Math 222A Lecture Notes
Partial Differential Equations

Professor: Daniel Tataru
Scribe: Daniel Raban

Contents

1 Introduction to Partial Differential Equations

We will not be strictly following a textbook, but the main outline will about the same as *Introduction to Partial Differential Equations* by Evans. The main difference will be when we cover the theory of **distributions**. For this, you could use the book by Strichartz.[1] A more in-depth reference would be the book by Friedlauder. The hardcore option is volume 1 of Hormander's books.

1.1 Basic notation for partial differential equations

Partial differential equations is the next level up from ordinary differential equations. For ODEs, we have functions $u : \mathbb{R} \to \mathbb{R}$ (or sometimes $\mathbb{C} \to \mathbb{C}$).

Definition 1.1. An **ordinary differential equation** is an equation $F(u, u', u'', \dots, u^{(n)}) = 0$, where F is a function.

Example 1.1. A **linear** ordinary differential equation is of the form $a_0 u + a_1 u' + \cdots + a_n u^{(n)} = f$, where the a_i are the **coefficients**.

Linear ODEs are generally covered in calculus, and nonlinear ODEs are generally covered in a class on ODEs. Further in this direction is the study of dynamical systems.

The above type of ODE is called a **scalar equation**.

Definition 1.2. A **system** of ordinary differential equations concerns $u : \mathbb{R} \to \mathbb{R}^n$ with

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}.$$

A partial differential equation concerns $u : \mathbb{R}^n \to \mathbb{R}$ (or $\mathbb{C}$) with $x = (x_1, \dots, x_n)$ and $\partial = (\partial_{x_1}, \dots, \partial_{x_n}) = (\partial_1, \dots, \partial_n)$. Higher order partial derivatives look like: $\partial_1^2, \partial_1\partial_2, \partial_2^2$. The solution for this notation is *multiindices*, where we denote $\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_n^{\alpha_n} =: \partial^\alpha$, where $\alpha = (\alpha_1, \dots, \alpha_n)$ is a **multiindex**. In this case, the **order** is $|\alpha| := \alpha_1 + \cdots + \alpha_n$.

Definition 1.3. A **partial differential equation** looks like $F(u, \partial u, \dots, \partial^{(m)} u) = 0$, which we may also write as $F(u^{(\le m)}) = 0$.

Example 1.2. Linear PDEs look like

$$\sum_{|\alpha| \le m} c_\alpha(x) \partial^\alpha u = f.$$

If $f = 0$, we call this a **homogeneous equation**, and if $f \neq 0$, we call this an **inhomogeneous equation**. where f is the **source term**.

[1] Strichartz is alive at 102 years old, so this book is quite old!

The course will be about 80% linear PDEs and about 20% nonlinear PDEs. Nonlinear PDEs are generally harder and require knowedledge of corresponding linear PDEs to understand. The exception, which we will study in this class, is **first order scalar equations**, $F(x, u, \partial u) = 0$. This will require ODEs.

1.2 Scalar equations vs systems

Every scalar equation is already a system, but we can do better: If we have a scalar equation $F(u^{(\leq m)}) = 0$, we can convert it to a first order system. We do this by basically changing notation, writing the derivatives of u as separate functions: $u_\alpha := \partial^\alpha u$, where $|\alpha| \leq m$. This changes the equation to $F(u_{\leq m}) = 0$, where $u_{\leq m}$ refers to the collection of α-partial derivatives of u with $|\alpha| \leq m$. The functions are related by $\partial_j u_\alpha = u_{\alpha+e_j}$, where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$, with a 1 in the j-th coordinate.

This conversion is not a correspondence. The class of first-order systems is much more complex than the class of scalar equations.

1.3 Important examples of linear PDEs

Example 1.3. The **transport equation** is the following linear PDE:

$$\sum_{j=1}^{n} a_j \partial_j u + bu = f.$$

Example 1.4. The **Laplace operator** Is $\Delta = \partial_1^2 + \cdots + \partial_n^2$. The **Laplace equation** is

$$-\Delta u = f.$$

Example 1.5. The **heat equation** is

$$\partial_t u - \Delta u = f.$$

Here, we think of $u(x, t)$ as a function of x and t. You may think this is just a notational convention, but in practice, it is very useful to think of one of the variables as representing time.

The heat equation describes a system evolving in time, so we call it an **evolution equation**. By contrast, the Laplace equation is a **static equation**. Notice that the Laplace equation and the heat equation are the same in the case that u is constant with respect to time. For this reason, we can interpret it as looking for the **steady states** of the evolution equation.

Example 1.6. The **wave equation** is another evolution equation, given by

$$(\partial_t^2 - \Delta)u = f.$$

The wave equation describes sound waves, electromagnetic waves, gravitational waves, elastic waves, and more.

The wave equation is an evolution equation, so it needs *initial data*. Write it as $\partial_t^2 u = \Delta u + f$. Compare this to Newton's law: $\partial_t^2 u$ corresponds to the acceleration, and $\Delta u + f$ corresponds to the force. We write our initial position as $u(t = 0, x) = u_0(x)$ and our initial velocity as $\partial_t u(t = 0, x) = u_1(x)$.

Example 1.7. The **Schrödinger equation** is a complex evolution equation of the form

$$(i\partial_t + \Delta)u = 0.$$

Here, u is a complex function. This is the fundamental equation of quantum mechanics.

1.4 What do we want to study?

In calculus, you learn methods to calculate solutions to differential equations. This is not the purpose of this course. We want to understand:

1. Existence: Does the equation have solutions?
2. Uniqueness: Is the solution the only one; i.e. does it definitively describe the behavior of the system we are studying, or can we not tell?
3. Continuous dependence on data: If we change our initial data, how does the solution change?

These three questions constitute **well-posedness theory**. We can also study:

4. Properties of solutions.

2 Function Spaces and Ordinary Differential Equations

We are interested in studying first order nonlinear scalar PDEs, equations of the form $F(x, u, \partial u) = 0$. Here is the battle plan: First, we will need to discuss function spaces and an provide an introduction to ODEs. Then we will be able to study nonlinear, scalar PDEs. We will study linear PDEs, then semilinear PDEs, and then work our way up to studying nonlinear PDEs.

2.1 Function spaces

What functions could be solutions to a PDE? How do we verify that a function is a solution? We need functions that are differetiable, but this is far from the only thing we will consider.

Suppose we have a function $u : \mathbb{R}^n \to \mathbb{R}$.

Definition 2.1. The set of **(bounded) continuous functions** are denoted $C(\mathbb{R}^n)$. It has the norm $\|u\|_{C(\mathbb{R}^n)} = \sup_{x\in\mathbb{R}^n} |u(x)|$, $C(\mathbb{R}^n)$.

For now, we will assume these functions are bounded, but we may not do so later. If $\Omega \subseteq \mathbb{R}^n$, we can similarly define $C(\Omega)$.

Definition 2.2. A **normed space** is a vector space equipped with a **norm** $u \mapsto \|u\| \geq 0$, which satisfies

1. $\|u + v\| \leq \|u\| + \|v\|$
2. $\|\lambda u\| = |\lambda|\|u\|$ for $\lambda \in \mathbb{R}$.
3. $\|u\| = 0 \implies u = 0$.

A **Banach space** is a normed space is a normed space which is **complete**, i.e. any Cauchy sequence is convergent.

That is, if $u_n \in X$ and $\lim_{n,m\to\infty} \|u_n - u_m\| = 0$, the sequence u_n must have a limit.

Example 2.1. $\mathbb{R}$ and $\mathbb{C}$ are complete.

Example 2.2. Equipped with the norm $\|u\|_{C(\mathbb{R}^n)} = \sup_{x\in\mathbb{R}^n} |u(x)|$, $C(\mathbb{R}^n)$ is a Banach space.

Example 2.3. $C^1 = \{u \in C : u \text{ differentiable everywhere}, \partial u \in C\}$ is the space of **continuously differentiable functions**. This space has the norm $\|u\|_{C^1} = \|u\|_C + \|\partial u\|_C$.

More generally, we may consider $C^m(\mathbb{R}^n)$. The set $\bigcap_{m=1}^{\infty} C^m(\mathbb{R}^n) =: C^\infty(\mathbb{R}^n)$ is the set of **smooth functions**. In general, the smooth functions is too small a class of functions to be the only focus of study in PDEs.

Here are examples of functions.

Example 2.4. Observe that

$$u(x) = \frac{1}{1+x^2} \in C(\mathbb{R}),$$

while

$$v(x) = x^2 \notin C(\mathbb{R})$$

because it is not bounded.

Definition 2.3. $C_{\text{loc}}(\mathbb{R})$ is the space of **continuous but not necessarily bounded functions**.

Example 2.5. If $I_N = [-N, N]$, we can try to use $\|u\|_{C(I_N)} = \sup_{x \in I_N} |u(x)|$. We would be able to get countably many of these to measure convergence of functions. But this is not a norm on all of $\mathbb{R}$, since it assigns 0 to nonzero functions. This is a *seminorm*.

Definition 2.4. A **seminorm** is a norm without the property that $\|u\| = 0 \implies u = 0$.

What does convergence look like with respect to seminorms? What happens is that $u_n \to u$ in C_{loc} if $\|u_n - u\|_{C(I_N)} \to 0$ for each N. So we extend the concept of a normed space to a **locally convex space**, where instead of a norm, we may have infinitely many seminorms.

Why is this called locally convex? In $\mathbb{R}^n$, we can specify convergence by a fundamental system of neighborhoods, balls around each point. Another property of balls is that they are convex. If we want to talk about convergence in a locally convex space, we can also do it using by specifying convex balls. We could have many different types of balls around any point defined by different seminorms.

From this point on, we will use C to refer to C_{loc}. So our functions may be unbounded.

Example 2.6. The seminorms for $C^m(\mathbb{R}^n)$ look like

$$p_{K,N}(u) = \sup_{x \in K} \sup_{|\alpha| \leq N} |\partial^\alpha u(x)|,$$

where $K \subseteq \mathbb{R}^n$ is compact.

Later, we will study more function spaces, such as Sobolev spaces.

2.2 Ordinary differential equations and Lipschitz functions

A (nonlinear) ODE regards a function $u : \mathbb{R} \to \mathbb{R}$ which solves an equation of the form

$$\begin{cases} u' = F(x, u(x)) \\ u(0) = u_0. \end{cases}$$

If we let the codomain be $\mathbb{R}^n$, we get a system of equations.

If this equation solvable? We are asking about existence of solutions, uniqueness of solutions, dependence of solutions on initial data, and local vs global solutions. At the minimum, we require that F is continuous and look for a C^1 local solution.

Theorem 2.1 (Peano)**.** *If F is continuous, then a local C^1 solution exists.*

However, uniqueness can fail, as the following example shows.

Example 2.7. Consider the equation $u'(x) = \sqrt{u}$ with $u(0) = 0$. One solution is $u = 0$. Alternatively, $u = x^2/4$ is another solution for $x > 0$. We can extend this second solution to a global solution by making it 0 for $x \leq 0$. Moreover, we can translate this solution to the left or right to get another solution. So there are infinitely many solutions.

Example 2.8. Consider the equation $u' = |u|^\alpha$. If we check $u = x^\beta$, we get that $\beta = \frac{1}{1-\alpha}$. We can consider this with a range of α, up to any $\alpha < 1$. What happens when $\alpha = 1$? The function $|u|^\alpha$ becomes Lipschitz.

Definition 2.5. A function F is **Lipschitz continuous** with **Lipschitz constant** L if

$$|F(x) - F(y)| \leq L|x - y| \qquad \forall x, y.$$

The Lipschitz functions form a Banach space when equipped with the norm $\|F\|_C + \|F\|_{\text{Lip}}$, where $\|F\|_{\text{Lip}} := \sup_{x,y} \frac{|F(x)-F(y)|}{|x-y|}$ which gives the "best" Lipschitz constant L.

Lipschitz functions have bounded slope, so it is reasonable to compare the spaces Lip and C^1. What is the relationship? We have $C^1 \subseteq \text{Lip}$. In 1 dimension, we can see this by the mean value theorem: $F(x) - F(y) = F'(c)(x - y)$ for some $x \in (x, y)$. For more than 1 dimension, we can still restrict the function to its values on a line connecting x, y to reduce to the 1 dimensional case.

However, $\text{Lip} \not\subseteq C^1$.

Example 2.9. The function $F(x) = |x|$ is 1-Lipschitz but not C^1.

Remark 2.1. It actually turns out that a Lipschitz function is differentiable outside a set of measure zero, but we will not use this.

This inclusion of Banach spaces is actually very nice because by the mean value theorem, we can use the same norm for both Lip and C^1.

2.3 Hölder continuous functions and fixed point methods

Starting from the continuous functions C^0, we have the subspaces $C^0 \supseteq \text{Lip} \supseteq C^1$. Is there anything in between C^0 and Lip?

Definition 2.6. The **α-Hölder continuous functions** are $C^\alpha(\mathbb{R}) = \{F : |F(x) - F(y)| \leq M|x - y|^\alpha\}$ for $0 < \alpha < 1$, equipped with the norm $\|F\|_{C^\alpha} := \sup \frac{|F(x)-F(y)|}{|x-y|^\alpha}$.

Remark 2.2. If $\alpha > 1$, the only functions that work are the constant functions.

Returning to our previous example, the function $|x|^\alpha$ is α-Hölder continuous.

Theorem 2.2. *If G is locally Lipschitz, then a local solution exists and is unique.*

Here is the beginning of the proof:

Proof. Restate the problem using the fundamental theorem of calculus. Integrating the equation gives

$$u(x) = u(0) + \int_0^x F(y, u(y))\, dy.$$

This allows us to think of the problem as a fixed point problem. Define the map $C^1 \ni u \mapsto N(u)(x) := u(0) + \int_0^x F(y, u(y))\, dy$. Observe that u solves our ODE if and only if $N(u) = u$. That is, we want u to be a fixed point of N.

In 1-dimension, if we have $f : \mathbb{R} \to \mathbb{R}$, when do we have fixed points $f(x) = x$? We can look for the points where the graph of f intersects the line $y = x$. One thing we can do to get fixed points is ask that the function does not increase very fast: $|f'| < 1$. In this case, f will have a unique fixed point.

We have just stated the following theorem:

Theorem 2.3. *If $f : \mathbb{R} \to \mathbb{R}$ with $|f'| < 1$, then f has a unique fixed point.*

This fact extends to Banach spaces.

Theorem 2.4. *Let B be a Banach space. If $f : B \to B$ is Lipschitz with Lipschitz constant $L < 1$ ($\|f(x) - f(y)\| \leq L\|x - y\|$), then f has unique fixed point.*

This is not sufficient for us because we are not looking at the entire space of C^1 functions. We only want local solutions.

Theorem 2.5 (Banach contraction principle)**.** *If $f : D \subseteq B \to D$ with D closed is Lipschitz with constant < 1, then f has a unique fixed point.*

Example 2.10. We need the domain D to be closed. If $D = (0, 1)$ and $f(x) = x/2$, then f has no fixed points. But adding the endpoints of the interval rectifies this.

Next time, we will further discuss this fixed point theorem.

3 Well-Posedness for ODEs

3.1 Local existence and uniqueness for ODEs

Last time, we were studying a local posedness theorem for ODEs.

Theorem 3.1. *Suppose F is locally Lipschitz; i.e. the restriction to any compact set is Lipschitz. Then the ODE*

$$\begin{cases} u' = F(x,u) \\ u(0) = u_0 \end{cases}$$

has a unique local solution $u \in \mathcal{C}^1([0,T])$.

Our main tool was Banach's contraction principle.

Lemma 3.1 (Contraction principle)**.** *Let $D \subseteq B$ be a closed subset of a Banach space, and let $N : D \to D$ be a contraction, i.e.* $\operatorname{lip}(N) < 1$. *Then N has a unique fixed point.*

This principle is useful not just in the study of ODEs but in PDEs as well. Here is a sketch of the proof.

Proof. We first prove uniqueness. Suppose $x = N(x)$ and $y = N(y)$. Then

$$\|x - y\| = \|N(x) - N(y)\| \leq \underbrace{L}_{<1} \|x - y\|.$$

This can only happen if $\|x - y\| = 0$, which implies $x = y$.

