epsilon$ for any $m(E) < \delta_\epsilon$.*

Proof. We we let $E_n = \{x : |f(x)| \leq N\}$ then $\lim_{\mathbb{R}} |f\chi_{E_N} - f| < \epsilon$. □

We note that this theory of Lebesgue integration can be regarded as a two dimensional measure theory. We can look at the graph and then define it as the measure of the set under the graph.

17.1 Existence of derivatives

The main tool in proving the fundamental theorem will be Vitali's covering technique(which was in the problem set). I have also included in the problem set the following statement.

Proposition 17.4. *If $E_k \nearrow E$, where the sets need not be measurable, then $m_*(E_k) \nearrow m_*(E)$.*

Theorem 17.5 (Vitali's covering)**.** *Let $E \subset \mathbb{R}$ with $m_*(E) < \infty$. Assume that for each $x \in E$ is associated a nonempty subset $A_x \subset \mathbb{R}_+$. Then given any ϵ there exist $x_1, \dots, x_k \in E$ and $r_j \in A_{x_j}$ such that $(x_j, x_j + r_j)$ are disjoint and*

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

Proof. We may assume that E is bounded because $m_*(E \cap [-n, n]) \nearrow m_*(E)$ (we are using the proposition above). We also may assume that there exists an $N > 0$ such that $\sup A_x > 1/N$ for all $x \in E$. Then pick a sufficiently small δ, and let $x_1 \leq \inf E + \delta$. Then pick an $r_1 \in A_{x_1}$ such that $r_1 \geq 1/N$, and then replace E by $E - (-\infty, x_1 + r_1)$. This ends after finitely many steps, and because we can set δ arbitrarily small, we are done. I guess everyone got this problem. □

The motivation for defining the differentiation is using this r instead of taking the difference quotient with respect to any number. Let us refine by taking an open $\mathcal{O} \supset E$. Assume that $\{r \in A_x : (x, x + r) \subset \mathcal{O}\}$ is nonempty for every x. Then given a collection of sets $\mathcal{A} = \{A_x\}$, we can find

$$\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}, \mathcal{O}) = \coprod_{j=1}^{k} (x_j, x_j + r_j) \subset \mathcal{O}$$

such that $m(E') \geq m(E) - \epsilon$. We can likewise define

$$\mathcal{V}_{\text{left}}(\epsilon, \mathcal{A}, \mathcal{O}) = \coprod_{j=1}^{k} (x_j - r_j, x_j).$$

Now we want to prove the fundamental theorem of calculus, which will take the form of

$$\frac{d}{dx} \int_a^x f = f \quad \text{(almost everywhere)}.$$

Assuming that f is integrable on $\mathbb{R}$, the integral is well-define. But is the derivative well-defined? At the least, we know from absolute continuity that the function $x \mapsto \int_a^x f$ is continuous. The idea is to split f into $f = f_+ - f_-$ so that $x \mapsto \int_a^x f$ is nondecreasing. Now we claim that if F is nondecreasing, then F' exists everywhere.

Assume that f is nondecreasing. We can define four different different quotients:

$$D^+f = \limsup_{r\to 0+} \frac{f(x+r)-f(x)}{r}, \quad D_+f = \liminf_{r\to 0+} \frac{f(x+r)-f(x)}{r}$$

and likewise D^-f and D_-f for the left derivative. We want to prove that all four D^+, D_+, D^-, D_- are equal. Automatically, we have $D_+f \le D^+f$ and $D_-f \le D^-f$. So it suffices to show $D_+ \ge D^-f$ and $D_-f \ge D^+f$.

For $\alpha < \beta \in$ ıı, let

$$E_{\alpha,\beta} = \{x : D_+f(x) < \alpha < \beta < D^+f(x)\}.$$

It suffices to show that $m(E_{\alpha,\beta}) = 0$, because then $\bigcup_{\alpha,\beta} E_{\alpha,\beta}$ will also have measure zero. Assume that $m_*(E_{\alpha,\beta}) > 0$. Then for any ϵ there exists an open $\mathcal{O}$ such that $m(\mathcal{O}) \le m_*(E_{\alpha,\beta}) + \epsilon$. If we let

$$A_+(x) = \Big\{ r \in \mathbb{R}_+ : \frac{f(x+r)-f(x)}{r} < \alpha, \quad (x, x+r) \subset \mathcal{O} \Big\}$$

then we can find an almost covering

$$\mathcal{O}' = \mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}_+, \mathcal{O}) = \bigcup_{j=1}^{k} (x_j, x_j + r) \subset \mathcal{O}$$

such that if we let $E' = E \cap \mathcal{O}'$ then $m(E') \ge m(E) - \epsilon$. Likewise we can define

$$A^+(x) = \Big\{ r \in \mathbb{R}_+ : \frac{f(x+r)-f(x)}{r} < \beta, \quad (x, x+r) \subset \mathcal{O}' \Big\}$$

and get a covering

$$\mathcal{O}'' = \mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}^+, \mathcal{O}').$$

Then $E'' = E' \cap \mathcal{O}''$ has positive measure. Now when we apply f, we have

$$f(\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}^+, \mathcal{O}')) \subset f(\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}_+, \mathcal{O}))$$

and it follows that

$$\beta m(\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}^+, \mathcal{O}')) \le m(f(\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}^+, \mathcal{O}'))) \le m(f(\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}_+, \mathcal{O}))) \le \alpha m(\mathcal{V}_{\text{right}}(\epsilon, \mathcal{A}_+, \mathcal{O})).$$

Letting ϵ sufficiently small, we get a contradiction.

So we get the following theorem.

Theorem 17.6. *A nondecreasing function is almost everywhere differentiable.*

18 March 31, 2016

Last time, we showed that the differentiation of an nondecreasing function is well-defined almost everywhere by using Vitali's covering technique.

18.1 Fundamental theorem of calculus

Theorem 18.1. *Assume that f is integrable on $\mathbb{R}$, and let $a \in \mathbb{R}$ be a real number. Then*

$$\frac{d}{dx}\int_a^x f = f$$

almost everywhere.

Because f is integrable, we may split f into $f = f_+ - f_-$, where both f_+ and f_- are nonnegative. So it suffices to prove for nonnegative f.

Let $F(x) = \int_a^x f$. Then clearly F is nondecreasing and thus F' is well-defined. The tool is to show that $\int_{x_1}^{x_2}(F' - f) = 0$ for any x_1, x_2. Then it follows that $\int_{\mathcal{O}}(F' - f) = 0$ for any open set $\mathcal{O}$ that is the disjoint union of finite intervals. We can then use the dominated convergence theorem to show that $\int_{\mathcal{O}}(F' - f) = 0$ for any open set $\mathcal{O}$. If $F' - f \neq 0$ on a set of positive measure, then it we can approximate some set from the inside and get a contradiction.

In our case, $F' \geq 0$. We will show that F' is integrable. Because F' is defined almost everywhere, we have

$$\frac{F(x + h_n) - F(x)}{h_n} \to F'$$

almost everywhere for a sequence $h_n \searrow 0$. Using Fatou's lemma, we see that

$$\begin{aligned}\int_a^b F' &\leq \liminf_{n\to\infty}\int_a^b \frac{F(x+h_n) - F(x)}{h_n} \\ &= \liminf_{n\to\infty}\frac{1}{h_n}\Big(\int_a^b F(x+h_n) - \int_a^b F(x)\Big) \\ &= \liminf_{n\to\infty}\frac{1}{h_n}\Big(\int_b^{b+h_n} F - \int_a^{a+h_n} F\Big) \\ &\leq \liminf_{n\to\infty}(F(b+h_n) - F(a)) = F(b) - F(a).\end{aligned}$$

So F is integrable and moreover we have $\int_{x_1}^{x_2} F' \leq F(x_2) - F(x_1)$. We further have to show that they are equal. We use truncation to achieve this. We let $f_n = \max(f, n)$. Then by applying the bounded convergence theorem instead of Fatou's lemma, actaully $\int_{x_1}^{x_2} F_n' \leq F_n(x_2) - F_n(x_1)$. Then by the monotone convergence theorem, we can look at the limit $n \to \infty$ and get $\int_{x_1}^{x_2} F_n' \nearrow \int_{x_1}^{x_2} F$.