For existence, start with $x_0 \in D$. Try to improve your guess successively by setting $x_1 = N(x_0)$, $x_2 = N(x_1)$, and so on. To see that this is convergent, observe that

$$\|x_2 - x_1\| = \|N(x_1) - N(x_0)\| \leq L\|x_1 - x_0\|.$$

Iterating this gives

$$\|x_{n+1} - x_n\| \leq L^n \|x_1 - x_0\|.$$

This suggests we can think of x_n as a sort of geometric series:

$$x_n = \underbrace{x_n - x_{n-1}}_{\leq L^{n-1}} + \underbrace{x_{n-1} - x_{n-2}}_{\leq L^{n-2}} + \cdots + x_0.$$

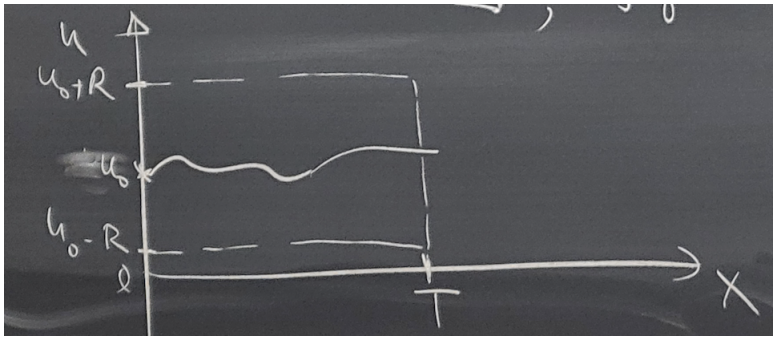
A geometric series is convergent, so the sequence x_n converges to some limit x. Since $x_{n+1} = N(x_n)$ taking the limit of both sides gives $x = N(x)$. □

This method of contraction is very useful when studying nonlinear PDEs. Now we can prove our ODE theorem:

Proof. We need N, B, and D. We obtain the map N by applying the fundamental theorem of calculus[2] to the ODE:

$$N(u)(x) = u_0 + \int_0^x F(y, u(y))\, dy.$$

Our Banach space will be $\mathcal{C}([0,T])$, where we need to figure out what is T. We want u to be locally Lipschitz, so we will define $D = \{u \in \mathcal{C}([0,T]) : \|u - u_0\|_{\mathcal{C}} \leq R\}$; we will also need to figure out what is R.



To figure out T, R, we have a few conditions:

1. We need N maps $D \to D$. For $u \in B(u_0, R)$,

$$\begin{aligned} |F(u)| &\leq |F(u_0)| + |F(u) - F(u_0)| \\ &\leq |F(u_0)| + LR \end{aligned}$$

   Suppose $R \leq 1$. Then

$$|N(u)(x) - u_0| \leq \int_0^x |F(y, u(y))|\, dy$$

   Bound this above by the length of the integral times the size of the integrand.

$$\leq T \cdot \underbrace{(|F(u_0)| + LR)}_{C}$$

   We can pick $T \ll 1$ is small enough such that

$$\leq \frac{R}{2}$$

2. N needs to be a contraction:

$$|N(u) - N(v)| \leq \int_0^x |F(y, u(y)) - F(y, v(y))|\, dy$$

[2]The idea is that the differential operator is unbounded, so you "lose" something when applying it. By contrast, when you integrate, you "gain" something.

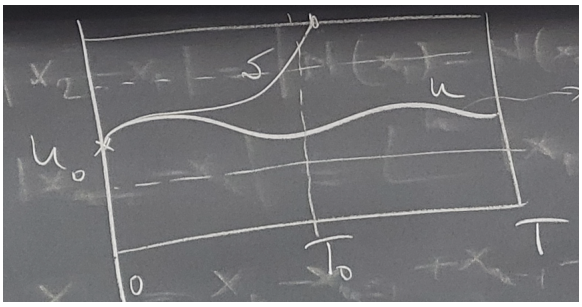
$$\leq \int_0^x L|u(y) - v(y)|\, dy$$
$$\leq T \cdot L \cdot \|u - v\|_{\mathcal{C}}.$$

Picking T small enough, we get

$$\|N(u) - N(v)\| \leq \underbrace{TL}_{<1} \|u - v\|.$$

By the contraction principle, there exists a unique solution u for the integral equation in D. If u solves the integral equation, then the right hand side of the integral equation is continuous. This implies that $u \in \mathcal{C}^1$ (as integrating a continuous function gives a $\mathcal{C}^1$ function).

The other issue is that our uniqueness statement is for functions in D. For uniqueness, is there any other solution which exits $B(u_0, R)$?

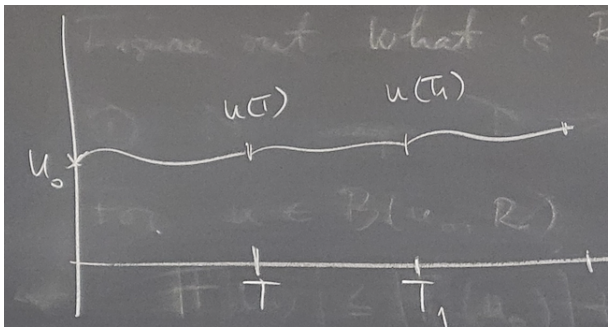


One solution is to find a T_0 small enough such that $u(T_0) \neq v(T_0)$ but $\|v - u_0\| \leq R$ in $[0, T_0]$ and apply the contraction principle in $[0, T_0]$. This gives $u = v$ in $[0, T_0]$.

Another solution is as follows. Denoting T_0 as the exit time of the ball of radius R, if $v : [0, T_0] \to B(u_0, R)$, our previous computation gives $\|v - u_0\| \leq R/2$. This is known as a **bootstrap argument**. □

3.2 Maximal solutions to ODEs

Now that we have proven existence and uniqueness of local solutions, let us move to the question of global solutions. Can we extend our local solution to global solutions?



This leads us to the idea of a maximal solution.

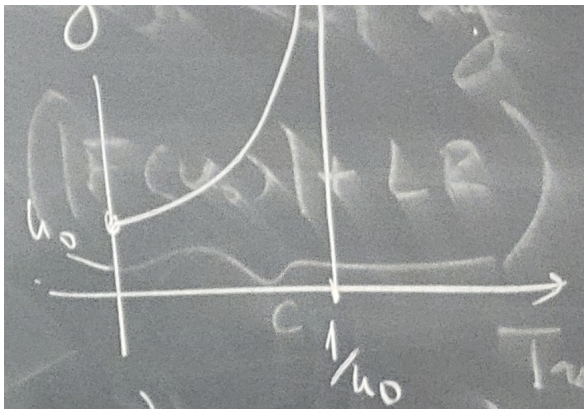
Definition 3.1. A **maximal solution** u is a solution to the differential equation that cannot be extended to a larger domain.

In general, global solutions may not exist!

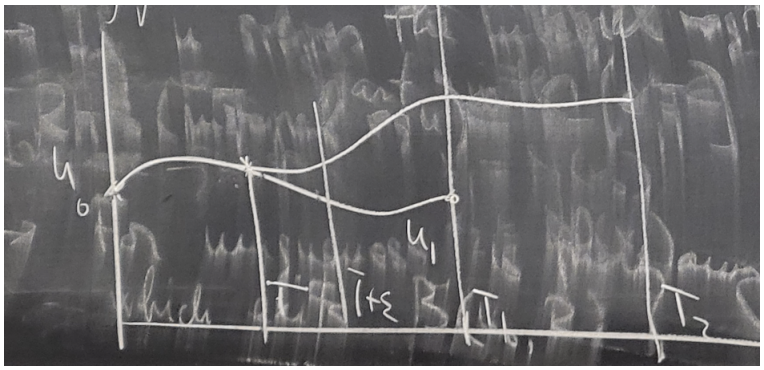
Example 3.1. Consider the equation

$$\begin{cases} u' = u^2 \\ u(0) = u_0 > 0. \end{cases}$$

By explicit computation, we can see $u(t) = \frac{1}{T-t}$, where $T = 1/u_0$.



How do we compute maximal solutions? Suppose $u_1 : [0, T_1] \to \mathbb{R}^n$ and $u_2 : [0, T_2] \to \mathbb{R}^n$ are two solutions. Can we compare them? Suppose $T_1 \leq T_2$. Then we can compare them up to time T_1.



Can this picture occur? Choose T to be maximal such that $u_1 = u_2$ in $[0, T]$. If $T < T_1$, then by local well-posedness, we must have $u_1 = u_2$ in $[T, T + \varepsilon]$. This contradicts the maximality of the choice of T, so we must have $T < T_1$. The conclusion is that as long as both solutions exists, the must be equal on the interval they share. The set of solutions is therefore ordered by inclusion, and a maximal solution exists.[3]

What can we say about maximal solutions? A maximal solution will look like $u : [0, T) \to \mathbb{R}^n$. The limit $\lim_{t \to T} u(t)$ cannot exist, or else we could solve the equation again from time T.

Proposition 3.1. *If $T < \infty$,*

$$\lim_{t \to T} |u(t)| = \infty.$$

Proof. Suppose not. Then there exists a sequence $t_n \to T$ such that $|u(t_n)| \leq M$. Start solving from t_n. We get a solution on the time interval $[t_n, t_n + T_n]$, where T_n is given by the local existence theorem. Since $|u(t_n)| \leq M$, the theorem gives $T_n = T_0$ not depending on n. If $t_n + T_0 > T$, then we get a contradiction because our solution extends beyond T. □

Remark 3.1. This proposition says nothing about what will happen to global solutions.

3.3 Continuous dependence on data

Suppose $u : [0, T] \to \mathbb{R}^n$ is our reference solution with data u_0, and we vary some v with initial data v_0. We want to know if $v_0 \to u_0$, does that mean $v \to u$ in $\mathcal{C}([0, T])$?

Theorem 3.2.

(a) If $|v_0 - u_0|$ is small enough, then v exists on $[0, T]$ and satisfies $\|v - u\|_\infty \leq 1$.

[3] We do not need the axiom of choice in this case because the time intervals are totally ordered, so we can just take the union.

(b) If $v_0 \to u_0$, then $v \to u$ in $\mathcal{C}([0,T])$.

Try to track $|u-v|^2$:

$$\begin{aligned}\frac{d}{dt}|u-v|^2 &= (u-v)\cdot\frac{d}{dt}(u-v)\\ &= (u-v)(F(u)-F(v))\\ &\le |u-v|\cdot L|u-v|\\ &= L|u-v|^2.\end{aligned}$$

We also have $|u-v|^2(0) = |u_0-v_0|^2$. Here, we have what might be called an **ordinary differential inequality** for $u-v$. If we had equality, then we would get $|u-v|^2 \le |u_0-v_0|^2e^{Lt}$. Otherwise, we hope to get $|u-v|^2 \le |u_0-v_0|^2e^{Lt}$. This step is the simplest form of what is known as Grönwall's inequality. Next time, we will discuss this inequality.

4 Continuous Dependence of ODEs on Initial Data and Classifications of PDEs

4.1 Continuous dependence of ODEs on initial data

Last time, we were discussing solving ODEs ofthe form

$$\begin{cases} u' = F(t,u) \\ u(0) = u_0. \end{cases}$$

We showed the following last time.

Theorem 4.1. *If F is locally Lipschitz, there exists a unique solution to the ODE.*

Today, we will talk more about continuous dependence of the solution on the initial data. So if we have $v' = F(t,v)$ with $v(0) = v_0$, we want to say that if $v(0)$ is close to $u(0)$, then v should be close to u.

Theorem 4.2. *Suppose that the solution u exists on $[0,T]$. Then there exists $\varepsilon > 0$ such that if $|v_0 - u_0| < \varepsilon$, then v exists on $[0,T]$ and*

$$\|u - v\|_C \le c|u_0 - v_0|.$$

That is, the map $u_0 \mapsto u|_{[0,T]}$ is locally Lipschitz.

Proof. We compute

$$\begin{aligned} \frac{d}{dt}|u-v|^2 &= 2(u-v)\cdot(u-v)_t \\ &= 2(u-v)\cdot(F(u)-F(v)) \end{aligned}$$

If F is Lipschitz,

$$\le 2L|u-v|^2.$$

So if $f(t) = |u-t|^2$, then $f'(t) \le 2Lf(t)$ with $f(0) = |u_0 - v_0|^2$. We claim that this implies that $f(t) \le f(0)e^{2Lt}$. This is called **Grönwall's inequality**.

Lemma 4.1 (Grönwall's inequality[4])**.** *If $f'(t) \le 2Lf(t)$, then $f(t) \le f(0)e^{2Lt}$.*

Proof. Let $g(t) = e^{-2Lt}f(t)$. It suffices to show that g is nonincreasing. We have $g'(t) = e^{-2Lt}f'(t) - 2Le^{-2Lt}f(t) \le 0$. □

The proof is finished except for:

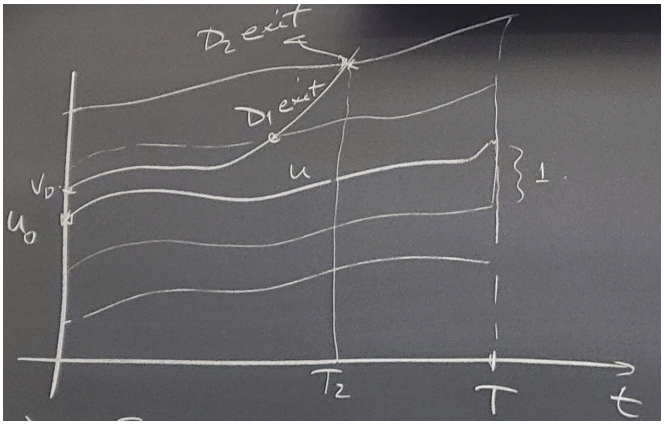
[4]More generally, we can prove this theorem with the same argument for $f'(t) \le h(t)f(t)$.

(a) If F is not globally Lipschitz.

(b) We do not know that v exists up to time T.

Suppose we have our solution u with initial data u. Consider two neighborhoods of u: a neighborhood $D_1 = \{v \in C([0,T]) : \|v-u\| \leq 1\}$ of size 1 and a neighborhood $D_2 = \{v \in C([0,T]) : \|v-u\| \leq 2\}$ of size 2.

Suppose we know that $v \in D_2$. Then v is defined on $[0,T]$, and stays in a compact set, so the above argument applies. How do we know v says in D_2? Suppose this is not true, so there is a time T_2 at which v exits D_2; then v must exit D_1 first.



By Grönwall's inequality applied to T_2, we have

$$\begin{aligned} |u(t)-v(t)|^2 &\leq |u_0-v_0|^2 \cdot e^{2LT_2}, \qquad t \in [0,T_2] \\ &\leq \varepsilon^2 e^{2LT} \end{aligned}$$

Choosing ε sufficiently small,

$$\leq 1.$$

This implies that v does not exit D_1, which is a contradiction; to exit D_2, v must first exit D_1. □

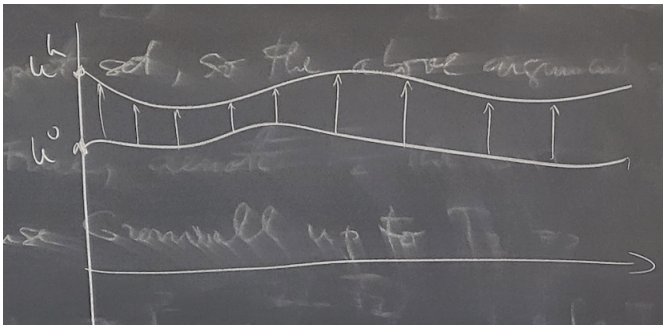
Remark 4.1. Suppose we want to prove that if $\varepsilon \ll 1$, then $\|u-v\| \leq 1$. We made a **bootstrap assumption** $\|u-v\| \leq 2$ and used this assumption to prove $\|u-v\| \leq 1$. This is called a **bootstrap argument**. These kind of bootstrap arguments are useful in nonlinear PDEs, when you don't even know whether a solution exists.

4.2 Linearizing an equation

Assume $F \in C^1$ and suppose we have initial data u_0^0. Take a one-parameter family of data u_0^h with h close to 0, so this is differentiable in h. Let u_0^0 give a solution u^0 and u_0^h give a

solution u^h. We can ask: how does u_h depend on h? We know that if $|u_0^h - u_0^0| \lesssim h$, then $|u^h - u^0| \lesssim he^{2LT}$ (with the notation $A \lesssim B$ meaning $A \leq cB$ for some constant c).

Here is a formal computation: If $\dot{u}^h = F(t, u^h(x))$, we want to compute an equation for $v^h = \frac{d}{dh}u^h$.



Apply $\frac{d}{dh}$ to get

$$\dot{v}^h = DF(t, u^h)v^h, \qquad v^h(0) = \frac{d}{dh}u_0^h.$$

This is a *linear* equation for v^h. It is called a **linearized equation**. This allows us to pass from one solution to another solution nearby.

Does the derivative actually exist? Let's compute:

$$\frac{d}{dt}(u^h - u^0) = F(t, u^h(T)) - F(t, u^0(t))$$

Think of this as a Taylor expansion

$$= DF(t, u^0(t))(u^h(t) - u^0(t)) + o(\underbrace{u^h(t) - u^0(t)}_{o(h)})^2.$$

Then

$$\frac{d}{dt}\frac{u^h - u^0}{h} = DF(t, u^0(t))\frac{u^h - u^0}{h} + o(h).$$

As $h \to 0$, $\frac{u^h - u^0}{h}(0) \to v^0$. So in the limit, we get $\frac{u^h - u^0}{h} \to v^0$, which is the solution to the linearized equation.

4.3 Classifications of first order scalar PDEs

We will study first order scalar PDEs. In these equations, we have $u : \mathbb{R}^n \to \mathbb{R}$, with

$$F(x, u, \partial u) = 0.$$

Evans' textbook uses Du instead of ∂u, but we will use this notation for something else later in the course.

Here is a classification by degree of difficulty:

- Linear:
$$\sum_j A_j(x)\partial_j u + B(x)u = f(x).$$
We can succinctly write this as $a \cdot \partial u + bu = f$.

- Semilinear:
$$\sum_j A_j(x)\partial_j u + b(x,u) = 0.$$
Here, the nonlinearity is only in u, not in the derivatives.

- Quasilinear:
$$\sum_j A_j(x,u)\partial_j u + b(x,u) = 0.$$

- Fully nonlinear:
$$F(x,u,\partial u) = 0.$$

If we differentiate a fully nonlinear PDE, we get a quasilinear PDE, but we get a system. For these equations, some things we know about scalar equations will not apply to systems.

What is our initial data? In $\mathbb{R}^n$, we take a surface Σ and specify $u|_\Sigma = u_0$ on the surface.

Definition 4.1. The equation plus our initial data is called an **initial value problem** or a **Cauchy problem**.

Another way we can classify partial differential equations is by static equations (at fixed time) and dynamic equations (evolution in time). This is a classification imposed less by the equations themselves and more by the motivation of the PDEs.

Example 4.1. The equation
$$u_t = F(x,u,\partial_x u)$$
with $u : \mathbb{R}_t \times \mathbb{R}_x \to \mathbb{R}$ is a dynamic or evolution equation. The **steady states** are solutions to the equation $0 = F(x,u,\partial_x u)$.

4.4 First order linear scalar PDEs

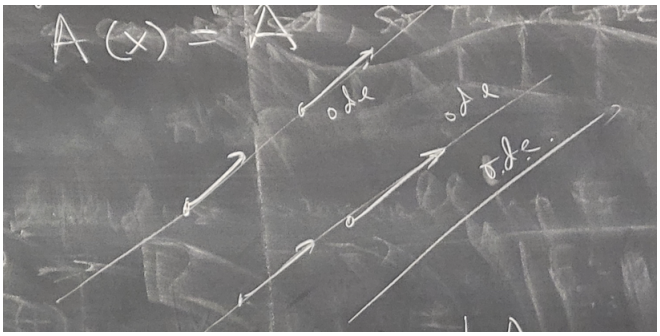
We are looking at the equation
$$\sum_j A_j(x) \cdot \partial_j u = bu + f,$$
which we can write as
$$A \cdot \nabla u = bu + f,$$

where $A \cdot \nabla u$ is the directional derivative of u in the direction A.

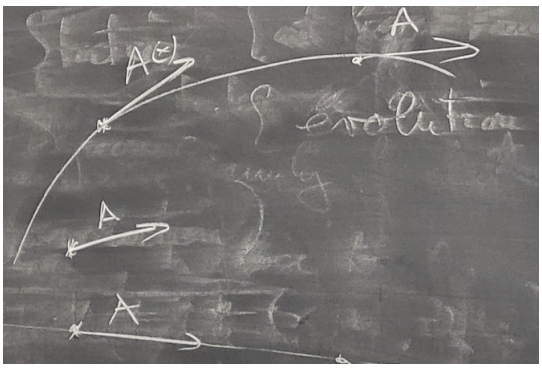
Let's start with a simpler case, where $A(x) = A$ does not depend on x. Then we can look at lines which point in the direction at A: $x = x_0 + tA$. Look at the function u along these lines: $u(x_0 + tA)$.

$$\begin{aligned}\frac{d}{dt}u(x_0 + tA) &= A\nabla u \\ &= bu(x_0 + tA) + f.\end{aligned}$$

This is a linear ODE for $u(x_0 + tA)$.



If A is not constant, can we do the same thing? Instead of straight lines, we need curves. In particular, we need curves which are tangent to A at each point.



Do such curves exist? The ODE $\dot{x}(t) = A(x(t))$ has C^1 solutions by ODE theory (where $A \in C^1$). So, given a point x, there is a unique curve starting from x that stays tangent to A. This is called an **integral curve** of A. We can calculate

$$\frac{d}{dt}u(x(t)) = \nabla u \cdot \dot{x}(t) = A\nabla u = bu(x(t)) + f,$$

which is an ODE for u. So if A is not constant, solving the PDE is like solving 2 ODEs: one that gives integral curves and one that tracks the solution u along each integral curve. Next time, we will look at what happens when we try to assign this initial data on a surface.

5 Local Solutions for Linear, Semilinear, and Quasilinear Scalar PDEs

5.1 Local solutions for linear, scalar PDEs

Last time, we were studying linear, scalar PDEs of the form

$$\underbrace{A_j \partial_j u}_{\text{directional derivative}} + bu = f.$$

The **initial curves** (or **characteristics**) of A were the solutions to the ODE

$$\dot{x} = A(x), \qquad x(0) = x_0$$

Along the integral curves, the PDE looks like

$$\frac{d}{dt} u(x(t)) + b(x(t))u(x(t)) = f(x(t)),$$

so solving the PDE is like solving two ODEs.

If we assume $A \in C^1$, then $x(t, x_0) \in C^1$. We want these characteristics to locally foliate $\mathbb{R}^n$; that is, we want them to cover the domain. One issue: what if $A(x_0) = 0$? Then $x(t) = x_0$ for all t!

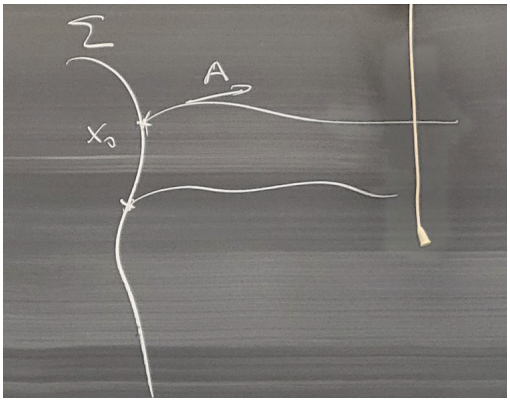
Example 5.1. Consider A that gives

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -x_1.$$

Then the integral curves will be circles, so $A(0) = 0$.

The fix for this problem is to assume that $A(x) \neq 0$ for any x.

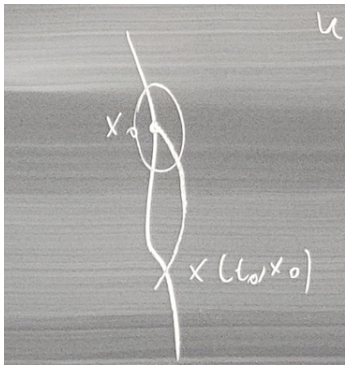
Now suppose we have initial data $u(x) = u_0(x)$ on a curve Σ. If we start at an x_0 on the curve or surface Σ, we can look at the integral curve starting from x_0.



From $x_0 \in \Sigma$ and $t \in [-\varepsilon, \varepsilon]$, we can construct $x(t, x_0)$. Once we know $u(x_0)$, we can solve the second ODE to get $u(x(t, x_0))$, where $x(t, x_0) \in C^1$. So by our ODE theorem, we will get $u \in C^1$.

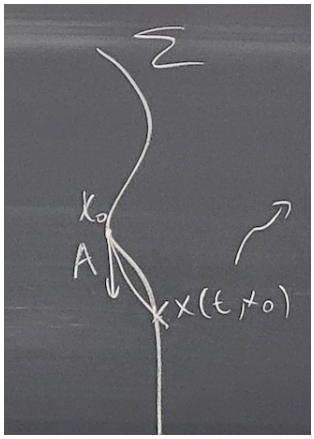
What are the bad cases?

- The integral curve may intersect Σ twice.



  We might still get a local solution if we look at a small enough neighborhood of x_0.

- A may be tangent to Σ, and re-intersection can happen arbitrarily close.



  Even if re-intersection is not arbitrarily close, there may be a more subtle issue with the solution not being C^1.

Here is how we avoid this issue.

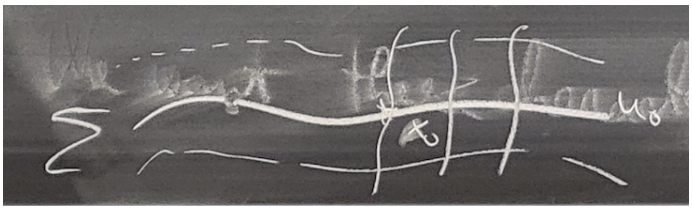
Definition 5.1. We say that Σ is **noncharacteristic** for our PDE if $A \cdot N \neq 0$ on Σ, where N is the normal to Σ.

This says that A is not tangent to Σ at any point.

Theorem 5.1. *Assume $A, b, f, \Sigma, u_0 \in C^1$, and suppose that Σ is noncharacteristic. Then the equation*

$$A_j \partial_j u + bu = f$$

with initial data u_0 has a unique C^1 local solution.



Proof. Step 1: For $x_0 \in \Sigma$, solve for the characteristic $\Sigma \times [-\varepsilon, \varepsilon] \ni (x_0, t) \mapsto x(x_0, t)$.

Step 2: Solve the ODE

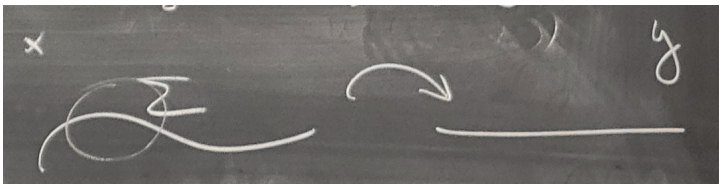
$$\frac{d}{dt} u(x(t)) + b(x(t))u(x(t)) = f(x(t))$$

along the characteristics to get $u(x(t, x_0))$, which is C^1 in t and x_0.

Step 3: Show that our characteristics foliate a neighborhood of Σ. What does this mean? Looking at the map $(x_0, t) \mapsto x(t, x_0)$. We want this to be a local diffeomorphism, i.e. a C^1 map with a C^1 inverse. Recall the following theorem from real analysis:

Theorem 5.2 (Local inversion theorem)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}^n \in C^1$. If $\det dF(x_0) \neq 0$, then F is a local diffeomorphism.*

We would like to change coordinates so that Σ is a hyperplane.



Since Σ is C^1, locally, Σ is the graph of a C^1 function, $x_n = f(x')$, $x' = (x_1, \ldots, x_{n-1})$ with $f \in C^1$. The new coordinates are $y = (x', x_n - f(x'))$. To check that this is a local diffeomorphism, the theorem says we should look at

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y'}{\partial x'} & \frac{\partial y'}{\partial x_n} \\ \frac{\partial y_n}{\partial x'} & \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} I_{n-1} & 0 \\ -df & 1 \end{bmatrix},$$

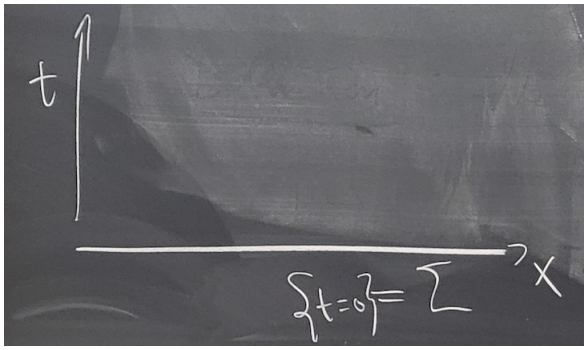
which has determinant 1. Check that the coefficients remain C^1 after changing coordinates.

In the new coordinates, $\Sigma = \{y_n = 0\}$, $y' = (y_1, \ldots, y_{n-1})$ are coordinate on Σ, and we are looking at the equation $\dot{y} = A(y)$. Here, $y = y(t, y_0')$. Look at $\frac{\partial y}{\partial(y_0', t)}$ at $t = 0$. When $t = 0$, $y(y_0', 0) = (y_0', 0)$. So

$$\frac{\partial(y', y_n)}{\partial(y_0', t)} = \begin{bmatrix} \frac{\partial y'}{\partial y_0'} & \frac{\partial y'}{\partial t} \\ \frac{\partial y_n}{\partial y_0'} & \frac{\partial y_n}{\partial t} \end{bmatrix} = \begin{bmatrix} I_{n-1} & 0 \\ A' & A_n \end{bmatrix}.$$

So $\det \frac{\partial y}{\partial(y_0', t)} = A_n \neq 0$, precisely from our noncharacteristic surface property. □

Remark 5.1. In the above proof, we reduced the situation to the case where Σ is a hyperplane. Let's use this to model the noncharacteristic case. Using coordinates (x, t), we can write $\Sigma = \{t = 0\}$.



Our equation looks like

$$A_t \cdot \partial_t u + A_1 \cdot \partial_1 u + \cdots + A_n \cdot \partial_n u + bu = f.$$

where $A_t \neq 0$ by the noncharacteristic assumption. So we may divide by it and just look at equations of the form

$$\partial_t u + A_1 \cdot \partial_1 u + \cdots + A_n \cdot \partial_n u + bu = f.$$

This is only a local modelling, however, not necessarily a global one.

5.2 Semilinear PDEs

Now we move on to solving semilinear PDEs, of the form

$$\begin{cases} A_j(x)\partial_j u + b(u, x) = 0 \\ u|_\Sigma = 0 \end{cases}$$

The characteristics are still $\dot{x} = A(x)$ (so $x = x(x_0, t)$), and our noncharacteristic initial surface condition is still $A \cdot N \neq 0$ on Σ. The evolution along the characteristics is

$$\frac{d}{dt} u(x(x_0, t)) = -b(u(x(x_0, t)), x(x_0, t)).$$

The difference from before is that our second equation is a nonlinear ODE, so it may have finite time blow-up. So local well-posedness is identical to the linear case, but global well-posedness may fail because the second ODE blows up.

5.3 Quasilinear PDEs

Now we look at the quasilinear problem

$$\begin{cases} A_j(x, u)\partial_j(u) + b(x, u) = 0 \\ u|_\Sigma = u_0. \end{cases}$$

Our characteristics now look like $\dot{x} = A(x, u)$. We cannot solve this because we do not know what u is outside of Σ. The second equation would read $\dot{u} = b(x, u)$. These two ODEs would be true if we already had a solution, but we cannot solve them. What if we put these two equations together into a system?

$$\begin{cases} \dot{x} = A(x, u) \\ \dot{u} = b(x, u) \end{cases}$$

We call this a **characteristic system**.

The initial data for the characteristic system is

$$\begin{cases} x(0) = x_0 \in \Sigma \\ u(0) = u(x(0)) = u_0(x_0), \end{cases}$$

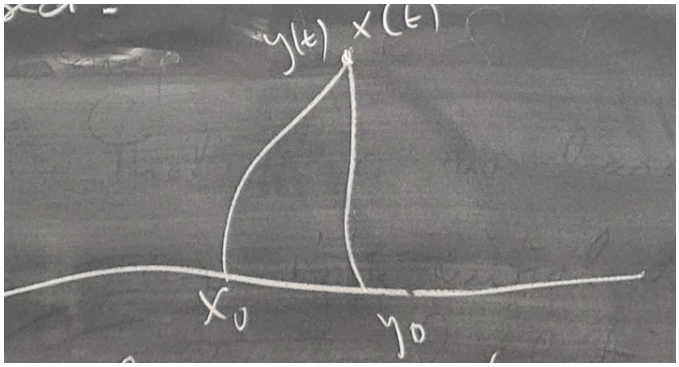
where the second initial condition depends on u_0. In this situation, our noncharacteristic Σ condition is

$$A(x_0, u_0(x_0)) \cdot N \neq 0.$$

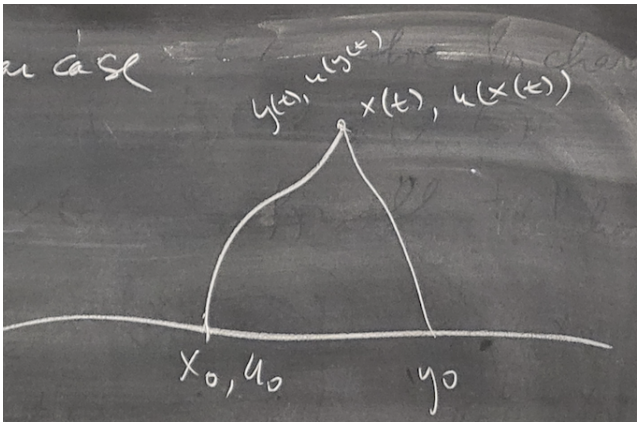
Our local well-posedness theorem is identical: If Σ is noncharacteristic and $u_0 \in C^1$, then there exists a unique local C^1 solution u.

The key difference is that the characteristics may now intersect. In the semilinear case, suppose two characteristics were to intersect. Then the characteristic equation would have

the same data, so the two characteristics must be the same.



In the quasilinear case, the initial data is both the location and the value of the function. Intersection means that $x(t) = y(t)$, but it does not necessarily mean $u(x(t)) = u(y(t))$. So we cannot say that the two characteristics must be the same.



Next time, we will talk about what might make characteristics intersect and what to do about it.

6 Quasilinear and Nonlinear First Order PDEs

6.1 Quasilinear PDEs and conversation laws

Last time, we were looking at first order, quasilinear, scalar PDEs

$$\sum_j A_j(x,u)\partial_j u + b(x,u) = 0.$$

We saw that our characteristics have to consider both x and u. We need to solve the characteristic system

$$\begin{cases} \dot{x} = A(x,u) \\ \dot{u} = -b(x,u) \end{cases}$$

to get a local solution. Because characteristics carry information about x and u, there was no prohibition against characteristics intersecting.

What is a noncharacteristic surface in this setting? If our initial data is $u|_\Sigma = u_0$, then the noncharacteristic condition becomes

$$A(x_0, u_0(x_0)) \cdot N \neq 0 \quad \text{on } \Sigma.$$

Remark 6.1. The condition of being noncharacteristic depends both on the surface and on the initial data on the surface. So the *problem* is noncharacteristic, rather than the surface (until we have a fixed set of initial data).