Thus theorem 18.1 is proved.

Theorem 18.2. *Let F be defined on $[a, b]$ where $a < b$, and assume that F is absolutely continuous. Then*

$$\int_{[a,b]} F' = F(b) - F(a).$$

Proof. We start out with F being absolutely continuous. We need to somehow show that F' has bounded variation and is integrable. Let $F' = f$. Then by the first part, if we let $G = \int_a^x f$ then $F' = G'$. It follows that $(F - G)' = 0$ almost everywhere and because $F - G$ is absolutely continuous, it is zero. (This we will prove in the next subsection.) □

18.2 More on absolute continuity

We have defined absolute continuity for an indefinite integral. We now extend it to any function.

Definition 18.3. A function F is **absolutely continuous** if given $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that for any finite disjoint intervals (a_j, b_j) with $m(\bigcup_j (a_j, b_j)) < \delta_\epsilon$,

$$\sum_j |F(b_j) - F(a_j)| < \epsilon.$$

Absolute continuity clearly implies bounded variation.

Proposition 18.4. *If g is absolutely continuous on $[a, b]$ and $g' = 0$ almost everywhere, then g is constant.*

Proof. Let $E = \{g' \neq 0\}$. Then $m(E) = 0$. That means that for any $\epsilon > 0$ there is an open set $\mathcal{O}$ containing E such that $m(\mathcal{O}) < \epsilon$. Then you can kind of cover the interval.

To be rigorous, we use Vitali's covering argument. Fix $\eta > 0$ and $x \in [a, b] - E$ so that $g'(x) = 0$. Let

$$A_{\eta,x} = \{r > 0 : |g(x + r) - g(x)| < \eta r\}$$

and then for any $\delta_\epsilon > 0$ you can cover the domain up by disjoint intervals $(a_j - b_j)$ whose measure is at least $(b - a) - \delta_\epsilon$. Then the complement of intervals have measure at most δ and hence the sum of the differences is at most ϵ. The intervals have some of the difference at most $\eta(b - a)$. Therefore $|g(b) - g(a)| \leq \eta(b - a) + \epsilon$. This shows that $g(b) = g(a)$. □

There is a counterexample to the second part of the fundamental theorem of calculus if F is not absolutely continuous. This shows that the condition cannot be replaced by bounded variation.

Example 18.5. We recall the Cantor function. This function $f : [0, 1] \to [0, 1]$ is given by

$$0.a_1 a_2 a_3 \ldots_{(3)} \mapsto 0.b_1 b_2 b_3 \cdots b_{m\,(2)}$$

where $a_m = 1$ and $a_1, \ldots, a_{m-1} \in \{0, 2\}$ and $b_j = \lceil a_j/2 \rceil$. This function clearly has bounded variation and is locally constant almost everywhere, but is not constant.

18.3 Hardy-Littlewood maximal function

When we proved the first part of the fundamental theorem of calculus, we tried to approximate $\int F'$ by $F(b) - F(a)$ by using truncation and the bounded convergence theorem. But can we use dominated convergence theorem to do this in one step? For instance, if

$$\sup_{(x_1,x_2)\ni x} \left| \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} f \right|$$

is dominated by an integrable function, then

$$\int F' = \int \lim_{x_1,x_2\to x} \frac{F(x_2) - F(x_1)}{x_2 - x_1} = \lim_{x_1,x_2\to x} \int \frac{F(x_2) - F(x_1)}{x_2 - x_1} = F(b) - F(a).$$

Definition 18.6. Let f be an integrable function. We define the **Hardy-Littlewood maximal function** as

$$f^*(x) = \sup_{x\in I} \frac{1}{m(I)} \int_I |f|$$

where I ranges over open intervals containing x.

If f^* were integrable, it would have been nice because we would have been able to apply dominated convergence theorem directly. But this is, in general, false. However, we can modify the proof and still get an alternative proof.

We know that $\int_{\mathbb{R}} |f| < \infty$. Then given any $\epsilon > 0$, we can approximate f by a compactly supported continuous function g such that

$$\int_{\mathbb{R}} |f - g| < \epsilon.$$

We have

$$\left| \frac{1}{m(I)} \int_I f - f(x) \right| \le \frac{1}{m(I)} \int_I |f - g| + \left| \frac{1}{m(I)} \int_I g - g(x) \right| + |g(x) - f(x)|.$$

Taking the limit superemum as I shrinks around x, we get

$$\limsup_{I\to x} \left| \frac{1}{m(I)} \int_I f - f(x) \right| \le (f - g)^*(x) + 0 + |g(x) - f(x)|$$

because g is contiuous. By definition, $|f(x) - g(x)|$ is integral. We want to show that as g approaches f, the right hand side approaches 0 almost everywhere. We make use of the following modified Tchebychev inequality.

Theorem 18.7 (Modified Tchebychev's inequality)**.** *Let f be an integrable function. The*

$$m\{f^* > c\} \leq \frac{3}{c} \int |f|.$$

Proof. We need the following variant of the covering lemma.

Lemma 18.8. *If $I_1, \dots, I_\ell$ are bounded open intervals in $\mathbb{R}$ then there exists a disjoint sub-collection $I_{i_1}, \dots, I_{i_k}$ such that*

$$m\Big(\bigcup_{\nu=1}^{\ell} I_\nu\Big) \leq 3m\Big(\bigcup_{\mu=1}^{k} I_{i_\nu}\Big).$$

Proof. Just pick the biggest interval and delete all the intervals intersecting this one. Do this and you get the disjoint intervals. □

We will contiue next time. □

19 April 5, 2016

Let us continue with our discussion on Lebesgue theory. When Foruier first developed his theory, we handwaved everything. Later on, Lebesgue came along and rigourized some of the theory. There was an open problem: If $|f|^p$ is integrable for some $p > 1$, then does the Fourier series necessarily converge almost everywhere? This was solved in 1966 by Carleson for $p = 2$ and Hunt proved the full problem. Their proofs uses the maximal function, and this is why I am doing the Hardy-Littlewood maximal function although we already have a proof of the fundamental theorem of calculus.

19.1 Hardy-Littlewood maximal function

Recall that the Hardy-Littlewood maximal function is defined as

$$f^*(x) = \sup_{I \ni x} \frac{1}{m(I)} \int_I |f|$$

for an integral function f.

Theorem 19.1 (Modified Tchebychev inequality)**.**

$$m\{f^* \geq c\} \leq \frac{3}{c} \in |f|.$$

Proof. It directly follows from the Vitali covering technique and the following lemma. ☐

Lemma 19.2. *Given a finite collection of open intervals $I_1, \ldots, I_\ell$, there exists a subcollection $I_{j_1}, \ldots, I_{j_k}$ that is disjoint and*

$$m\Big(\bigcup_{\nu=1}^{\ell} I_\nu\Big) \leq 3m\Big(\bigcup_{\mu=1}^{k} I_{j_\mu}\Big).$$

Proof. We use the algorithm of picking the largest interval and removing all the intervals intersecting it. By maximality, all the intervals intersecing it would be covered by the 3 times larger interval. Then we see that all the intervals will contained inside at least one of the 3 times larger intervals. ☐

We want to prove the following theorem:

Theorem 19.3. *Let f be an integrable funciton. For almost every x,*

$$\left|\frac{1}{m(I)} \int_I f - f(x)\right| \to 0$$

as $m(I) \to 0$, where I is an open interval.

The main technique we are going to use is to approximate f by a sequence of functions that behave nicely, and then approximate both terms using the sequence.

Proof. Let g_n be continuous with compact support so that $g_n \to f$ almost everywhere and

$$\int |f - g_n| \to 0$$

as $n \to \infty$. Then we have

$$\left| \frac{1}{m(I)} \int_I f - f(x) \right| \le \left| \frac{1}{m(I)} \int_I f - \frac{1}{m(I)} \int_I g_n \right| + \left| \frac{1}{m(I)} \int_I g_n - g_n(x) \right| + |g_n(x) - f(x)|.$$

Taking the limit supremeum as $I \to x$, we get

$$\limsup_{I \searrow x} \left| \frac{1}{m(I)} \int_I f - f(x) \right| \le (f - g_n)^*(x) + |g_n(x) - f(x)|,$$

because g_n is continuous. Next taking the limit infimum as $n \to \infty$, we get

$$\limsup_{I \searrow x} \left| \frac{1}{m(I)} \in_I f - f(x) \right| \le \liminf_{n \to \infty} (f - g_n)^*(x)$$

almost everywhere. We now need to show that the right hand side is zero almost everywhere.

For any $\epsilon > 0$, consider the set

$$E_\epsilon = \big\{ x : \liminf_{n \to \infty} (f - g_n)^*(x) > \epsilon \big\}.$$

It suffices to show that $m(E_\epsilon) = 0$, becasue we can take the union where ϵ ranges over positive rationals. Let

$$\tilde{E}_{\epsilon,n} = \big\{ x : (f - g_n)^*(x) > \epsilon \big\}$$

and then

$$m(E_\epsilon) = m\Big(\limsup_{n \to \infty} \tilde{E}_{\epsilon,n} \Big) = \limsup_{n \to \infty} m(\tilde{E}_{\epsilon,n}) \le \frac{3}{\epsilon} \limsup_{n \to \infty} \int |f - g_n| = 0$$

by the Tchebychev inequality. □

19.2 The Lebesgue set

Recall that we have proved the convergence of the Fourier series by showing that the convolution of f with the Dirichlet-Dini kernel converges to f.

Definition 19.4. A point x is a **Lebesgue point** of a function f if and only if

$$\lim_{\substack{m(I) \searrow 0 \\ x \in I}} \frac{1}{m(I)} \int_I |f - f(x)| = 0.$$

This is slightly stronger than what we have shown above to be almost everywhere. But still the following is true.

Theorem 19.5. *The complement of the Lebesgue set (the set of all Lebesuge points) is measurable and of measure zero.*

Proof. We apply the above thereom to the function $f(x) - r$. For each $f - r$, denote the set on which the first fundamental theorem of calculus fails by E_r. Then $m(E_r) = 0$.

We claim that the complement of the Lebesgue set is in $\bigcup_{r\in\mathbb{Q}} E_r$. This is because if $x \notin \bigcup_r E_r$ then

$$\frac{1}{m(I)} \int_I |f - f(x)| \le \frac{1}{m(I)} \int_I |f - r| - |f(x) - r|.$$

Then we can make the both terms on the right hand side small. □

19.3 Hilbert spaces

The reason people look at this is because they want to solve differential equations.

Suppose we have a function f and a Fourier series of f. We have to convolute f with a function K_ϵ that approaches the Dirac delta. People were not happy with this becasue this is not defined as function. So came the notion of a weak solution. This is something that can evaluate a so-called test functions. For instance, x is a weak solution to $Ax = b$ if $(A^*\varphi, x) = (\varphi, Ax) = (\varphi, b)$ for every test function. Then the question becomes finding an x such that maps $A^*\varphi \to (\varphi, b)$ for every φ.

Definition 19.6. An $\mathbb{R}$-vector space X with an inner product is a **Hilbert space** if satisfies the following properties:

(1) (completeness) Every Cauchy sequence in X converges in X with respect to the metric $d(x, y) = \|x - y\|_X$.

(2) (separability) There exists a countable dense set in X.

Example 19.7. Consider the set $L^2(\mathbb{R})$ of square integrable functions on $\mathbb{R}$; $f \in L^2(\mathbb{R})$ if and only if f is measurable and $|f|^2 = f^2$ is integrable. The inner product is given by

$$(f, g)_{L^2(\mathbb{R})} = \int_{\mathbb{R}} fg.$$

More generally, consider $L^2(E)$, where E is a measruable set in $\mathbb{R}$.

Proposition 19.8. *The space $L^2(\mathbb{R})$ is a Hilbert space.*

Proof. We first prove completeness. Supose that we have a sequence of functions such that $\int_{\mathbb{R}} |f_n - f_m|^2 \to 0$ as $n, m \to \infty$. We need to show that $f_n \to f$ for some $f \in L^2(\mathbb{R})$.

Pick an increasing sequence k_ν such that $\|f_n - f_m\| < 2^{-\nu}$ for every $m, n \geq k_\nu$. Then by the triangle inequality,

$$\sum_{j=1}^{\infty} \|f_{k_j} - f_{k_{j+1}}\| < 1.$$

Clearly

$$f_{k_m} = f_{k_1} + \sum_{j=1}^{m-1} (f_{k_{j+1}} - f_{k_j}).$$

We want to make it sure this converges. Let

$$g_m = |f_{k_1}| + \sum_{j=1}^{m-1} |f_{k_{j+1}} - f_{k_j}|, \quad g = \lim_{m \to \infty} g_m.$$

Then $|f_{k_m}| \leq g_m \leq g$. By the triangle inequality again,

$$\|g_m\| \leq \|f_{k_1}\| + \sum_{j=1}^{m-1} \|f_{k_{j+1}} - f_{k_j}\| \leq \|f_{k_1}\| + 1.$$

But because $g_m^2 \nearrow g^2$ and all $\int g_m^2$ are bounded, we see that g is well-defined as an integrable function by the monotone convergence theroem. Then we have $f_{k_m} \to f$ almost everywehre. Using the dominated convergence theorem, it further follows that $\int |f_{k_n} - f|^2 \to 0$ as $n \to \infty$. This proves the completeness. □

Consider a closed subset $Y \subset X$. Given a $x \in X$, we can find an onrthogonal projection u of x onto Y. This will satisfiy $\|x - u\|_X \leq \|x - y\|_X$ for every $y \in Y$.

Theorem 19.9 (Riesz representation theroem)**.** *Let X be a Hilbert space an $f : X \to \mathbb{R}$ be a bounded linear functional, i.e., there exists a C_f such that $|f(x)| \leq C_f \|x\|_X$ for every $x \in X$. Then there unqiuely exists a $v_f \in X$ such that $f(x) = (x, v_f)_X$ for every $x \in X$.*

20 April 7, 2016

There will be a test next Tuesday.

We gave an example

$$L^2(\mathbb{R}^2) = \{f \text{ mesurable on } \mathbb{R} \text{ s.t. } \int_{\mathbb{R}} |f|^2 < \infty\}.$$

By definition, the space comes with an inner product. There are two important things: completeness and separability. The completeness is to make the series converge, and separability is to make sure we can choose a basis.