The model problem is

$$\begin{cases} \partial_t u + \sum_j A_j(x,u)\partial_j u + b(x,u) = 0 \\ u|_{t=0} = u_0 \end{cases}$$

Since we are already using t, let's use s as the parameter along the chracteristics. We have

$$\begin{cases} \dot{t} = 1 \\ \dot{x} = A(x,u) \\ \dot{u} = -b(x,u) \end{cases}$$

The first equation tells us that we can choose $s = t$. This corresponds to a dimensionality reduction of our problem.

Example 6.1. A special case of this is what we call **conservation laws**:

$$u_t + \partial_j F_j(u) = 0.$$

We can equivalently write this as

$$u_t + F_j'(u)\partial_j u = 0.$$

Using the first form is not important for scalar equations, but it is for scalar systems because it is not always the case that we can write the second version with a divergence term.

The first version is called **density flux notation**. This is because the u_t tells how the density of some quantity changes in time, and the flux term, $\partial_j F_j(u)$, tells you how the mass is moving with velocity $F_j'(u)$.

6.2 Burgers' equation

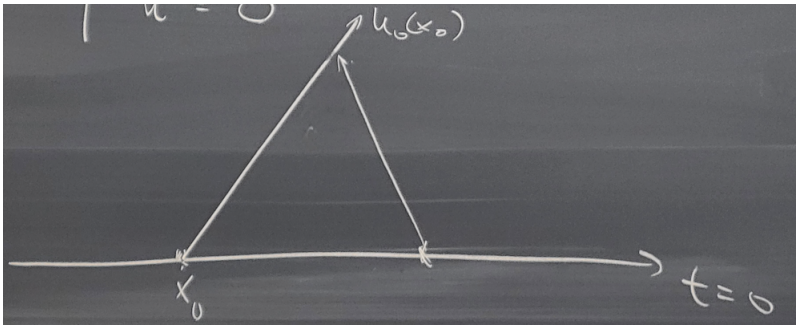
Example 6.2. The simplest quasilinear problem is the **Burgers' equation**

$$\begin{cases} u_t + uu_x = 0 \\ u|_{t=0} = u_0. \end{cases}$$

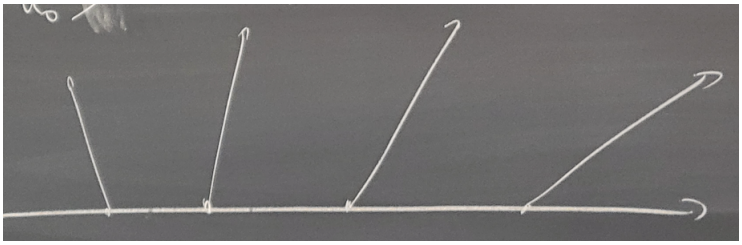
This equation seems simple, but it ends up being a model problem for more complicated equations. Here are the characteristics:

$$\begin{cases} \dot{x} = u \\ \dot{u} = 0. \end{cases}$$

Thus, the characteristics are $x(t) = x_0 + tu_0(x_0)$. Here, the characteristics may intersect as follows:



How would we choose our data so the lines don't intersect? If u_0 is increasing, the picture looks like this:



So we get a global solution forward in time, but we don't get a global solutions backward in time. So the only global solutions are constant.

Remark 6.2. In physics, we expect there to be *causality*. That is, we expect the future to be determined by the past but not the past to be determined by the future. Later we will see what we will do after the point where characteristics intersect.

Let's give an equation for u_x:

$$u_{tx} + uu_{xx} + u_x^2 = 0.$$

If we write $u_x = v$, then this equation is just talking about the derivative along the characteristics:

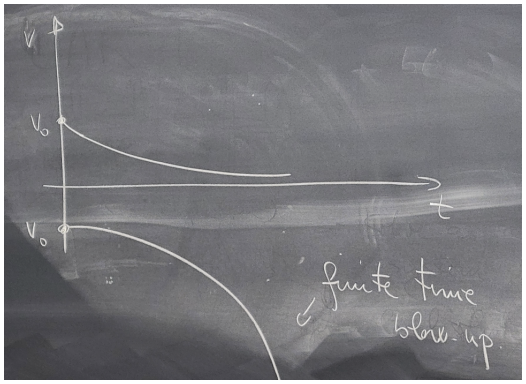
$$(\partial_t + u\partial_x)v + v^2 = 0.$$

We may also write this as

$$\dot{v} + v^2 = 0,$$

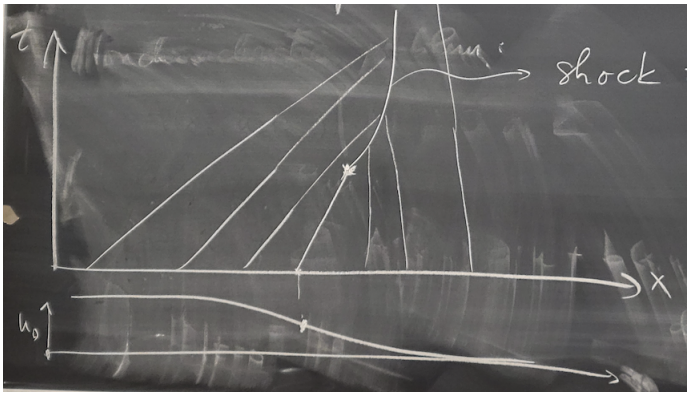
where the dot is the derivative along the characteristic. This equation tells us how the slope of the solution is evolving.

If $v_0 > 0$, the slope decreases toward 0. However, if $v_0 < 0$, we get finite time blow-up.



The smallest slope means the fastest blow-up. Suppose the initial data u_0 is decreasing, so we will get intersections of characteristics. Then the time with the most negative slope

will be the time of blow up for x.



Because things intersect, there is no unique way to continue the equation. Here, we have a **shock**, or a jump discontinuity. We will see later how to find what the equation for the shock curve looks like.

Conservation laws is still a very active area, with a number of hard problems.

6.3 Fully nonlinear problems

We now look at PDEs of the form

$$\begin{cases} F(x, u, \partial u) = 0 \\ u|_\Sigma = u_0, \end{cases}$$

where the dependence on ∂u is nonlinear. Where do we start? Before, we had a vector field that let us interpret the equation using a directional derivative.

Let's look at the linearized equation: Suppose we have not just a solution but a 1-parameter family of solutions u^h to our problem with solution v^h to our linearized equation given by

$$\frac{d}{dh} u^h = v^h.$$

Differentiate the equation with respect to h to get the linearized equation:

$$0 = \frac{\partial}{\partial h} F(x, u^h, \partial u^h) = F_u \cdot v^h + F_{p_j} \cdot \partial_j v^h,$$

where we write $F = F(x, u, p)$ and $p = (p_1, \ldots, p_n)$ (in $\mathbb{R}^n$).

This linearized equation is a linear transport equation. So we get a vector field $A_j = F_{p_j}(x, u, \partial u)$. We should try to use this vector field to find characteristics. Our equation looks like

$$\begin{cases} \dot{x}_j = F_{p_j}(x, u, \partial u) \\ \dot{u} = \cdots \end{cases}$$

The first equation depends on ∂u, so we may try to add an equation $\dot{\partial u} = \cdots$. But then we would get $\partial^2 u$ in this equation, and we would be in the same situation. How do we get past this issue?

Suppose $F(u, \partial u) = 0$. We say that this equation is invariant with respect to translations. This means that if $u(x)$ is a solution, $u(x + hy)$ is a solution, as well. This produces a 1-parameter family of solutions. This implies that $y \cdot \partial u$ solves the linearized equation. In particular, we can use this as our equation for $\dot{\partial u}$. Here is the computation:

$$\begin{cases} \dot{x}_j = F_{p_j}(x, u, \partial u) \\ \dot{u} = F_{p_j}(x, u, \partial u) \cdot \partial u \qquad = \text{directional derivative of } u \text{ in the } F_{p_j} \text{ direction} \\ \dot{\partial_j u} = -F_{x_j}(x, u, \partial u) - F_u(x, u, \partial u) \cdot \partial_j u, \end{cases}$$

where we calculate the last equation by

$$\begin{aligned} 0 &= \partial_j F(x, u, \partial u) \\ &= F_{x_j}(x, u, \partial u) + F_u(x, u, \partial u)\partial_j u + \underbrace{F_{p_k}(x, u, \partial u)\partial_k \partial_j u}_{\dot{\partial_j u}}. \end{aligned}$$

We still have a problem. Suppose we solve the above system. We are treating the function u and its derivatives as separate objects, so how do we know that the solutions are still related to each other? First, let's summarize what we have done so far in a proposition:

Proposition 6.1. *If $u \in C^2$, then $(x, u, \partial_j u)$ solve the characteristic system*

$$\begin{cases} \dot{x}_j = F_{p_j}(x, u, \partial u) \\ \dot{u} = F_{p_j}(x, u, \partial u) \cdot \partial u \\ \dot{\partial_j u} = -F_{x_j}(x, u, \partial u) - F_u(x, u, \partial u) \cdot \partial_j u. \end{cases}$$

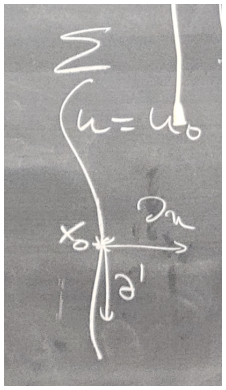
When we solve the system, use the notation z instead of u and p_j instead of $\partial_j u$ because we are solving this equation without enforcing the relationship between these objects. The characteristic system becomes

$$\begin{cases} \dot{x}_j = F_{p_j}(x, z, p) \\ \dot{z} = F_{p_j}(x, z, p) \cdot p_j \\ \dot{p}_j = -F_{x_j}(x, z, p) - F_z(x, z, p) \cdot p_j. \end{cases}$$

What is the initial data for this system? We had $x(0) = x_0$ and $u(0) = u_0$ before, but now we have

$$\begin{cases} x(0) = x_0 \\ u(0) = u_0 \\ \partial u(0) = ? \end{cases}$$

We need the information of *all* the derivatives of u at x_0. In particular, we need both $n-1$ tangential derivatives to Σ and 1 normal partial derivative to Σ.



If we frame this in the tangent space, we want the tangent derivative $\partial' = (\partial_1, \ldots, \partial_{n-1})$ and the normal derivative ∂_n. We know ∂', but what about ∂_n? We know that

$$F(x_0, u_0, \partial' u_0, \partial_n u) = 0,$$

so we would like to solve for $\partial_n u$. This tells us that

$$\partial_n u = G(x_0, u_0, \partial' u_0)$$

for some function G. We can do this if

$$F_{p_n}(x_0, u_0, \partial' u_0, p_n) \neq 0.$$

If we did not put our equation in this special frame, this condition reads as

$$F_p(x_0, u_0, p) \cdot N \neq 0,$$

the condition that the equation is noncharacteristic.

Remark 6.3. What if this equation has more than 1 solution? We may not get uniqueness; the answer may depend on our choice here of initial data.

7 Existence of Solutions to Nonlinear First Order Scalar PDEs

7.1 Proving existence and uniqueness given initial data

Last time, we were looking at fully nonlinear equations

$$\begin{cases} F(x, u, \partial u) = 0 \\ u = u_0 \text{ on } \Sigma. \end{cases}$$

If u solves this equation, then $(x, u, \partial_j u)$ solves the characteristic system

$$\begin{cases} \dot{x} = F_p(x, z, p) \\ \dot{z} = F_p(x, z, p) \cdot p \\ \dot{p} = -F_x(x, z, p) - F_z(x, z, p) \cdot p. \end{cases}$$

The initial data for the characteristic system on Σ is

$$\begin{cases} x(0) = x_0 \\ z(0) = u_0(x_0) \\ p(0) = p_0, \end{cases}$$

where p_0 has a tangential component $\partial_\tau u_0$ and a normal component given by solving $F(x_0, u_0, p_0)$. In this last part, we had a local solvability condition $F_p \cdot N \neq 0$, where N is the normal to Σ. This is the same as the noncharacteristic condition.

Our objective is to turn this into an existence proof.

Theorem 7.1. *Assume that F is of class C^2, Σ is C^2, $u_0 \in C^2$, and the problem is noncharacteristic, i.e. there exists p_0 on Σ such that $F_{p_0} \cdot N \neq 0$, $F(x_0, u_0, p_0) = 0$, and $(p_0)_\tau = \partial_\tau u_0$. Then there exists a unique local solution $u \in C^2$ near Σ such that $u|_\Sigma = u_0$ and $\partial u|_\Sigma = p_0$.*

Proof. First, an outline:

Step 1: Solve the characteristic system with initial data (x_0, u_0, p_0) on Σ. This gives us

$$(x(s, x_0), u(s, x_0), p(s, x_0)),$$

which we can solve by using ODE theory.

Step 2: Show that the map

$$\Sigma \times [-\varepsilon, \varepsilon] \ni (x_0, s) \mapsto x(x_0, s) \in \mathbb{R}^n$$

is a local diffeomorphism with inverse

$$x \mapsto (x_0, s).$$

Step 3: Define

$$u(x(s, x_0)) = z(s, x_0).$$

This is true if a solution u exists.

The main difficulty is that at the end of our construction, we get the functions

$$z(s, x_0) = u(x), \qquad x = x(s, x_0), \qquad p_j(s, x_0) \stackrel{?}{=} \partial_j z(x).$$

Our final goal is to prove that $p_j(s, x_0) = \partial_j z(s, x_0)$. By construction of our initial data, we know this is true at $s = 0$. Ideally, we might want to show that $\frac{\partial}{\partial s}(p_j - \partial_j z) = 0$. Instead, we will have a weaker version that works:

$$\frac{\partial}{\partial s}(p_j - \partial_j z) = \text{coeff}(p_j - \partial_j z),$$

which is a linear ODE for $p_j - \partial_j z$.

Our preliminary step is to show that $F(x, z, p) = 0$. This is true on Σ, i.e. when $s = 0$.[5] Compute

$$\frac{d}{ds} F(x, z, p) = F_x \cdot \dot{x} + F_z \cdot \dot{z} + F_p \cdot \dot{p} = 0.$$

Next, compute $\frac{\partial}{\partial s}(p_j - \partial_j z)$. We have

$$\frac{\partial}{\partial s} = (-F_{x_j} - F_z \cdot p_j),$$

but to calculate $\frac{\partial}{\partial s}\partial_j z$, we need to use $\dot{z} = F_p \cdot p$. We have $\frac{\partial}{\partial s} = F_{p_k} \cdot \frac{\partial}{\partial x_k}$, where F_{p_k} has variable coefficients. So the derivatives do not commute. We can explicitly compute

$$\frac{\partial}{\partial s}\partial_j z = F_{p_k}\partial_k \partial_j z,$$

$$\partial_j \dot{z} = \partial_j (F_{p_k}\partial_k z) = F_{p_k}\partial_j \partial_k z + \partial_j F_{p_k} \cdot \partial_k z,$$

which gives

$$\frac{\partial}{\partial s}\partial_j z = \partial_j \dot{z} - \partial_j F_{p_k} \cdot \partial_k z.$$

So we get

$$\begin{aligned} \frac{\partial}{\partial s}(p_j - \partial_j z) &= -F_{x_j} - F_z \cdot p_j - \partial_j \dot{z} + \partial_j (F_{p_k}) \cdot \partial_k z \\ &= -F_{x_j} - F_z \cdot p_j - \partial_j (F_{p_k} \cdot p_k) + \partial \cdot (F_{p_k})\partial_k z \end{aligned}$$

[5] This is the same thing we wanted to do with $p_j - \partial_j z$, but that is more difficult to work with because that is a vector equation, rather than just a scalar equation.

$$= -F_{x_j} - F_z \cdot p_j - F_{p_k} \partial_j p_k \underbrace{-p_k(F_{x_j p_k} + F_{z p_k} \partial_j z + F_{p_\ell p_k} \partial_j p_\ell) + \partial_k z(\text{same})}_{-(p_k - \partial_k z) \cdot \partial_j F_{p_k}}$$
$$= -F_{x_j} - F_z \cdot p_j - F_{p_k} \partial_j p_k + \text{good}.$$

We also have

$$F_{x_j} + F_z \cdot \partial_j z + F_{p_k} \partial_j p_k = 0$$

by taking $\frac{\partial}{\partial x_j}$ of our earlier computation. This last term $F_{p_k} \cdot \partial_j p_k$ is the same worst term in the above expression. If we substitute, we get

$$\frac{\partial}{\partial_s}(p_j - \partial_j z) = -F_z(p_j - \partial_j z) - \partial_j F_{p_k}(p_k - \partial_k z),$$

which is a linear system.