Let us consider the linear equation $Ax = b$ where there is the compatibility condition $Sb = 0$. Then there is an exact sequence

$$V_1 \xrightarrow{A} V_2 \xrightarrow{S} V_3$$

and the solution is given by

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

The case of differential equaitons is the same. In this case, all vector spaces are replaced by Hilbert spaces.

$$H_1 \xrightarrow{A} H_2 \xrightarrow{S} H_3$$

20.1 The projection map

Consider a Hilbert space X and consider a closed subspace Y. Then there is a projection map $\pi_Y : X \to Y$ such that $\pi_Y(x)$ is the element in Y closed to X.

Of course, we need to prove that this actually exists and is unique. Given $x \in X$, let $\mu = \inf_{y \in Y} \|x - y\|$. We want to show that μ is realized. There exists a sequence y_n such that

$$\mu = \inf_n \|x - y_n\|.$$

Becasue

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

and $\|y_n + y_m - 2x\|^2 \geq 4\mu^2$ by definition of μ, it follows that $\|y_n - y_m\|$ is small. Hence $\{y_n\}$ is Cauchy and thus they converge to some y^*. Since $\|-\|$ is contiuous, $\|y^* - x\| = \mu$.

The unqiueness comes from the same parallelogram law. So we can finally define:

Definition 20.1. We define the **projection map** $\pi_Y : X \to Y$ as

$$\pi_Y(x) = y^*.$$

where y^* is the unique y^* that minimzes $\|x - y^*\|$ as

If you have a finite dimenisonal subspace Y, then there exists an orthonormal set $e_1, \dots, e_k$ such that

$$Y = \bigoplus_{j=1}^{k} \mathbb{R} e_j.$$

It automatically follows that Y is closed, and moreover

$$\pi_Y x = \sum_{j=1}^{k} (x, e_j) e_j.$$

For instance, the Fourier series is the special case.

20.2 Riesz representation theroem

Theorem 20.2. *Let X be a Hilbert space over $\mathbb{R}$ and let $f : X \to \mathbb{R}$ be a bounded $\mathbb{R}$-linear functional.*[3] *Then there exists a unique $v_f \in X$ such that $f(x) = (x, v_f)_X$.*

Proof. Assume that f is not identically zero, because then it is trivial. We look at the kernel $Y = \ker f$ and any element $u \notin Y$. Without loss of generality, assume that $u \perp Y$, because we can just look at $u - \pi_Y(u)$. We can further assume that $\|u\| = 1$ by scaling.

Let $v_f = f(u)u$. We claim that $f(x) \equiv (v_f, x)$. This is becuase we can decompose $x = (x, u)u + \pi_Y(x)$ and then

$$f(x) = (x, u)f(u) + 0 = (x, f(u)u) = (x, v_f).$$

Uniqueness follows from $(v, v - w) \neq (w, v - w)$ for any $v \neq w$. □

Now let us go back to our diagram.

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

In the finite-dimensional case, we have $x = T^*(TT^* + S^*S)^{-1}b$. We always know that $B = (TT^* + S^*S)$ is positive definite in the finite dimensional case. But in the infinite dimensional case, things are more difficult, and thus we give an ***a priori* estimate**

$$\|T^*y\|^2 + \|Sy\|^2 \geq C\|y\|^2$$

for any $y \in H_2$.

Definition 20.3. Let $T : X \to Y$ be an $\mathbb{R}$-linear map between Hilbert spaces, and assume that T is continuous. Then we define

$$\|T\| = \sup_{|x| \leq 1} \|Tx\|.$$

[3]Such a function is bounded, i.e., $|f(x)| \leq C\|x\|$ for some absolute constant C, if and only if it is continuous. That is because X is a vector space and then we can bring everything to the origin and rescale things.

Proposition 20.4. *Let $T : X \to Y$ be a bounded operator. Then there exists a unique bounded operator $T^* : Y \to X$ such that*

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

for every $x \in X$ and $y \in Y$.

We call this T^* as the **adjoint** of T.

Proof. Fix $y \in Y$. The map $x \mapsto (y, Tx)$ is bounded because

$$|f(x)| = |(Tx, y)| \le \|Tx\|\|y\| \le \|T\|\|x\|\|y\|.$$

Then by Riesz representation theorem, there is a unique u such that $(x, u) = (Tx, y)$. We define $u = T^* y$.

This Y^* is easily checked to be linear. To prove boundedness, we use the inequality we obtained above. In the Riesz representation theroem, we always have $\|v_f\| \le \|f\|$. Then $\|T^* y\| \le \|T\|\|y\|$ for all y and hence $\|T^*\| \le \|T\|$. □

Moreover, by uniqueness, we have $(T^*)^* = T$. Therefore $\|T^*\| = \|T\|$.

Theorem 20.5. *Let*

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

be bounded operators, and assume the a priori *estimate. Then given $y \in H_2$ such that $Sy = 0$, there exists a unique $x \in H_1$ such that $Tx = y$ and $x \perp \ker T$.*

Proof. Solving $Tx = y$ is equivalent to solving $(Tx, z) = (y, z)$ for every $z \in H_2$. Then we are solving $(x, T^* z) = (y, z)$.

Consider the linear functional $f : H_1 \to \mathbb{R}$ given by

$$T^* z \mapsto (z, y).$$

There are two problems: there might be elements in H_1 that is not expressed as $T^* z$, even they can be, it might not be unique. We will continue next time. □

21 April 14, 2016

Let I be an open interval of length A in $\mathbb{R}$. We look at the domain $\Omega = I \times \mathbb{R}^n \subset \mathbb{R} \times \mathbb{R}^n$ and consider the operator

$$L = B\frac{\partial^m}{\partial x^m} + \sum_{v=0}^{m-1} L_v \frac{\partial^v}{\partial x^v}$$

where $B \neq 0$ is constant and

$$L_v = \sum_{\lambda_1+\cdots+\lambda_n \le p_v} a_{\lambda_1\cdots\lambda_n} \frac{\partial^{\lambda_1+\cdots+\lambda_n}}{\partial y_1^{\lambda_1}\cdots\partial y_n^{\lambda_n}}$$

where again each $a_{\lambda_1\cdots\lambda_n}$ is a constant. We want to find a solution to $Lu = f$ on Ω, for a given f on Ω. If $f \in L^2(\Omega)$ then we will try to find a solution in the *weak* sense so that $(Lu, g) = (u, L^*g)$ where $g \in C_0^\infty(\Omega)$. Because C_0^∞ is dense in L^2, we will get a full solution.

21.1 The toy model

Consider Hilbert spaces and operators

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

Then we can calculuate the minimal solution

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

Because we don't really know the invertibility of $AA^* + S^*S$ is invertible, we assume the *a priori* estimate. That is, assume that there exists a $c > 0$ such that

$$((TT^* + S^*S)g, g) \geq c\|g\|^2.$$

It automatically follows that $\|T^*g\|^2 + \|Sg\|^2 \geq c\|g\|^2$.

Proposition 21.1. *Suppose*

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

are bounded linear maps and also suppose that $ST = 0$. Asuume that there is an a priori *estimate. Then given an $f \in H_2$ with $Sf = 0$, there exists a $u \in H_1$ such that $Tu = f$.*

Before proving the proposition, let me recall the finite dimensional toy model. We have $SA = 0$ and $Sb = 0$, and we want to find a solution to $Au = b$. Let $v = (AA^* + S^*S)^{-1}b$. Then $(AA^* + S^*S)v = b$. Because we want to set $u = A^*v$, we want to show that $S^*Sv = 0$. This follows from $(S^*Sv, S^*Sv) = (SS^*Sv, Sv) = (0, Sv) = 0$.

Motivated by this baby toy model, we now move on to the teenage toy model. We shall prove the proposition by using the Riesz representation theroem.