Therefore, z is the solution to our equation, and we are done. □

7.2 Problems in standard form

Example 7.1. Begin with the equation

$$u_t + F(t, x, u, \partial u) = 0$$

We will label u_t as τ, u as z, and ∂u as p. So we get the equation

$$\widetilde{F}(t, x, z, \tau, p) = \tau + F(t, x, z, p) = 0,$$

and the system

$$\begin{cases} \dot{t} = 1 \text{ (so } s = t) \\ \dot{x} = F_p \\ \dot{z} = \tau + F_p \cdot p = F_p \cdot p - F \\ \dot{p} = -F_x - F_z \cdot p \\ \dot{\tau} = -F_t - F_z \cdot \tau \end{cases}$$

In the middle 3 equations, we have no τ terms, so we can discard the last equation. Another way to think of this is that $\widetilde{F} = 0$, so τ is already given as $-F$. So we get a smaller system

$$\begin{cases} \dot{x} = F_p \\ \dot{z} = F_p \cdot p - F \\ \dot{p} = -F_x - F_z \cdot p. \end{cases}$$

The price we pay is the extra F term in the second equation, compared to before.

Remark 7.1. Solutions are local, near Σ, until characteristics may intersect. There is no way to continue solutions in general past this intersection of characteristics. For specific classes of problems, however, there is hope.

Example 7.2. Suppose we have an equation $H(x, \partial u) = 0$ which does not depend directly on u. Then we get

$$\begin{cases} \dot{x} = H_p \\ \dot{p} = -H_x \\ \dot{z} = H_p \cdot p - H. \end{cases}$$

The first two equations do not depend on z, so we can discard the last equation, solve the first two equations first, and integrate the last equation at the end.

This type of system is called a **Hamilton flow**.[6] Many PDEs can be interpreted as Hamiltonian flows. The **Hamilton-Jacobi** equations are of the form

$$u_t + H(x, \partial u) = 0.$$

Next time, we will do a bit of variational calculus to not only solve Hamilton-Jacobi equations but to also see how we may extend a solution past a point where characteristics intersect. In a Hamilton flow, the characteristics only depend on (x, p). When characteristics intersect, they may have the same x but different $p = \partial u$. We will try to continue the solution in a way such that ∂u has a jump discontinuity.

Example 7.3. Consider the equation

$$\begin{cases} u_t + \frac{1}{2}|\partial_x u|^2 = 0 \\ u(0) = u_0. \end{cases}$$

Here, $H(p) = \frac{1}{2}p^2$, and we get the system

$$\begin{cases} \dot{x} = p \\ \dot{p} = 0. \end{cases}$$

Here, the characteristics are straight lines, with $p(0) = \partial_x u_0$.

Example 7.4 (Eikonal equation)**.** The equation

$$|u_t|^2 - |\partial_x u|^2 = 0.$$

is not in the form we have talked about already. This gives

$$u_t = \pm|\partial_x u|,$$

so we will get 2 solutions.

[6] Hamilton flows play a role in symplectic geometry.

8 Expressing Hamilton-Jacobi Equations in Terms of Calculus of Variations

8.1 Recap: Hamilton-Jacobi equations

Last time, we started talking about Hamilton-Jacobi equations, as an example of first order PDEs:

$$\begin{cases} u_t + H(x, Du) = 0 \\ u(0) = u_0 \end{cases}$$

The characteristics for this system were given by

$$\begin{cases} \dot{u} = H_p(x, p) \\ \dot{p} = -H_x(x, p) \\ \dot{z} = H_p(x, p) \cdot p - H(x, p) \end{cases}$$

with initial data

$$\begin{cases} x(0) = x_0 \\ p(0) = \partial_x u_0. \end{cases}$$

The equations for $\dot{u}$ and $\dot{p}$ are called the **Hamilton equations**. We noticed that we only need to solve them first to get the characteristics, and then we can integrate the $\dot{z}$ equation to solve it after the fact.

8.2 Calculus of variations

Today, we will be looking at the calculus of variations. Here is the setup: We have a function $L(x, q)$ we call the **Lagrangian**, and to each function $x : [0, T] \to \mathbb{R}$, we associate to this function an **action functional**

$$\mathcal{L}(x) = \int_0^T L(x, \dot{x})\, dt.$$

The question we want to ask is: what are the minimizers of $\mathcal{L}$? We are looking for

$$\min_{x:[0,T]\to\mathbb{R}} \mathcal{L}(x).$$

We can think of $\mathcal{L}$ giving the cost of the trajectory x. So we want to find the most efficient trajectory x.

If we were just minimizing a function in $\mathbb{R}^n$, we would look for critical points. In particular, for $f : \mathbb{R}^n \to \mathbb{R}$, a minimum point in a critical point if $\nabla f = 0$. How do we do this in the case of our functional? We can talk in terms of directional derivatives. Replace

x by $x+hy$ and look at the map $h \mapsto \mathcal{L}(x+hy)$, where $h=0$ is a minimum point. Assume that our perturbation y is compactly supported. In this case, at $h=0$, we have

$$\begin{aligned} 0 &= \frac{d}{dh}\mathcal{L}(x+hy) \\ &= \frac{d}{dh}\int_0^T L(x+hy, \dot{x}+h\dot{y})\,dt \\ &= \int_0^T L_x(x,\dot{x})\cdot y + L_q(\dot{x})\cdot \dot{y}\,dt, \end{aligned}$$

where we are using q as a placeholder for the second variable, as we did with p before. This holds for all $y \in C_0^\infty([0,T])$. To deal with the $\dot{y}$ term, we integrate by parts (using the compact support assumption):

$$= \int_0^T y(L_x(x,\dot{x}) - \frac{d}{dt}L_q(x,\dot{x}))\ dt$$

when integrated against any function with compact support, the part inside the parentheses gives 0. So it must equal 0, Thus, we have actually proven a theorem:

Theorem 8.1 (Euler-Lagrange equation)**.** *x is a critical point for $\mathcal{L}$ if and only if it solves*

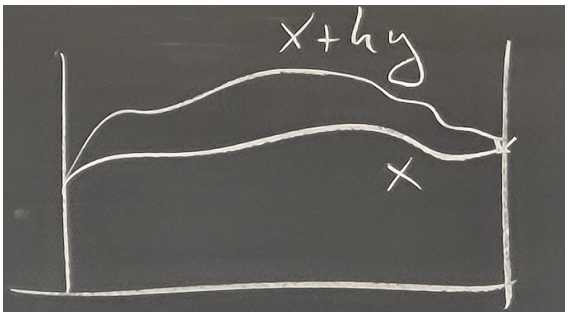
$$L_x(x,\dot{x}) - \frac{d}{dt}L_q(x,\dot{x}) = 0.$$

Remark 8.1. The PDE analogue takes a function $u : \mathbb{R}^n \to \mathbb{R}$ and gives the Euler-Lagrange equation

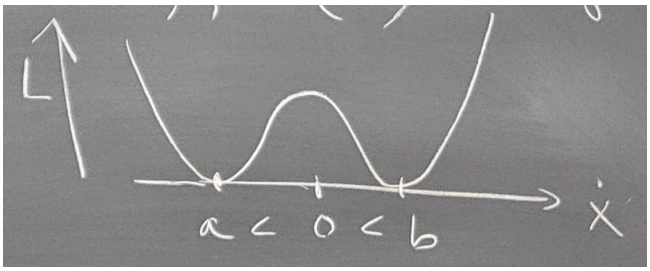
$$L_x(u,\partial u) - \partial_j L_{q_j}(u,\partial u) = 0,$$

which is a second order PDE.

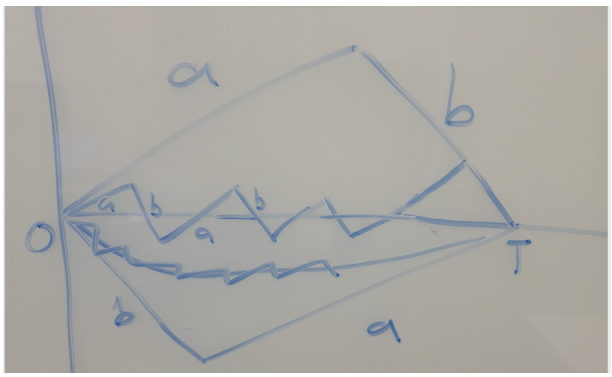
Remark 8.2. Our perturbation does not change the values at the endpoints $x(0), x(T)$, so it gives critical points in a context where $x(0)$ and $x(T)$ are fixed.



Remark 8.3. Suppose $L = L(\dot{x})$ is the following "double well potential."

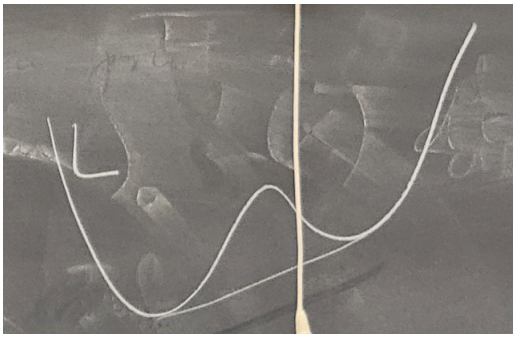


Suppose also that $x(0) = x(T)$. We want to minimize $\int_0^T L(\dot{x})\,dt \geq 0$. Can we achieve 0? We can make a line with slope a and then a line with slope b to get 0 as the minimum (notice that this is not differentiable!). Alternatively, we can alternate between lines of slope a and b in any number of ways as follows:

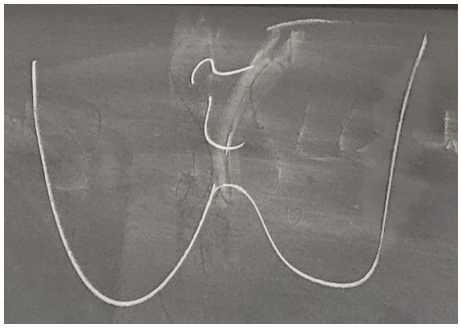


So we get that the infimum is 0 (since we can approximate any piecewise function by smoothing out the corners), and the minimum is 0 if we allow for any Lipschitz function x. In fact, all trajectories with slopes between $[a, b]$ are limiting minimizers. This means we are actually dealing with an **effective Lagrangian** L_{eff} with the hump between a and b flattened out. The effective Lagrangian L_{eff} is the **convex envelope** of L.

If we had another Lagrangian like the following, could we again look at the convex envelope?



Suppose we add a linear constant to get $\widetilde{L}(q) = L + c \cdot q$. Then we get the following picture, which is the same as before:



So the effective Lagrangian must be convex as a function of q. For PDEs, convexity is no longer required. Instead, we require **rank one convexity**, which is given by convexity in one variable at a time.

Example 8.1. Here is an example that comes from classical mechanics. Suppose we have a particle with trajectory $x(t)$ moving in a conservative force field $F = \nabla\phi$, where ϕ is the potential. Then we have the Lagrangian

$$L(x, q) = \underbrace{\frac{1}{2}mq^2}_{\text{kinetic energy}} - \underbrace{\phi(x)}_{\text{potential energy}},$$

where we have $\phi_x = \frac{d}{dt}(m\dot{x})$, which we can write as $m \cdot \ddot{x} = F(x)$, which is Newton's law.

8.3 Connecting the Hamilton-Jacobi equations to the Euler Lagrange equations

Returning to Hamilton-Jacobi equations, we have x, p with the function H, and we want to relate this to the $x, q = \dot{x}$ and L in the Euler-Lagrange equation. We can think of the Euler-Lagrange equation as a system for x and q via

$$\begin{cases} \dot{x} = q \\ \frac{d}{dt}L_q(x, q) = L_x. \end{cases}$$

We want to let $p = L_q(x, q)$. For this to make sense, we need $q \mapsto L_q(x, q)$ to be a diffeomorphism from $\mathbb{R}^n \to \mathbb{R}^n$ for fixed x.

Proposition 8.1. *If $L : \mathbb{R}^n \to \mathbb{R}$ is strictly convex and **coercive** (meaning $\lim_{q\to\infty} \frac{L(q)}{|q|} = \infty$), then $q \mapsto L_q$ is a diffeomorphism.*

Proof. Injectivity: L is strictly convex, so the graph of L is above its tangent lines at points of nonintersection:

$$L(y) > L(x) + (y - x)DL(x), \qquad y \neq x.$$

We can use this to write

$$(y - x)(DL(y) - DL(x)) > 0, \qquad y \neq x.$$

This gives injectivity.

Surjectivity: We want to minimize $L(x, q) - p \cdot q$. If a minimum exists, then the gradient must equal 0:

$$L_q(x, q) = p,$$

which is our surjectivity. Why must the minimum exist? This is because $\lim_{q\to\infty} L(x, q) - p \cdot q = \infty$ by coercivity.

To check that this is a local diffeomorphism, the differential of $q \mapsto L_q(x, q)$ is $L_{qq} \geq 0$. In fact, by strict convexity, this is > 0. □

So we have $p = L_q(x, q)$. We will define $H(x, p) = \max_q p \cdot q - L(x, q)$, Note that this is the same quantity we dealt with in the above proof. The functions $p \cdot q - L(x, q)$ are linear in p, so this maximum is convex.

Proposition 8.2. *H is convex and coercive.*

Proof. This comes from the strict convexity and coercivity of L. □

Proposition 8.3.

$$q = H_p(x, p).$$

Proof. This is a maximum, so $H(p) + L(q) - pq \leq 0$, with equality if $p = L_q(x, q)$. Now fix q and vary p! Then p is a maximum point for this expression when the derivative $H_p(p) - q = 0$. □

Now let's change our variables: The Euler-Lagrange equations say

$$L_x(x, q) - \frac{d}{dt}\underbrace{L_q(x, q)}_{p} = 0$$

So we get

$$\begin{cases} \dot{p} = L_x(x, q) \stackrel{?}{=} -H_x(x, p) \\ \dot{x} = q = H_p(x, p). \end{cases}$$

We have

$$H(x, p) + L(x, q) - p \cdot q \leq 0,$$

If we think of $p = p(x, q)$, we can take $\frac{d}{dx}$ to get

$$H_x(x,p) + L_x(x,q) + \underbrace{(H_p(x,p) - q)}_{=0} \cdot \frac{\partial p}{\partial q} = 0.$$

So this gives us our relationship between H_x and L_x.

9 Solutions to Hamilton-Jacobi Equations via Calculus of Variations

9.1 Recap: Connecting Hamilton-Jacobi equations to calculus of variations using the Legendre transform

Last time, we wanted to compare Hamilton-Jacobi equations to calculus of variations. The Hamilton-Jacobi equations are of the form

$$\begin{cases} u_t + H(x, \partial u) = 0 \quad \text{in } \mathbb{R} \times \mathbb{R}^n \\ u(0) = u_0 \quad \text{in } \mathbb{R}. \end{cases}$$

The characteristics given to this equation are

$$\begin{cases} \dot{x} = H_p \\ \dot{p} = -H_x \\ \dot{z} = p \cdot H_p - H, \end{cases}$$

with initial data $x(0) = x_0$ and $p(0) = \partial u_0$. The first two equations are called the **Hamilton flow**.

In calculus of variations, we have a **Lagrangian** $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, and we want to minimize an **action functional**

$$\min_{x \in \mathcal{A}} \underbrace{\int_0^T L(x, \dot{x})\, dt}_{\mathcal{L}(x)},$$

where $\mathcal{A} = \{x : [0, T] \to \mathbb{R} \text{ Lipschitz} \mid x(0) = x_0, x(T) = x_T\}$. Minimizers satisfy the **Euler-Lagrange equation**

$$L_x(x, \dot{x}) - \frac{d}{dt} L_q(x, \dot{x}) = 0.$$

Last time, we connected these two setups. We saw that

- L is strictly convex and coercive if and only if H is strictly convex and coercive.
- $$H(x, p) = \max_{q \in \mathbb{R}^n} -L(x, q) + p \cdot q,$$
  which is maximized at $p = L_q(x, q)$. This relation gives
  $$H(x, p) + L(x, q) \geq p \cdot q$$
  with equality when $p = L_q(x, q)$. This expression is symmetric in p and q, so it allows us to cast q in terms of p: $q = H_p(x, p)$. This relationship is known as the **Legendre transform**.

Remark 9.1. The Legendre transform well-defined and is an involution, only assuming convexity.

Example 9.1. If we remove strict convexity and coercivity, we can get functions which are not defined everywhere. For example, take

$$\begin{cases} L(0) = 0 \\ L(q) = \infty \quad q \neq 0. \end{cases}$$

What is H in this case?

We have not incorporated the initial data of the Hamilton-Jacobi equations into our calculus of variations. We will do this by adding $u_0(x_0)$ to the minimization problem (so when $T = 0$, we get $u_0(x_0)$) and removing the condition $x(0) = x_0$ from our set $\mathcal{A}$. So we are minimizing

$$\min_{x \in \mathcal{A}} \int_0^T L(x, \dot{x})\, dt + u_0(x_0) = u(T, x_T),$$

with $\mathcal{A} = \{x : [0, T] \to \mathbb{R} \text{ Lipschitz} \mid x(T) = x_T\}$.

9.2 Existence of minimizers for the Euler-Lagrange equation

We want to prove the following:

Theorem 9.1. *The minimal value function $u(T, x_T)$ in the calculus of variations is the solution to the Hamilton-Jacobi equations.*

First, we should ask: Does a minimum solution to the Euler-Lagrange equation exist? The answer is yes, as long as L is convex, coercive, and Lipschitz in x and if $u_0 \in \text{Lip}$. However, there is no guarantee of uniqueness. We will not prove this, but here is some intuition:

Here is the trivial case:

Proposition 9.1. *Suppose we have a continuous function $F : K \to \mathbb{R}$ with K compact. Then* $\min F$ *is attained.*

Proof. Let x_n be a minimizing sequence: $F(x_n) \to \inf F$. Then $x_n \to x_0$ along a subsequence. Then $F(x_n) \to F(x_0)$, so x_0 is the minimizer. □

What if we try to apply this to calculus of variations? Suppose we have a minimizing sequence $x_n : [0, T] \to \mathbb{R}^n$. Then $\mathcal{L}(x_n) \to u(T, x_T)$ but in what topology? Is x_n in a bounded set? We know that $\mathcal{L}(x_n)$ is bounded. If $L(x, q) = q^2$, for example, we could conclude that $\int_0^T (\dot{x}_n)^2 \leq c$. Then $\dot{x}_n$ is bounded in $L^2([0, T])$. This would imply that x_n is bounded in $C^{1/2}$ using Hölder's inequality: $(|x_n(t) - x_n(s)| \leq c|t - s|^{1/2})$. This implies

that x_n is equicontinuous (and equibounded by the $x(T) = x_T$ assumption). So the Arzelà Ascoli theorem says that $x_n \to x$ uniformly. Then

$$\lim_{n\to\infty} \mathcal{L}(x_n) = \lim_{n\to\infty} \int_0^T L(x_n, \dot{x}_n)\, dt + \underbrace{u(x_{n,0})}_{\to u_0(x_0)}$$

We can pass to the limit without a problem for x, but convergence with respect to $\dot{x}$ is trouble.

The limit of the integral may not exist, but maybe we can hope for

$$\int_0^T L(x, \dot{x})\, dt \le \liminf_{n\to\infty} \int_0^T L(x_n, \dot{x}_n)\, dt.$$

This is lower semicontinuity for the map $x \mapsto \mathcal{L}(x)$. The key observation is that convexity of $\mathcal{L}$ implies lower semicontinuity of $\mathcal{L}$:

Proof. The convexity inequality tells us that

$$L(\dot{x}_n) \ge L(\dot{x}) + L_q(\dot{x})(\dot{x}_n - \dot{x}).$$

Integrating gives us

$$\int_0^T L(\dot{x}_n)\, dt \ge \int_0^T L(\dot{x})\, dt + \int_0^T L_q(\dot{x})(\dot{x} - \dot{x}_n)\, dt$$

We are done if $\lim_{n\to\infty} \int L_q(\dot{x})(\dot{x}_n - \dot{x}) = 0$. We have replaced our nonlinear dependence on $\dot{x}_n - \dot{x}$ by a linear property.

Since $\dot{x} \in L^2$, we can approximate $L_q(\dot{x})$ by smooth functions. Suppose $y_k \in C^\infty$ with $y_k \to L(\dot{x})$ in L^2. It is enough to see that

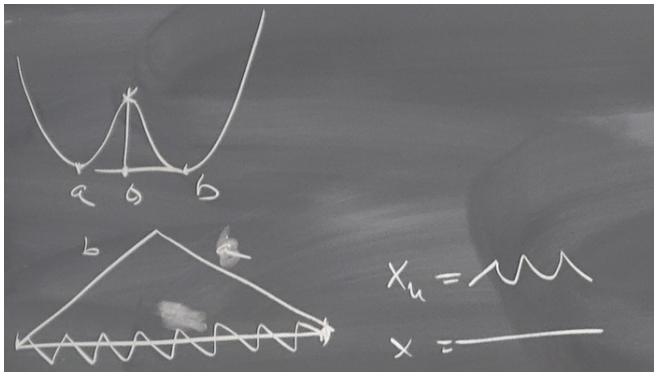
$$\lim_{n\to\infty} \int_0^T y_k(\dot{x}_n - \dot{x})\, dt = 0$$

In this context, we can integrate by parts. The integral equals

$$\int_0^T y_k(\dot{x}_n - x)\, dt = \int_0^T \dot{y}_k(x_n - x)\, dt + y_k(x_n - x)|_0^T \xrightarrow{n\to\infty} 0$$

by uniform convergence of $x_n \to x$. □

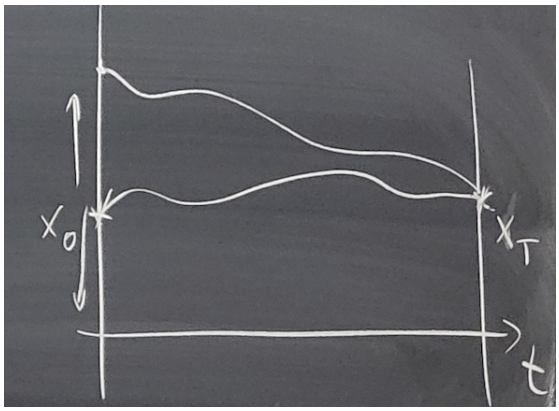
Example 9.2. Recall our double well potential.



In this case, if x_n is a wiggle approximating the 0 trajectory, we have $L(\dot{x}_n) = 0$ by $L(\dot{x}) = L(0) > 0$.

Remark 9.2. The Hamilton-Jacobi equation can be solved for a short time using characteristics. In calculus of variations, the analogue turns out to be that minimizers are unique for a short time.

We want to think of two minimizers in calculus of variations as characteristics that intersect.

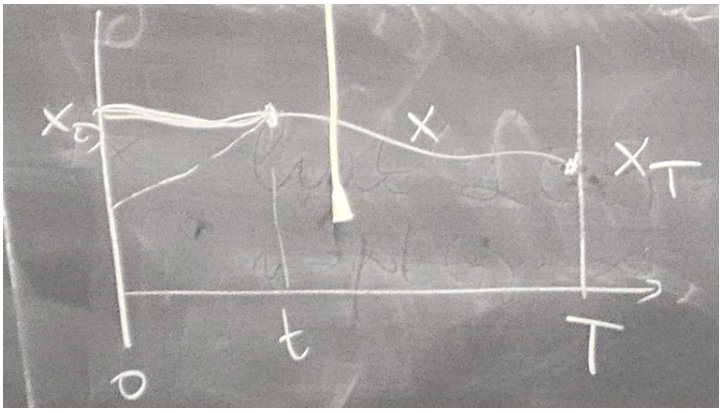


9.3 Proving that Euler-Lagrange equation minimizers solve Hamilton-Jacobi equations

Here is the "proof" of our theorem.

Proof. Suppose x is a minimizer for the action functional. We can choose a intermediate

point t, and first minimize relative to the time t.



$$\min_x \int_0^T L(x,\dot{x})\,dt + u_0(x_0) = \min_x \int_0^t L(x,\dot{x})\,ds + u_0(x_0) + \int_t^T L(x,\dot{x})\,ds$$

If $x|_{[0,T]}$ is a minimizer, then $x|_{[0,t]}$ is also a minimizer. So

$$u(x_T,x_0) = \min u(x_t,x_0) + \int_t^T L(x,\dot{x})\,ds.$$

This is called the **dynamic programming principle**.[7] This principle tells us that for minimizers,

$$u(x_T,x_0) = u(x_t,x_0) + \int_t^T L(x,\dot{x})\,ds,$$

which we can differentiate with respect to t to get

$$\begin{aligned}\frac{d}{dt}u(x_t,x_0) &= L(x,\dot{x})\\ &= p\cdot q - H(x,p)\\ &= p\cdot H_p - H.\end{aligned}$$

We conclude that $u(t,x_t)$ from the calculus of variations is the same as the $u(t,x_t)$ from the Hamilton-Jacobi equation because they solve the same equation with the same initial data at time 0. □

Remark 9.3. This is not an entirely correct proof. How do we know that there is an optimal trajectory starting at x_0? If the time is short enough, we can guarantee a minimizer starting at x_0, but this is exactly the issue of uniqueness of minimizers. This proof can be made rigorous for short times.

[7] This is discussed near the end of Evans' book.

Remark 9.4. More generally, this is related to **control theory**, where we try to find

$$u(x_0, T) = \min \int_0^T L(x, u)\, dt + u_0(x(0)), \qquad \dot{x} = h(x, f)$$

Here, we can choose some weight of influence by changing f, and we are trying to optimize some cost functional. The function $u(x_0, T)$ solves a Hamilton-Jacobi equation.

We can think of our calculus of variations problem as the case where the ODE for x is given by $\dot{x} = f$.

Remark 9.5. Calculus of variations allows us to obtain meaningful solutions for Hamilton-Jacobi equations after characteristics begin to intersect. Instead of picking which characteristic to continue, we can just look for a minimizer for a calculus of variations problem in longer time.

10 The Hopf-Lax Solution to Hamilton-Jacobi Equations

10.1 The Hamiltonian in classical mechanics

Last time, we were solving the Hamilton-Jacobi equation

$$\begin{cases} u_t + H(x, Du) = 0 \\ u(0) = u_0 \end{cases}$$

using the calculus of variations:

$$u(x,t) = \inf_{y(t)=x} \int_0^t L(y(s), \dot{y}(s))\, ds + u_0(y(0)).$$

Theorem 10.1. *The function u solves the Hamilton-Jacobi equation for as long as the solutions stay smooth.*

In the proof, we had the convex duality

$$H(x,p) = \max_q p \cdot q - L(x,q)$$

for the Hamiltonian $H(x,p)$ and the Lagrangian $L(x,q)$.

Example 10.1. Here is an example from classical mechanics. Consider the Lagrangian

$$L(x,q) = \frac{1}{2}mq^2 - \phi(x),$$

where $\frac{1}{2}mq^2$ is kinetic energy and $\phi(x)$ is potential energy. Then

$$H(x,p) = \sup_q p \cdot q - \frac{1}{2}mq^2 + \phi(x)$$

Complete the square to get

$$\begin{aligned} &= \sup_q \frac{1}{2m}p^2 - \frac{1}{2m}(p - mq)^2 + \phi(x) \\ &= \frac{1}{2m}p^2 + \phi(x) \end{aligned}$$

In the physical interpretation, the Hamiltonian $H(x,p)$ plays the role of the energy of the system.

10.2 The Hopf-Lax formula

Now we will consider a special case, where $L = L(q)$ does not depend on x (and consequently $H = H(p)$). Assume that L, H are strictly convex and coercive (i.e. $\lim_{q\to\infty} \frac{L(q)}{|q|} = \infty$). The Euler-Lagrange equation tells us that

$$\cancel{L_x(y,\dot{y})} + \frac{d}{dt} L_q(y, \dot{y}) = 0.$$

So we get that $L_q(\dot{y})$ is constant. Since L_q is a local diffeomorphism, we get that $\dot{y}$ is constant. That is, the solutions to the Euler-Lagrange equation are linear.

We claim that fixing the endpoints $y(0), y(t)$, the minimum is attained for linear trajectories.

Theorem 10.2 (Hopf-Lax formula[8])**.** *If $L = L(q)$ is convex, then*

$$u(x,t) = \inf_y u_0(y) + tL\left(\frac{x-y}{t}\right).$$

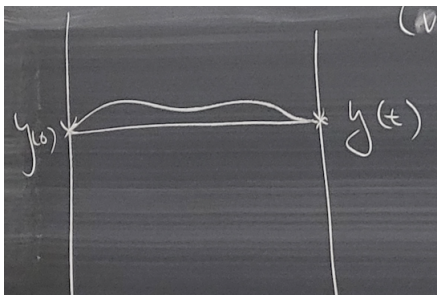
Proof. Since

$$\int_0^t \dot{y}(s)\, ds = y(t) - y(0),$$

we can average to get

$$\frac{1}{t}\int_0^t \dot{y}(s)\, ds = \frac{y(t) - y(0)}{t},$$

where the right hand side is the average velocity for a straight path.



Then

$$\int_0^t L(\dot{y}(s))\, ds = t \cdot \frac{1}{t}\int_0^t L(\dot{y}(s))\, ds$$

[8]This is from the 50s or the 60s. Professor Tataru was actually able to meet Lax a few times.

Convexity says that $L(\frac{x+y}{2}) \leq \frac{1}{2}(L(x) + L(y))$. More generally, we get that $L(hx + (1-h)y) \leq hL(x) + (1-h)L(y)$. If we use n variables, this is $L(\frac{x_1+\cdots+x_n}{n}) \leq \frac{1}{n}(L(x_1) + \cdots + L(x_n))$. If we increase the number of variables, this says that $L(\text{avg}(z)) \leq \text{avg}(L(z(s)))$, where we are taking average integrals. This is called **Jensen's inequality**, and it gives us

$$\geq t \cdot L\left(\frac{y(t) - y(0)}{t}\right)$$

In other words, the cost of an arbitrary path is $\geq$ the cost of the straight path. □

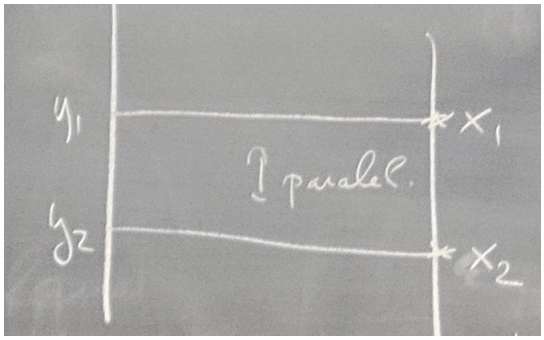
We are not done yet. We still need to minimize $u_0(y(0))$ over the choice of $y(0)$.

10.3 Properties of the Hopf-Lax solution

Assume L is convex and coercive. For simplicity, also assume that u_0 is bounded. Observe that if $t > 0$, then we can restrict $q = \frac{x-y}{t}$ to a compact set. So if u_0 is also continuous, then the infimum is attained.

Proposition 10.1. *If $u_0 \in$ Lip, then $u \in$ Lip.*

Proof. Here is a proof by picture. Suppose we have points x_1, x_2, and we want to compare $u(x_1)$ and $u(x_2)$. It is enough to consider parallel trajectories with y_1, y_2.



Take $x_1 - y_1 = x_2 - y_2$. Then $y_1 - y_2 = x_1 - x_2$. We have

$$u(x_1, t) = \inf_{y_1} u_0(y_1) + tL\left(\frac{x_1 - y_1}{t}\right),$$

$$u(x_2, t) = \inf_{y_2} u_0(y_2) + tL\left(\frac{x_2 - y_2}{t}\right).$$

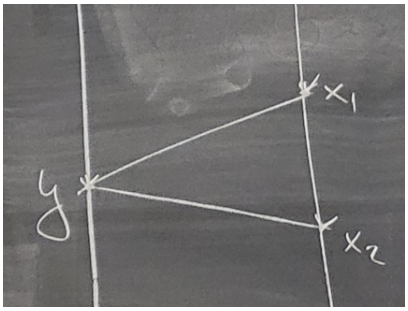
Using the Lipschitz condition, $|u_0(y_1) - u_0(y_2)| \leq L|y_1 - y_2| = L|x_1 - x_2|$. So the conclusion is that

$$|u(x_1, t) - u(x_2, t)| \leq L|x_1 - x_2|.$$ □

What if we don't assume u is Lipschitz? Can we still conclude that u is Lipschitz?

Proposition 10.2. *If u_0 is continuous, then $u(t)$ is Lipschitz.*

Proof. In this case, compare x_1 and x_2 to the same y:



We have

$$u(x_1) = \inf_y u_0(y) + L\left(\frac{x_1 - y}{t}\right),$$

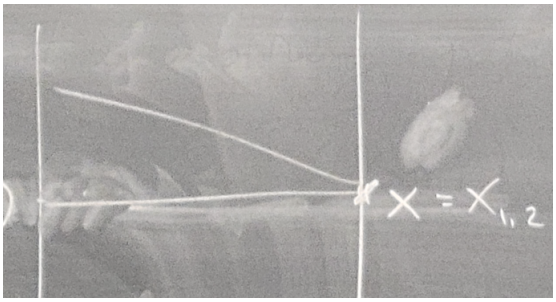
$$u(x_2) = \inf_y u_0(y) + L\left(\frac{x_2 - y}{t}\right).$$

The difference

$$\left| L\left(\frac{x_1 - y}{t}\right) - L\left(\frac{x_2 - y}{t}\right) \right| \leq C \cdot \frac{|x_1 - x_2|}{t},$$

where the Lipschitz constant $C = C(t)$ in the set where $\frac{x_1 - y}{t}$ and $\frac{x_2 - y}{t}$ live.

Where should we look? $\frac{y - x_1}{t}, \frac{y - x_2}{t}$ cannot be too large. Let $x = x_1 = x_2$, and compare the straight trajectory to an arbitrary trajectory.

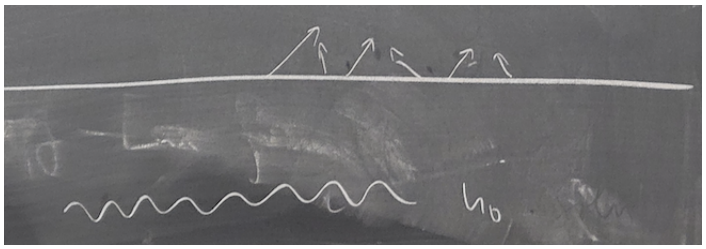


The oblique trajectory loses if $u_0(x) + tL(0) \leq u_0(y) + tL\left(\frac{x-y}{t}\right)$. This is when $\frac{2M}{t} \leq L(\frac{x-y}{t})$. So we can restrict to y such that $L(\frac{x-y}{t}) \leq \frac{2M}{t}$. So $\frac{x-y}{t}$ is in a compact set depending on t. Then the conclusion is that

$$|u(x_1, t) - u(x_2, t)| \leq C(t) \cdot \frac{|x_1 - x_2|}{t},$$

where $C(t)$ is the Lipschitz constant for L in the region $L(q) \leq \frac{C}{t}$. This Lipschitz constant goes to ∞ as $t \to 0$. □

In terms of the Hamilton-Jacobi equation, there will be lots of velocities with different speeds. So there is only an average velocity that survives.