Proof. By the Riesz representation theroem, we need only find a u such that $(T^*g, u) = (g, f)$ for all $g \in H_2$. That is, we need to show that there is a functional that maps $T^* \mapsto (g, f)$. The first question is, is the map well-defined? Also, we need to check that the functional is bounded. We shall do that in one step.

Now let us analyze further. We see that

$$|(g, f)| \leq \|g\|\|f\| \leq \frac{1}{\sqrt{c}}(\|T^*g\|^2 + \|Sg\|^2)^{1/2}\|f\|.$$

If we prove that $\|Sg\| = 0$, then we are done.

Recalling what we did in the finite dimensional case, we decompose g into $g = g_1 + g_2$, where $g_1 \in \ker S$ and $g_2 \in (\ker S)^\perp$. The key is to look at the estimate separately. From the compatibility condition, we have $f \in \ker S$ and hence $(g_2, f) = 0$. Also, from $(\ker S)^\perp \subset (\operatorname{im} T)^\perp \subset \ker T^*$ we get $T^*g_2 = 0$. We then have

$$\begin{aligned} |(g, f)| \leq |(g_1, f)| &\leq \frac{1}{\sqrt{c}}(\|T^*g_1\|^2 + \|Sg_1\|^2)^{1/2}\|f\| \\ &= \frac{1}{\sqrt{c}}\|T^*g\|\|f\|. \end{aligned}$$

□

After some more work, we get the adult toy model.

Proposition 21.2. *Suppose*

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

are bounded linear maps and also suppose that $ST = 0$. Asuume that there is an a priori *estimate. Then given an $f \in H_2$ with $Sf = 0$, there exists a $u \in H_1$ such that $Tu = f$ in the weak sense and $\|u\| \leq \|f\|/\sqrt{c}$.*

21.2 Fourier transforms

Given a differential operator T, let's say we want to solve $Tu = f$. We can take the Fourier transform to make it into $\hat{T}\hat{u} = \hat{f}$. Then $\hat{T}$ is a polynomial and hence we get $\hat{u} = \hat{f}/\hat{T}$. But if we do it like this, $\hat{T}$ might have a zero, and it is hard to make sense out of this. We are going to try and resolve this.

Consider the interval $[0, 2\pi]$ and a function $f \in L^2([-\pi, \pi])$. We define

$$\hat{f}(n) = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}dx.$$

Then we kind of want the identity

$$f(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{inx}$$

in some sense, probably in $L^2([-\pi,\pi])$.

The first step is rescaling. Because we always have $1/2\pi$ hanging around, we get rid of it by using $e^{2\pi inx}$ istead of e^{inx}. Then f will be in $L^2([0,1])$ and we will have

$$\hat{f}(n) = \int_0^1 f(x)e^{-2\pi inx}dx, \qquad f(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{2\pi inx}.$$

If we rescale it further, and send L to infty, the we will get

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

Now the whole theory of Fourier transform is to invert the process.

The first question we need to address is whether $\hat{f}$ is defined. It is actually defined for only specific f. Because we want to use it as test functions, things like $\int_{-\infty}^{\infty}|f| < \infty$ won't do. (When we differentiate it, it becomes something that is not a test function.) So we need a better space.

Schwartz was the first person to come up with the following space.

Definition 21.3. The **Schwartz space** $\mathcal{S}(\mathbb{R})$ is the set of all $C^\infty(\mathbb{R})$ functions f such that

$$\sup_{x\in\mathbb{R}}|f^{(k)}(x)(1+x^{2l})| < \infty$$

for all $k, l \geq 0$.

Then the fundemantal theorem of Fourier series can be formulated as the following.

Theorem 21.4. *Let us define the **Fourier transform** as*

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

Then for any $f \in \mathcal{S}(\mathbb{R})$, we have

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

Example 21.5. Let us consider the Gauss distribution $f(x) = e^{-\pi x^2}$. Then we have $\int_{-\infty}^{\infty} e^{-\pi x^2}1$. We denote

$$K(x) = e^{-\pi x^2}, \qquad K_\delta(x) = \frac{1}{\sqrt{\delta}}e^{-\pi(\frac{x}{\sqrt{\delta}})^2} = \frac{1}{\sqrt{\delta}}e^{-\frac{\pi x^2}{\delta}}.$$

Then we easily see that $\int K_\delta(x) = 1$. This will play the role of the approximate identity. The key is that K is its Fourier transform, i.e., $\hat{K} = K$. To see this, we compute

$$\begin{aligned}\hat{K}(\xi) &= \int e^{-\pi x^2 - 2\pi i\xi x}dx = \int e^{-\pi(x+i\xi)^2}e^{-\pi\xi^2}dx \\ &= e^{-\pi\xi^2}\end{aligned}$$

after looking at the contour integral over the rectangle.

22 April 19, 2016

The basic theorem in partial differential equations is the following.

Theorem 22.1. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and let $L = \sum_{|\alpha|\le m} a_\alpha (\partial/\partial x)^\alpha$ be a differential operator, where $|(\alpha_1, \dots, \alpha_n)| = \alpha_1 + \cdots + \alpha_n$. Then for every $L^2(\Omega)$, there exists an $u \in L^2(\Omega)$ such that $Lu = f$ in the weak sense and $\|u\| \le c\|f\|$ for some $c > 0$.*

By "weak sense", we mean that for every $\psi \in C_0^\infty(\Omega)$, we have the equality

$$(u, L^*\psi)_{L^2(\mathbb{R}^n)} = (f, \psi).$$

Here, L^* is just the formal adjoint, defined by

$$L^* = \sum_{|\alpha|\le m} (-1)^{|\alpha|} \overline{a_\alpha} \Big(\frac{\partial}{\partial x}\Big)^\alpha.$$

If u is indeed in C_0^∞, then it follows that u is a solution even in the classical sense. To make things easier, we restrict our attention to the 1-dimensional case.

22.1 The Gauss distribution, and operations with Fourier transforms

Let $\mathcal{S}(\mathbb{R})$ be the Schwartz space. It can be checked that the Fourier transform sends $\mathcal{S}(\mathbb{R})$ to itself. Recall that for a $f \in \mathcal{S}(\mathbb{R})$, the Fourier transform is defined as

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

The key is that the inverse process is given quite nicely as

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

There are many ways to prove this, and I have outline two of the ways in the homework assignment. But the slick way people use nowadays is to use the Gauss distribution.

There are nice properties we can use. First of all, the Fourier transform of $e^{-\pi x^2}$ is itself. The second is that its rescaling is an approximate identity. The third is that

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

There are two ways in showing that the Fourier transform fixes the function $K(x) = e^{\pi x^2}$. The first way is to use complex analysis and say

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

because we can use Cauchy's integral formula on the rectangle $[-R, R] \times [0, i\xi]$.

Before doing the real analysis proof, let me digress a bit and look at the effect of three things on the Fourier transform.

(1) Translation
Let $(T_h f)(x) = f(x+h)$. Then we see that
$$\widehat{(T_h f)}(\xi) = \int f(x+h)e^{-2\pi i\xi x}dx = e^{2\pi i h\xi}\hat{f}(\xi).$$

(2) Differentiation
We have
$$\widehat{f'}(\xi) = \int f'(x)e^{-2\pi i x\xi}dx = (2\pi i\xi)\hat{f}(\xi).$$

(3) Rescaling
Let us define $f_\delta(x) = f(\delta x)$. Then
$$\widehat{f_\delta}(\xi) = \frac{1}{\delta}\hat{f}\Big(\frac{\xi}{\delta}\Big).$$

Going back to the Fourier transform of the Gaussian distribution, we have
$$\begin{aligned}(\hat{K})'(\xi) &= \int_{x\in\mathbb{R}} (-2\pi i x)e^{-\pi x^2}e^{-2\pi i x\xi}dx \\ &= i\int K'(x)e^{-2\pi i x\xi}dx \\ &= i(i2\pi\xi)\int K(x)e^{-2\pi i x\xi}dx = -2\pi\xi\hat{K}(\xi).\end{aligned}$$

Hence solving the differential equation, we get what we want.

We can easily prove that the Gauss distribution approximates the identity, because for $|x| \geq \eta$, we have
$$K_\delta(x) \leq \frac{e^{-\pi(\eta/\sqrt{\delta})^2}}{\sqrt{\delta}}$$
and hence goes to zero as $\delta \to 0$. Then $f * K_\delta$ approximates f as $\delta \to 0$.

Again, let me digress and talk about the Fourier transform of the convolution. We have
$$(f * g)(x) = \int_{y\in\mathbb{R}} f(y)g(x-y)dy$$
by definiion, and it follows that
$$\widehat{(f * g)}(\xi) = \int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} f(y)g(x-y)e^{-2\pi i x\xi}dydx$$
and hence
$$\widehat{(f * g)} = \hat{f}\cdot\hat{g}.$$

Now let use the fact that K_δ approximates the identity. We have

$$f(0) = \lim_{\delta\to 0}\int_{\mathbb{R}} f(x)K_\delta(x)dx = \lim_{\delta\to 0}\int \hat{f}(\xi)\hat{K}_\delta(\xi)d\xi = \int_{\mathbb{R}} \hat{f}(\xi)$$

by the monotone convergence theorem at the end, since $\hat{K}_\delta(\xi) = e^{-\pi\delta\xi^2} \nearrow 1$. Now, applying this identity to the translate $T_x f$, we get

$$f(x) = T_x f(0) = \int \widehat{(T_x f)}(x)dx.$$

So this is the slick proof of the inverse Fourier transform.

22.2 Plancherel's theorem

The main difficulty in solving differential equations is that if you take the Fourier transform of $Lu = f$ and make it into $Q\hat{u} = \hat{f}$, then you cannot divide by Q because Q might have zeros. So we first use the toy model to change it into a Riesz representation theorem form, and then use the Fourier transform. It is a roundabout way, but this is why people were not able to solve it for a hundred years.

Now when we take the Fourier transform, we need the following theorem.

Theorem 22.2. *For any $f \in \mathcal{S}(\mathbb{R})$, we have*

$$\int |f|^2 = \int |\hat{f}|^2.$$

Proof. The idea is to use a special case of Fourier inversion formula. We find a function g such that $\hat{g}(\xi) = |\hat{f}(\xi)|^2$ and hope that $g(0)$ is the same as both terms. Because $|\hat{f}(\xi)|^2 = \hat{f}(\xi)\overline{\hat{f}(\xi)}$, we define $h(x) = \overline{f(-x)}$ and $g = f * h$. Then we see that

$$g(0) = \int f(y)h(0-y)dy = \int f(y)\overline{f(y)}dy = \int |f(y)|^2 dy.$$

On the other hand, $g(0) = \int |\hat{g}|^2 = \int |\hat{f}|^2$. □

22.3 Solving constant coefficient differential equations

Now we are in business. We first modify the baby toy model argument with the Riesz representation theorem. We let

$$L = \sum_{|\alpha|\le m} a_\alpha \Big(\frac{\partial}{\partial x}\Big)^\alpha, \qquad L^* = \sum_{|\alpha|\le m} (-1)^{|\alpha|}\overline{a_\alpha}\Big(\frac{\partial}{\partial x}\Big)^\alpha.$$

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain, and assume that for some $c > 0$, we have

$$\|L^*\psi\| \ge c\|\psi\|$$

for every $\psi \in \mathcal{S}(\mathbb{R}^n)$. We want to show that $Lu = f$ on Ω can be solved in the weak sense.

We first remark that L^* is not defined on $L^2(\mathbb{R}^n)$. So we need to produce a alternative "Hilbert space" in order to do what we can do. We give up the completeness property and just look at the **pre-Hilbert space** $\mathcal{H}_0 = C_0^\infty(\Omega) \subset L^2(\Omega)$. We define the inner product as

$$(\phi, \psi)_{\mathcal{H}_0} = (L^*\phi, L^*\psi)_{L^2(\Omega)}.$$

This is so far a pre-Hilbert space, and we now take its completion to produce the Hilbert space $\mathcal{H}$ we are going to work in. (That is, an element of $\mathcal{H}$ is is a Cauchy sequence of $\mathcal{H}_0$.

Now we define a linear functional $\ell_0 : \mathcal{H}_0 \to \mathbb{C}$ given by

$$\ell_0(\psi) = (\psi, f)_{L^2(\Omega)}.$$

Then we see that

$$\begin{aligned}|\ell_0(\psi)| &= |(\psi, f)_{L^2(\Omega)}| \le \|\psi\|_{L^2(\Omega)}\|f\|_{L^2(\Omega)} \\ &\le \frac{1}{c}\|L^*\psi\|_{L^2(\Omega)}\|f\|_{L^2(\Omega)} \\ &= \frac{1}{c}\|\psi\|_{\mathcal{H}_0}\|f\|_{L^2(\Omega)}.\end{aligned}$$

Hence ℓ_0 is bounded, and therefore we can extend the functional $\ell_0 : \mathcal{H}_0 \to \mathbb{C}$ to $\ell : \mathcal{H} \to \mathbb{C}$ by using Cauchy sequences. This new linear functional should also satisfy

$$\|\ell\|_{\mathcal{H}} \le \frac{1}{c}\|f\|_{L^2(\Omega)}.$$

We can then use the Riesz representation theorem and find a $\tilde{u}$ such that

$$(\psi, f)_{L^2(\Omega)} = (L^*\psi, L^*\tilde{u})_{L^2(\Omega)}.$$

Going back to the differential equation, it suffices to show that

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

for every $\psi \in C_0^\infty(\Omega)$. Using the Plancherel's formula, we can change it to

$$\|Q\hat{\psi}\|_{L^2(\mathbb{R})} \ge c\|\hat{\psi}\|_{L^2(\mathbb{R})}.$$

We are going to use the mean value property with polynomial weight, which you have proved in one of the assignment.

Proposition 22.3. *Let $Q(z)$ be a monic polynomial with complex coefficients, let F be a holomorphic function defined on the closed disc $\{|z| \le 1\}$. Then*

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

23 April 21, 2016

Let $A > 0$ and I be an open interval of length A in $\mathbb{R}$. The domain Ω is an open subset of $I \times \mathbb{R}^n \subset \mathbb{R}$, and we have a differential operator

$$L = B\frac{\partial^m}{\partial x^m} - \sum_{\nu=0}^{m-1} L_\nu \frac{\partial}{\partial x^\nu}, \quad B \neq 0, \qquad L_\nu = \sum_{\lambda_1+\cdots+\lambda_n \leq p_\nu} a_{\lambda_1,\dots,\lambda_n} \frac{\partial^{\lambda_1+\cdots+\lambda_n}}{\partial y_1^{\lambda_1}\cdots \partial y_n^{\lambda_n}}.$$

Let $c = e^{\pi A}/|B|(2\pi)^n$ be the constant. Our conclusion is that for every $f \in L^2(\Omega)$, there exists a weak solution u of $Lu = f$ on Ω with $\|u\|_{L^2(\Omega)} \leq c\|f\|_{L^2(\Omega)}$. In our case, we are only going to consider $n = 0$.