We say that this PDE has a mild **regularizing effect**.

10.4 Almost everywhere solvability of the Hamilton-Jacobi equation

Recall the following theorem from real analysis (which requires measure theory).

Theorem 10.3. *If u is a Lipschitz function, then u is differentiable almost everywhere.*

So we get the following conclusion.

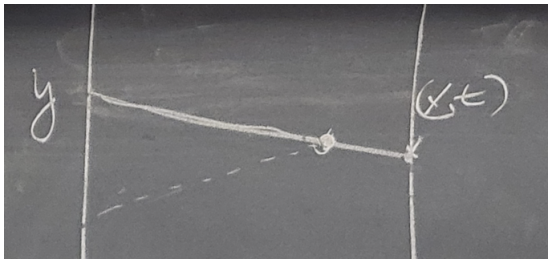
Corollary 10.1. *The solution u is differentiable almost everywhere.*

Proposition 10.3. *Let (x,t) be a differentiability point for u. Then the Hamilton-Jacobi equation holds at (x,t).*

Corollary 10.2. *The function u solves the Hamilton-Jacobi equation almost everywhere.*

Let's prove the proposition.

Proof. We can think of the Hamilton-Jacobi equation as proving two separate inequalities. If our trajectory is optimal, then it is optimal if we only look at the trajectory at a shorter length of time. Look at the optimal trajectory, ending at y and with slope $\frac{x-y}{t}$.



Then

$$u(x,t) = u_0(y) + tL\left(\frac{x-y}{t}\right),$$

so

$$u\left(x - h\frac{x-y}{t}, t-h\right) = u_0(y) + (t-h)L\left(\frac{x-y}{t}\right)$$

The first equation tells us that y is the optimal trajectory for (x,t), and the second says that y is optimal for $(x \cdot h\frac{x-y}{t}, t-h)$.

Let $q = \frac{x-y}{t}$. Then dividing by h gives

$$\frac{u(x,t) - u(x-hq, t-h)}{h} = hL(q).$$

Letting $h \to 0$ gives

$$\partial_x u \cdot q + \partial_t u = L(q).$$

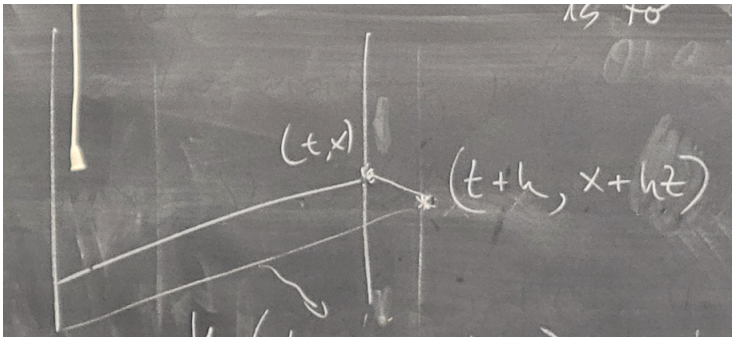
So for this special q we have chosen,

$$\partial_t u + \partial_x u \cdot q - L(q) = 0.$$

We want to think of this in terms of the Legendre transform. Since $H(p) = \sup p \cdot q - L(q)$, the latter half of our equation, $\partial_x u \cdot q - L(q)$, is $\leq H(\partial_x u)$. So we get

$$\partial_t u + H(\partial_x u) \geq 0.$$

Now we want to produce the other inequality. Notice that for the previous inequality, it was enough to work with a specific value of q, whereas for this direction, we will need to look at all values of q. Instead of looking at the past of (t,x), look at the future of (t,x). Our picture looks like



One trajectory from $(t+h, x+hz)$ is to go through x, but this may not be optimal. So

$$u(t+h, x+hz) \leq u(t,x) + \underbrace{hL(z)}_{=\int_t^{t+h} L(z)\,ds}$$

As before, subtract the right hand side, divide by h, and let $h \to 0$. Then we get

$$\frac{u(t+h, x+hz) - u(t,x)}{h} \leq L(t) \implies \partial u + \partial_x u z \leq L(z).$$

So we have proven that for all z,

$$\partial_t u + \partial_x u \cdot z \leq 0.$$

Taking the supremum over all z, we get

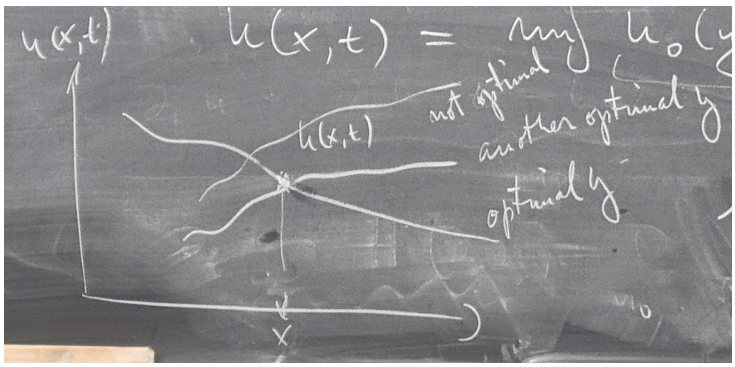
$$\partial_t u + H(\partial_x u) \leq 0. \qquad \square$$

Now we will tell a story. The details are in Evans' book, but the overall story is more important. We want to ask a question: Does solving the Hamilton-Jacobi equation almost everywhere suffice to guarantee uniqueness for Hamilton-Jacobi? Equivalently, does this guarantee that u is the minimal value function? The answer is no.

Are there other interesting properties for the function u? Look at the Hopf-Lax formula

$$u(x,t) = \inf u_0(y) + tL\left(\frac{x-y}{t}\right).$$

Observe that this is an infimum of functions which are smooth in x. We can compare what this looks like for different optimal/nonoptimal y:



Since we are taking a minimum, we can see that our curve could have a corner pointing upwards, but a corner pointing downwards is not possible. This points to a concavity property of our solution.

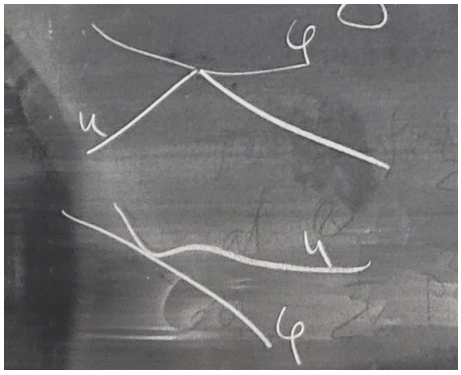
Proposition 10.4. *u is semiconcave.*

Concave means that $u(t,x) \geq \frac{u(t,x+y)+u(t,x-y)}{2}$. **Semiconcave** means that

$$u(t,x) \geq \frac{u(t,x+y)+u(t,x-y)}{2} - c \cdot |x-y|^2.$$

Theorem 10.4. *The optimal value function u is the unique semiconcave solution to the Hamilton-Jacobi equation.*

The proof is in Evans, but it is a little hard to follow. There is a better way to do things! Instead of plugging in u to check whether it satisfies the equation, if we have a corner, draw a tangent test function φ with $\varphi_t + H(\partial_x \phi) \geq 0$ or $\varphi_t + H(\partial_x \phi) \leq 0$.



These are called **viscosity solutions** for Hamilton-Jacobi equations.

11 Introduction to Distribution Theory

11.1 Weak solutions to PDEs

For the next month or so, our goal will be to study linear, constant coefficient PDEs

$$P(\partial)u = f, \qquad P(\partial) = \sum_{|\alpha| \le m} c_\alpha \partial^\alpha.$$

We will first take a detour to study the theory of distributions. First, some motivation:

Example 11.1. Recall the transport equation

$$\begin{cases} (\partial_t + A_j \partial_j)u = 0 \\ u(0) = u_0 \end{cases}$$

with constant coefficients A_j. The characteristics are given by $\dot{x} = A$, which gives $x(t) = x(0) + tA$. This means that $\dot{u} = 0$ along these characteristics, so $u(x(t), t) = u(x(0), 0)$. In other words,

$$u(x, t) = u_0(x - tA).$$

Classically, if $u_0 \in C^1$, then $u \in C^1$. What if $u_0 \in C$? It doesn't make sense to say that the solution u is continuous because we need to take derivatives. If we interpret the equation as a directional derivative, $u_0 \in C$ gives a solution. This interpretation relies strongly on the specific problem. Can we treat this problem in general?

Suppose we have a smooth function $\varphi \in C_0^\infty$. We can write the equation as the condition

$$\int_{\mathbb{R}^{n+1}} (\partial_t + A_j \partial_j) u \varphi \, dx = 0,$$

where a function is 0 if it integrates to be 0 against all $\varphi \in C_0^\infty$. Now integrate by parts to get

$$-\int_{\mathbb{R}^{n+1}} u(\partial_t + A_j \partial_j)\varphi \, dx = 0, \qquad \forall \varphi \in C_0^\infty,$$

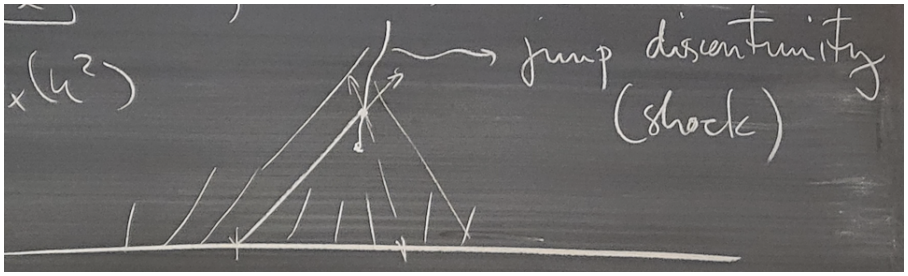
which applies to all $u \in C$. Our continuous solution will be a solution to this integral equation.

Definition 11.1. u is a **weak solution** to a PDE if the corresponding integral equation holds for all $\varphi \in C_0^\infty$.

Example 11.2. Recall the Burgers equation

$$u_t + uu_x = 0, \qquad u(0) = u_0.$$

The characteristics are given by $\dot{x} = u$ and $\dot{u} = 0$. The characteristics will intersect, and at the point of intersection of characteristics, the solution will start to develop a jump discontinuity, known as a **shock**.



In this problem, if we think of the equation as a directional derivative, the derivative along the characteristics are different when they intersect, so we cannot get a solution. However, we can similarly look for a weak solution by integrating by parts as before. When we do this, we want to think of uu_x as $\frac{1}{2}\partial_x(u^2)$.

11.2 Topologies on vector spaces

The key idea in the theory of distribution is that we can think of a function $u : \mathbb{R}^n \to \mathbb{R}$ as a linear map on all $\varphi \in C_0(\mathbb{R}^n)$ via

$$u(\varphi) := \int_{\mathbb{R}^n} u \cdot \varphi \, dx.$$

Observe that if $u(\varphi) = 0$ for all φ, then $u = 0$.

We will use the notation $\mathcal{D} = C_0^\infty$ to refer to the smooth functions with compact support. Obesrve that $\mathcal{D}$ is a linear space. What is the topology of $\mathcal{D}$? Recall that C is a normed space, with

$$\|u\|_C = \sup_{x \in \mathbb{R}^n} |u(x)|.$$

Recall:

Definition 11.2. A **normed space** is a vector space V, with a norm map $\|\cdot\| : V \to \mathbb{R}$ (or $\mathbb{C}$) satisfying

(a) $\|u\| \geq 0$, with equality iff $u = 0$.

(b) $\|\lambda u\| = |\lambda| \|u\|$ for all $\lambda \in \mathbb{R}$ (or $\mathbb{C}$).

(c) $\|u + v\| \leq \|u\| + \|v\|$.

We obtain a metric space structure, given by $d(u, v) = \|u - v\|$. Recall that **completeness** of a metric space means that every Cauchy sequence is convergent.

Definition 11.3. A **Banach space** is a complete normed space.

Here is a special class:

Definition 11.4. A **Hilbert space** is a vector space with a complete inner product $\langle u, v\rangle = u \cdot v$.

In a Hilbert space, we get a norm by

$$u \cdot u = \|u\|^2 \geq 0.$$

Example 11.3. The L^2 **space** is given by

$$L^2(\mathbb{R}^n) = \left\{ u : \mathbb{R}^n \to \mathbb{R} \mid \int |u|^2 \, dx < \infty \right\}.$$

This space is a Hilbert space, given the inner product

$$u \cdot v = \int_{\mathbb{R}^n} uv \, dx$$

(with v replaced by $\overline{v}$ in the complex case).

Hilbert spaces are a special case of Banach spaces, but a single space can have different norm structures on it.

Example 11.4. We can equip $\mathbb{R}^n$ with the norm $\|v\|^2 = \sum_j v_j^2$ which comes from the usual dot product (a Hilbert space structure). We can also equip $\mathbb{R}^n$ with the L^p norm $\|v\|^p = \sum_j |v_j|^p$ with $1 \leq p < \infty$, which gives a Banach space structure.

Example 11.5. C^k is a Banach space with the norm

$$\|u\| = \sup_{|\alpha| \leq K} \sup_{x \in \mathbb{R}^n} |\partial^\alpha u(x)|.$$

Returning to our objective, what norm can we give $C^\infty(\mathbb{R}^n)$? We can define

$$\|u\|_\alpha = p_\alpha(u) := \sup_{x \in \mathbb{R}^n} |\partial^\alpha u(x)|.$$

The problem is that we have infinitely many of these. What would $u_n \to u$ mean in $C^\infty(\mathbb{R}^n)$? We want to say that $\partial^\alpha u_n \to \partial^\alpha u$ uniformly for all α.

The solution is to use all the $\|\cdot\|_\alpha$ as **seminorms**, which satisfy all the norm conditions except for $\|u\| = 0 \implies u = 0$.

Definition 11.5. Locally convex spaces are vector spaces equipped with a family of seminorms. A complete, locally convex space is called a **Fréchet space**.

In a locally convex space,

$$p_\alpha(u) = 0 \quad \forall \alpha \implies u = 0.$$

Why is this called a "locally convex space"? The idea is that each seminorm gives you neighborhoods of points, which may not be nested in each other for different seminorms. But these are all convex neighborhoods, and we can intersect these neighborhoods to get more convex neighborhoods around every point.

The picture of our function spaces looks like

$$\text{Hilbert spaces} \subsetneq \text{Banach spaces} \subsetneq \text{Fréchet spaces}.$$

Example 11.6. We will use the notation $\mathcal{E} = \{u \in \mathbb{R}^n \to \mathbb{R} \mid u \text{ is smooth}\}$. Here, we ask for nothing at ∞. What does $u_n \to u$ mean in $\mathcal{E}$? We can define this as $\partial^\alpha u_n \to \partial^\alpha u$ uniformly on compact sets. For this space, we need to use the collection of seminorms

$$p_{\alpha,K}(u) = \sup_{x \in K} |\partial^\alpha u(x)|, \qquad \alpha \in \mathbb{N}^d, K \text{ compact}.$$

We don't need to check *all* compact sets; it suffices to take nested balls with radius going to ∞. With this topology, $\mathcal{E}$ is a locally convex space.

For $\mathcal{D}$, we have an issue: if we have a sequence of functions of compact support, the support may grow to not be compact in the limit. To solve this, there is a notion called the **inductive limit** of locally convex spaces, essentially cooked up only to describe $\mathcal{D}$. To make a long story short, we describe convergence in $\mathcal{D}$ as $u_n \to u$ in $\mathcal{D}$ if

(a) $\partial^\alpha u_n \to \partial^\alpha u$ uniformly.

(b) There is a compact set K such that $\operatorname{supp} u_n \subseteq K$.

Remark 11.1. If $u \in C(\mathbb{R}^n)$ and $\varphi \in \mathcal{D}$, the map $\varphi \mapsto u(\varphi) = \int u\varphi \, dx$ is continuous.

Definition 11.6. The space of **distributions**, denoted $\mathcal{D}'$ or $\mathcal{D}^*$ is the space of linear, continuous $f : \mathcal{D} \to \mathbb{R}$.

This seems to separate us from our original goal. If we have a function, we can get a distribution, but if we have a distribution, we can't always get a function back; instead, we get *generalized functions*.[9]

[9] The term "distribution" comes from the French school, whereas the term "generalized functions" comes from the Russian school.

11.3 Examples of distributions

Here are some examples of distributions.