23.1 The last estimate

To do this, we have to use the mean value property with the monic polynomial coefficient.

Proposition 23.1. *Let $P(z) = z^m + \sum_{k=0}^{m-1} b_k z^k$ be a monic polynomial and let $F(z)$ be a holomorphic function on $\{|z| \leq 1\}$. Then*

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

Proof. If $P(z) = \prod_{|\alpha_j|\geq 1}(z-\alpha_j)\prod_{|\beta_j|<1}(z-\beta_j)$, we define

$$\tilde{P}(z) = \prod_{|\alpha_j|\geq 1}(z-\alpha_j)\prod_{|\beta_j|<1}(1-\bar{\beta}_j z)$$

and apply the mean value property to $\tilde{P}F$ instead of PF. ☐

Last time I have showed that if we have the estimate

$$\|T^* g\| \geq c\|g\|,$$

for every $g \in \operatorname{Dom} T^*$, then we can use the Riesz representation theorem on the Hilbert space constructed as the completion of the pre-Hilbert space to get our claim. So the problem is reduced to showing

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

for every $\psi \in C_0^\infty(\Omega)$.

We will use Plancherel's identity. Let $\widehat{L^*\psi} = Q\hat{\psi}$. Then Q can be assumed to be a monic polynomial. Here we use the polynomial mean value property and translation. We have

$$\begin{aligned}\int_{\xi\in\mathbb{R}} |Q(\xi+i\eta)\hat{\psi}(\xi+i\eta)|^2 d\xi &= \int_{\xi\in\mathbb{R}} |\widehat{L^*\psi}(\xi+i\eta)|^2 d\xi \\ &= \int_{x\in(-M,M)} |(L^*\psi)(x)e^{i\pi\eta x}|^2 dx \leq e^{4\pi|\eta|M}\int_{x\in\mathbb{R}} |(L^*\psi)(x)|^2 dx.\end{aligned}$$

(For simplicity we let $A = (-M, M)$.) Then we get

$$\int_{\xi\in\mathbb{R}} |Q(\xi + \cos\theta + i\sin\theta)\hat{\psi}(\xi + \cos\theta + i\sin\theta)|^2 d\xi \le e^{4\pi M} \int_{x\in\mathbb{R}} |(L^*\psi)(x)|^2 dx.$$

If we integrate this over θ, we get

$$\begin{aligned}\int_{\xi\in\mathbb{R}} |\psi(\xi)|^2 d\xi &= \int_{\xi\in\mathbb{R}} |\hat{\psi}(\xi)|^2 d\xi \\ &\le \frac{1}{\pi}\int_{\theta=0}^{2\pi}\int_{\xi\in\mathbb{R}} |Q(\xi\cos\theta + i\sin\theta)\hat{\psi}(\xi + \cos\theta + i\sin\theta)|^2 d\xi\, d\theta \\ &\le \frac{1}{2\pi}\int_{\theta=0}^{2\pi} e^{4\pi M}\int_{x\in\mathbb{R}} |(L^*\psi)(x)|^2 dx \le c\|L^*\psi\|^2.\end{aligned}$$

This is the end of proof.

23.2 Argument principle

There is still time left, so I want to do a few things I wanted to do. Here the argument means the angle. Suppose $f(z)$ is holomorphic on the closure of a domain Ω with a piecewise smooth boundary. If we assume f is nowhere zero on the boundary $\partial\Omega$ then the number of zeros of f inside Ω is equal to $1/2\pi$ times the increment of the argument of f around $\partial\Omega$. The argument is given as

$$f(z) = |f(z)|e^{i\arg(f(z))}$$

and is defined up to an element of $2\pi\mathbb{Z}$. But still the increment around a closed loop is well-defined, and is denoted by $\Delta_{\partial\Omega}\arg f$.

Proof. We apply Stokes' theorem to $d\log f(z) = f'(z)/f(z)dz$. The behavior of f near a zero can be described easily. We can consider the power series expansion

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n$$

on $|z-a| < r$, and if you plug this in, you get that the contour integral of $d\log f(z)$ around a root is $2\pi i$ times the multiplicity. Then we get the result. □

Example 23.2. We can try to compute the number of zeros of $z^4 + z^3 + 4z^2 + 2z + 3 = 0$ in each quadrant of the complex plane. After computation, you see that the 1st quadrant has 0 roots.

23.3 Singularities of holomorphic functions

We say that a is an isolated singularity if f is holomorphic in a deleted neighborhood of a point a. Let z be a point in the ring $r < |z-a| < R$. Then we can use the Cauchy integral formula to get

$$f(z) = \frac{1}{2\pi i}\oint_{|\xi-a|=R}\frac{f(\xi)}{\xi - z}d\xi - \frac{1}{2\pi i}\oint_{|\xi=a|=r}\frac{f(\xi)}{\xi - z}d\xi.$$

Then we get a **Laurent series expansion** by expanding the integral in terms of the Taylor series. It will be of the form

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z-a)^n,$$

where $c_n = 0$ for sufficiently small n.

24 April 26, 2016

So we have been looking at the fundamental solution of differential equations, differentiation and the fundamental theorem of calculus. For the higher dimensional case, there was the Stokes theorem which came from the boundary of the boundary being empty. In the more technical side, there was the problem of which functions are differentiable and integrable. This came with the hierarchy of functions: integrable functions (or measurable functions), functions with boundary variation (those that can be differentiated) and absolutely continuous functions (those on which the fundamental theorem of calculus hold).

After all this, you want to know how to actually solve differentiation. This involves the method of Fourier and eigenfunctions. The way we use Hilbert spaces is to make sure we can use the Riesz representation theorem to get the solution.

Going back to singularities of holomorphic functions, let $f(z)$ be holomorphic on $\{R_1 < |z-a| < R_2\}$. Then

$$f(z) = \frac{1}{2\pi i}\int_{|\zeta|=R_1} \frac{f(\zeta)d\zeta}{\zeta - z} - \frac{1}{2\pi i}\frac{f(\zeta)d\zeta}{\zeta - z}$$

by the Cauchy integral formula. Using the Taylor expansions

$$\frac{1}{\zeta - z} = \frac{1}{(\zeta-a)-(z-a)} = \frac{1}{\zeta - a}\frac{1}{1-\frac{z-a}{\zeta-a}} = \sum_{n=0}^{\infty}\frac{(z-a)^n}{(\zeta-a)^{n+1}},$$

$$\frac{1}{z-\zeta} = \frac{1}{(\zeta-a)-(z-a)} = \frac{-1}{z-a}\frac{1}{1-\frac{\zeta-a}{z-a}} = -\sum_{n=0}^{\infty}\frac{(\zeta-a)^n}{(z-a)^{n+1}},$$

we get

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z-a)^n.$$

Here, the coefficient will be given by

$$c_n = \frac{1}{2\pi i}\int_{|\zeta-a|=r}\frac{f(\zeta)}{(\zeta-a)^{n+1}}d\zeta.$$

This is call the **Laurent series** of f.

Now let us look at the case of an isolated singularity, that is, the case in which $f(z)$ is holomorphic in $\{0 < |z-a| < R\}$. When we look at the Laurent series expansion about a, we will get one of the following.

(1) (Essential singularity) There are infinitely many nonzero c_n with $n < 0$.

(2) (Pole) There are finitely many nonzero c_n with $n < 0$.

(3) (Removable singularity) There are no nonzero c_n with $n < 0$.

Let us first look at the case of the pole. If the Laurent expansion looks like

$$f(z) = \frac{c_{-k}}{(z-a)^k} + \cdots + \frac{c_{-1}}{z-a} + \sum_{n=0}^{\infty} c_n (z-a)^n$$

with $c_{-k} \neq 0$. This k is called the order of the pole at a. In this case, the image of $f(z)$ on $\{0 < |z-a| < \epsilon\}$ is contained in a neighborhood $\{|z| > A_\epsilon\}$ of ∞ in S^2, and $A_\epsilon \to \infty$ as $\epsilon \to 0$.

In the case of a essential singularity, the image of $f(z)$ on $\{0 < |z-a| < \epsilon$ is dense in $\mathbb{C}$ for every $\epsilon > 0$. To prove it, it suffices to show that the intersection with any nonempty disk in $\mathbb{C}$ is nonempty. Suppose that the intersection of f on some punctured neighborhood and the disk $|f(z) - b| < \beta$ is empty. Then there exists a function g that is holomorphic at a and

$$f(z) = b + \frac{1}{g(z)}.$$

This contradicts the fact that f has a essential singularity.

Also, there is the residue around a singularity defined as

$$\mathrm{Res}_a\, f = \frac{1}{2\pi i} \int_{|z-a|=r} \frac{f(z)}{z-a} dz = c_{-1}.$$

24.1 Baire category, Uniform boundedness, Open mapping theorems

We introduced Hilbert spaces to use the Riesz representation theorem to solve partial differential equations with constant coefficients. I want to first talk about estimates in the weak norm, and eigenfunctions of compact self-adjoint operators in a Hilbert space. The reason we need stuff in more generality is that the Fourier's e^{inx} functions work only in the case of differential equations with *constant* coefficients.

In the theory of Hilbert spaces, things work because of this Baire category theorem. The Baire category theorem works also works generally on **Banach spaces**, i.e., normed vector spaces that are complete, and even on any complete metric space.

Theorem 24.1 (Baire category theorem)**.** *Given a complete metric space* (X, dist_X)*, the intersection of a countable number of open dense sets is again dense.*

Proof. The proof actually quite simple. Let $\mathcal{O}_n$ be the open dense subsets. To show the intersection is dense, we need to show that

$$U \cap \bigcap_{n \geq 0} \mathcal{O}_n \neq \emptyset.$$

First pick a $x_1 \in U \cap \mathcal{O}_1$, which we can do because $\mathcal{O}_1$ is dense. Then there exists a $r_1 > 0$ such that $\overline{B_{r_1}(x)} \subset U \cap \mathcal{O}_1$. Likewise, we can inductively define x_n and $r_n > 0$ such that

$$\overline{B_{r_n}(x_n)} \subset B_{r_{n-1}}(x_{n-1}) \cap \mathcal{O}_n.$$

Then because $\{x_n\}$ is Cauchy, it converges to some point, and our balls are nested, we are done. $\square$

Using this, we have the uniform boundedness theorem.

Theorem 24.2 (Uniform boundedness theorem)**.** *Let X be a Banach space and Y be any (not even necessarily complete) normed vector space. Let J be a family continuous (i.e., bounded) linear operators $T : X \to Y$. If J is pointwise bounded, i.e.,*

$$\sup_{T\in J} \|T(x)\|_Y < \infty \text{ for every } x \in X,$$

then J is uniformly bounded, i.e.,

$$\sup_{\|x\|_X \leq 1, T\in J} \|T(x)\|_Y < \infty.$$

Proof. Let **nowhere dense** mean to be the complement of a open dense set. Then Baire's category theorem says that $\bigcup F_n$ must also be nowhere dense and hence cannot be X.

Now let

$$F_n = \{x \in X : \sup_{T\in J} \|T(x)\|_Y \leq n\}.$$

Then F_n is closed and $\bigcup F_n = X$, and hence one of it must contain an open set. Then we get uniform boundedness. $\square$

Theorem 24.3 (Open mapping theorem)**.** *Let $T : X \to Y$ be a surjective and continuous linear map between two Banach spaces. Then T is open, i.e., T(neighborhood of $0 \in X$) contains some neighborhood of $0 \in Y$.*

Proof. Let B_X be a open unit ball of X. By surjectivity, we have

$$\bigcup_{n\geq 0} T(nB_X) = Y.$$

Then some closure $\overline{T(n_0 B_X)}$ must contain an open ball in Y. If we translate and rescale, we get $\overline{T(n_1 B_X)} \supset B_Y$. Fix an $0 < \epsilon < 1$. Then for any $y \in Y$ there exists an $x_1 \in X$ such that $\|x_1\|_X \leq n_1$ and $\|y - Tx_1\|_Y < \epsilon$. Then you can do a successive approximation to get the result. $\square$

24.2 The spectral theorem

Definition 24.4. Let X be a Hilbert space and $T : X \to X$ be a linear map. We say that T is **compact** if for any given bounded sequence x_n, there exists a subsequence x_{n_k} such that Tx_{n_k} converges in X.

Theorem 24.5 (The spectral theorem)**.** *Let T be a self-adjoint operator, i.e., $(Tx, y) = (x, Ty)$. Then the following are true.*

(1) The eigenvalues of T are real.

(2) *If x and y are eigenvectors with eigenvalues λ, μ, and $\lambda \neq \mu$, then $x \perp y$.*

(3) *The dimension of the eigenspace Λ_λ for eigenvalue λ has dimension $\dim \Lambda_\lambda < \infty$ if $\lambda \neq 0$.*

(4) *The closure of the linear span of all eigenvectors is X.*

How is this related to Fourier transform? Let T be the differential operator d/dx on the space C^∞ on $\mathbb{R}$ with period 2π. The problem is that T is not bounded, so if you apply it to $(TT^* + 1)^{-1}$ then you get the basis.

Proof. (1) and (2) are trivial. We see that

$$\|T\| = \sup_{\|x\| \leq 1} |Tx| = \sup_{|x|, \|y\| \leq 1} |(Tx, y)| = \sup_{\|x\| \leq 1} |(Tx, x)|.$$

The last equality comes from

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

So we need to produce eigenvectors. By the equality we have, there is a sequence $x_\nu \in X$ such that $\|x_\nu\| = 1$ and $|(Tx_\nu, x_\nu)| \to \|T\|$. Without loss of generality, we can suppose that $(Tx_\nu, x_\nu) \to \|T\|$. The goal is to show that either $\|T\|$ or $-\|T\|$ is an eigenvalue with eigenvector $\lim_{\nu \to \infty} Tx_{n_\nu}$. (Here compactness is used.) Let $\lambda = \|T\|$. We see that

$$\|Tx_n - \lambda x_n\|^2 = \|Tx_n\|^2 - 2\lambda \Re(Tx_n, x_n) + \lambda^2 \|x_n\|^2 \leq 2\lambda^2 - 2\lambda \Re(Tx_n, x_n) \to 0.$$

Because we may assume that $\lambda \neq 0$ (if it is zero, then there is nothing to prove), we see that

$$x_n = \frac{1}{\lambda}(Tx_n - (Tx_n - \lambda x_n)) \to \frac{1}{\lambda} y,$$

where $Tx_n \to y$. Then $Ty = \lambda y$.

So we have one eigenvector. For any $c > 0$, we claim that

$$\dim_{\mathbb{C}}(\text{linear span of all } \Lambda_\lambda \text{ with } |\lambda| \geq c) < \infty.$$

It it is infinite dimensional, then there would be x_n and $|\lambda_n| \geq c$ with $Tx_n = \lambda_n x_n$, also with $\{x_n\}$ orthonormal. Then

$$\|Tx_n - Tx_m\|^2 = \lambda_n^2 + \lambda_m^2 \geq 2c^2$$

But this cannot be possible because T is compact. ∎

Index