Example 11.7. The **Dirac mass** at 0 is

$$\delta_0(\varphi) = \varphi(0).$$

Example 11.8. Another distribution is

$$\delta_0'(\varphi) = -\varphi'(0).$$

The reason for the minus sign will become apparent later on. In general, we can define

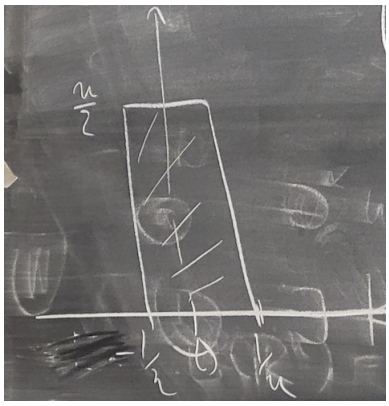
$$\delta_x^{(\alpha)}(\varphi) = (-1)^{|\alpha|}\partial^\alpha \varphi(x).$$

The space $\mathcal{D}'$ of distributions is a linear space.[10] It has the topology of **weak convergence**: $f_n \to f$ in $\mathcal{D}'$ if

$$f_n(\varphi) \to f(\varphi) \qquad \forall \varphi \in \mathcal{D}.$$

Example 11.9. Can we approximate δ_0 with functions? This may shed some light on what generalized functions look like. Let

$$u_n(x) = \begin{cases} n/2 & x \in [-1/n, 1/n] \\ 0 & \text{otherwise.} \end{cases}$$



Here, $\int u_n = 1$ for all n. If we try to take the limit in the sense of distributions, we get (in 1 dimension):

$$u_n(\varphi) = \int u_n \cdot \varphi \, dx$$

[10] You should think of the prime as a notion of duality of vector spaces.

$$= \frac{n}{2} \int_{-1/n}^{1/n} \varphi(x)\, dx$$
$$= \frac{1}{2} \int_{-1}^{1} \varphi(y/n)\, dx$$
$$\xrightarrow{n\to\infty} \varphi(0),$$

so $u_n(\varphi) \to \delta_0(\varphi)$. That is, $u_n \to \delta_0$.

Remark 11.2. In $\mathbb{R}^n$, we could use

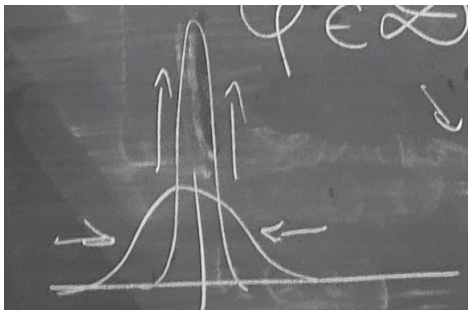
$$u_\varepsilon = \frac{1}{\varepsilon^n c_n} \mathbb{1}_{B(0,\varepsilon)}, \qquad c_n = |B(0,1)|.$$

In n dimensions, this has size $\sim 1/\varepsilon^n$.

Remark 11.3. We could also use $\mathcal{D}$ functions. If $\varphi \in \mathcal{D}$ with $\int \varphi = 1$, then we can define the rescaled function (at scale ε)

$$\varphi_\varepsilon(x) = \frac{1}{\varepsilon} \varphi(x/\varepsilon).$$

Here is the picture:



By the same argument, $\varphi_\varepsilon \to \delta_0$ in $\mathcal{D}'$.

Next time, we will see how we can think of distributions as solutions to PDEs. This will require knowing things like how to differentiate distributions.

12 Operations on Distributions and Homogeneous Distributions

12.1 Operations on distributions

Last time, we introduced distributions. We had the set $\mathcal{D} = C_0^\infty$ of **test functions** and the set $\mathcal{D}'$ of **distributions**, continuous linear maps $\mathcal{F} : \mathcal{D} \to \mathbb{R}$. If u is a function, we interpreted it as a distribution via

$$u(\phi) = \int u\phi \, dx.$$

So we can think of distributions as generalized functions. We also saw distributions as a limit of functions, in this weak sense.

Now, we want to see distributions as solutions to PDEs, so we need to think about operations with distributions.

12.1.1 Differentiation

We want to define $u \mapsto \partial_j u$ for distributions. First suppose u is a function. Then $\partial_j u$ is a function with

$$\begin{aligned}\partial_j u(\phi) &= \int \partial_j u\phi \, dx \\ &= -\int u \cdot \partial_j \phi \, dx \\ &= -u(\partial_j \phi).\end{aligned}$$

We can take this as a definition.

Definition 12.1. If $u \in \mathcal{D}'$, define $\partial_j u$ by $\partial_j u(\phi) = -u(\partial_j \phi)$.

Remark 12.1. If $u \in C^1$, then u is the same classically and as a distribution.

Example 12.1. Consider the Heaviside function

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0. \end{cases}$$

in 1 dimension. Then $\partial_x H = 0$ away from 0, in the classical sense. We can check that

$$\begin{aligned}\partial_x H(\phi) &= -H(\partial_x \phi) \\ &= -\int H(x)\partial_x \phi \, dx\end{aligned}$$

$$
\begin{aligned}
&= -\int_0^\infty \partial_x \phi(x)\, dx \\
&= -\phi|_0^\infty \\
&= \phi(0) \\
&= \delta_0(\phi),
\end{aligned}
$$

so $\partial_x H = \delta_0$ as a distribution. The idea is that when we have a jump discontinuity, differentiating gives us a Dirac mass.

Example 12.2. What is the derivative of the Dirac mass?

$$
\begin{aligned}
\partial_x \delta_0(\phi) &= -\delta_0(\partial_x \phi) \\
&= -\partial_x \phi(0) \\
&= \delta_0'(0).
\end{aligned}
$$

So the derivative of δ_0 is what we previously called δ_0'. Similarly, we can have $\partial^\alpha \delta_0 = \delta_0^{(\alpha)}$ for a multi-index α.

12.1.2 Multiplication by smooth functions

Suppose $\psi \in \mathcal{E}$ and u is a function. Then ψu is a function. What if $u \in \mathcal{D}'$? If u is a function, then

$$
\begin{aligned}
\psi u(\phi) &= \int \psi u \phi \, dx \\
&= \int u \underbrace{\psi \phi}_{\in \mathcal{D}} \, dx \\
&= u(\psi \phi).
\end{aligned}
$$

We can again take this as a definition.

Definition 12.2. If $u \in \mathcal{D}'$ and $\psi \in \mathcal{E}$, define ψu by $\psi u(\phi) = u(\psi \phi)$.

The Leibniz rule for derivatives says

$$
\partial(\psi u) = \partial \psi \cdot u + \psi \cdot \partial u.
$$

Using these definitions, this rule also holds for $u \in \mathcal{D}'$ and $\psi \in \mathcal{E}$.

If we have the equation $P(x, \partial) u = f$ with $P(x, \partial) = \sum c_\alpha(x) \partial^\alpha$, then all these operations are well-defined for distributions, so we can think of distribution solutions to PDEs.

12.2 The support of a distribution

Recall that if u is a function, its **support** is the largest closed set "where u is nonzero." In particular,

$$x_0 \notin \operatorname{supp} u \iff u = 0 \text{ in } B(x_0, r) \text{ for some } r > 0.$$

Definition 12.3. If $u \in \mathcal{D}'$, its **support** is the closed set defined by

$$x_0 \notin \operatorname{supp} u \iff u(\phi) = 0 \text{ for all } \phi \in \mathcal{D} \text{ with } \operatorname{supp} \phi \subseteq B(x_0, r)$$

Example 12.3. The support of the Dirac mass is $\operatorname{supp} \delta_0 = \{0\}$: If $x_0 \neq 0$, then there is a ball $B(x_0, r) \notin 0$. Then if we let $\phi \in \mathcal{D}$ have $\operatorname{supp} \phi \subseteq B(x_0, r)$, then $\delta_0(\phi) = \phi(0) = 0$.

Let $\mathcal{E}'$ denote the **compactly supported distributions**.

Proposition 12.1. *If $f \in \mathcal{E}'$, then f extends "naturally" to a continuous linear function on $\mathcal{E}$.*

Proof. We know $f(\phi)$ when $\phi \in \mathcal{D}$. Because $\operatorname{supp} f \subseteq B(0, R)$, $f(\phi) = 0$ if ϕ is supported outside $B(0, R)$. We can truncate ϕ outside B as follows: Replace ϕ by $\chi\phi$, where χ is a **cutoff function** with compact support, $\operatorname{supp} \chi \subseteq B(0, 2R)$, and $\chi = 1$ in $B(0, R)$. Then

$$\begin{aligned} f(\phi) &= f(\chi\phi) + f((1 - \chi)\phi) \\ &= f(\chi\phi). \end{aligned}$$

So for $\phi \in \mathcal{E}$, define $f(\phi) := f(\chi\phi)$. □

We have the following picture:

$$\begin{array}{ccc} \mathcal{D} & \xrightarrow{\text{dual}} & \mathcal{D}' \\ {\scriptstyle\subseteq}\downarrow & & \uparrow{\scriptstyle\subseteq} \\ \mathcal{E} & \xrightarrow[\text{dual}]{} & \mathcal{E}' \end{array}$$

We will extend this picture later when we learn about the Fourier transform.

12.3 Homogeneous distributions

Example 12.4. The polynomial $f(x) = x^n$ is a homogeneous polynomial. We can express this homogeneity by

$$f(\lambda x) = \lambda^n f(x),$$

where n is the homogeneity index.

Example 12.5. The homogeneity index does not have to be an integer. If we have $f(x) = |x|^\alpha$, then

$$f(\lambda x) = \lambda^\alpha f(x)$$

for $\lambda > 0$. If α is not an integer, this is not smooth at 0. Is $|x|^\alpha$ a distribution? This is related to the question of whether $|x|^\alpha$ is integrable (away from infinity). In 1 dimension, $\int |x|^\alpha \, dx$ exists if $\alpha > -1$. In n dimensions, we can use polar coordinates:

$$\int |x|^\alpha \, dx = c_n \int r^\alpha r^{n-1} \, dr,$$

where c_n is the volume of the unit ball in n-dimensions. Here, we need $\alpha + n - 1 > -1$, i.e. $\alpha > -n$. So $\frac{1}{|x|^n}$ is borderline.

Example 12.6. The Heaviside function is homogeneous of index 0:

$$H(\lambda x) = \lambda^0 H(x)$$

for $\lambda > 0$.

Example 12.7. In 2 dimensions (expressed in polar coordinates (r, θ)), the function

$$f(x) = r^\alpha g(\theta)$$

is homogeneous of index α.

For functions, the homogeneity condition $f(\lambda x) = \lambda^\alpha f(x)$ has a distributional interpretation:

$$\int f(\lambda x)\phi(x) \, dx = \lambda^\alpha \int f(x)\phi(x) \, dx$$

Applying a change of variables on the left,

$$\int f(y)\phi(y/\lambda)\frac{1}{\lambda^n} \, dy = \lambda^\alpha \int f(x)\phi(x) \, dx.$$

Denoting $\phi_\lambda(x) = \lambda^{-n}\phi(x/\lambda)$, we get the relation

$$f(\phi_\lambda) = \lambda^\alpha f(\phi),$$

which is meaningful for distributions.

Definition 12.4. A distribution $f \in \mathcal{D}'$ is **homogeneous of order** α if

$$f(\phi_\lambda) = \lambda^\alpha f(\phi)$$

for $\phi \in \mathcal{D}$.

Example 12.8. Can we think of the Dirac mass δ_0 as a homogeneous distribution?

$$\delta_0(\phi_\lambda) = \phi_\lambda(0) = \lambda^{-n}\phi(0) = \lambda^{-n}\delta_0(\phi),$$

so δ_0 has homogeneity $-n$.

In calculus, we have $\partial_x x^n = nx^{n-1}$. That is, we differentiate something which is homogeneous of order n and get something which is homogeneous of order $n-1$.

Proposition 12.2. *If $f \in \mathcal{D}'$ is homogeneous of order α, then $\partial_x f$ is homogeneous of order $\alpha - 1$.*

Proof. The chain rule works for functions, so it also works using the definition for distributions by passing the derivative to the test function. □

Example 12.9. The Heaviside function is homogeneous of order 0, and $\partial_x H = \delta_0$ is homogeneous of order -1. Similarly, $\partial_x \delta_0 = \delta_0'$ is homogeneous of order -1.

In 1 dimension, we want to classify homogeneous distributions. Start with functions and $\alpha > -1$. We need to assign $f(-1)$ and $f(1)$, so this is a linear space of dimension 2. Here is a basis:

$$x_+^\alpha = \begin{cases} 0 & x < 0 \\ x^\alpha & x > 0, \end{cases} \qquad x_-^\alpha = \begin{cases} |x|^\alpha & x < 0 \\ 0 & x > 0. \end{cases}$$

Then $|x|^\alpha = x_+^\alpha + x_-^\alpha$, and

$$\partial_x x_+^\alpha = \alpha x_+^{\alpha-1}, \qquad \partial_x x_-^\alpha = -\alpha x_-^{\alpha-1}.$$

Now look at when $\alpha \in (-2, -1)$. We can define

$$\partial_x x_+^{\alpha+1} := (\alpha+1)x_+^\alpha.$$

If we repeat this, we can get homogeneous distributions to all noninteger negative αs.

What about $\alpha = -1$? We have δ_0. At order 0, we have 2 homogeneous distributions: H and the constant 1 function. But differentiating these gives δ_0 and 0, which do not have a 2 dimensional span. Other candidates are $\frac{1}{|x|}$ or $\frac{1}{x}$. We can look at the integrals

$$\int \frac{1}{|x|}\phi(x)\,dx \qquad \int \frac{1}{x}\phi(x)\,dx.$$

On the left, there may be no cancelation at 0, but we may be able to get some cancelation at 0 for the right integral. We may try to define

$$\int_{\mathbb{R}} \frac{1}{x}\phi(x)\,dx := \lim_{\varepsilon \to 0} \int_{R\setminus[-\varepsilon,\varepsilon]} \frac{1}{x}\phi(x)\,dx.$$

Does this limit exist? We can look at

$$\int_{[-1,1]\setminus[-\varepsilon,\varepsilon]} \frac{1}{x}\phi(x)\,dx = \int_{-1}^{-\varepsilon} \frac{1}{x}\phi(x)\,dx + \int_{\varepsilon}^{1} \frac{1}{x}\phi(x)\,dx$$

Use the change of variables $y = -x$ on the left integral to get

$$= \int_{\varepsilon}^{1} \frac{\phi(x) - \phi(-x)}{x}\,dx.$$

$\phi(x) - \phi(-x)$ is $o(x)$, so this converges.

Thus, we can define the **principal value** $\mathrm{PV}\,\frac{1}{x}$ by

$$\mathrm{PV}\,\frac{1}{x}(\phi) = \lim_{\varepsilon\to 0} \int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} \frac{\phi(x)}{x}\,dx,$$

which is homogeneous of order -1.

13 Homogeneous Distributions of Order -1, Convolution, and Fundamental Solutions

13.1 Special homogeneous distributions of order -1

13.1.1 The principal value of $1/x$ as a complex limit

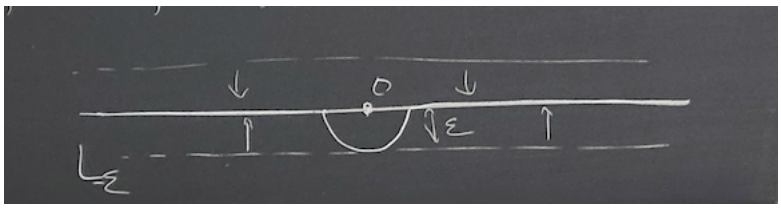
Last time, we were discussing homogeneous distributions. When classifying homogeneous distributions of order -1 in 1 dimension, we saw two interesting distributions:

$$\delta_0, \qquad \mathrm{PV}\,\frac{1}{x}.$$

If you like complex analysis, you can consider the function

$$f(z) = \frac{1}{z} = \frac{1}{x+iy}.$$

Then $f(z) = \frac{1}{x-i\varepsilon}$ on the line $L_{-\varepsilon}$ below the real line:



What is $\lim_{\varepsilon\to 0} \frac{1}{x-i\varepsilon}$? Apply this to a test function:

$$\begin{aligned}
\frac{1}{x-i\varepsilon}(\varphi) &= \int \frac{\varphi(x)}{x-i\varepsilon}\,dx \\
&\approx \int_{\mathbb{R}\setminus[\varepsilon,\varepsilon)} \frac{\varphi(x)}{x-i\varepsilon} + \int_{\frac{1}{2}C_\varepsilon} \frac{\varphi(z)}{z}\,dz \\
&\approx \mathrm{PV}\,\frac{1}{x}(\varphi) + \varphi(0)\cdot\int_{\frac{1}{2}C_\varepsilon}\frac{1}{z}\,dz
\end{aligned}$$

Write $\ln z = \ln|z| + i\arg z$. Then $z = \varepsilon e^{i\theta}$ for $\theta \in [\pi, 2\pi]$

$$\begin{aligned}
&= \mathrm{PV}\,\frac{1}{x}(\varphi) + \varphi(0)\cdot\int_{\pi}^{2\pi}\frac{i\varepsilon e^{i\theta}}{\varepsilon e^{i\theta}}\,d\theta \\
&= \mathrm{PV}\,\frac{1}{x}(\varphi) + \varphi(0)\pi i.
\end{aligned}$$

So

$$\lim_{\varepsilon\to 0}\frac{1}{x-i\varepsilon} = \mathrm{PV}\,\frac{1}{x} + \pi i\delta_0.$$

If we do the same approximation from the line L_ε above the real line, we get

$$\lim_{\varepsilon\to 0}\frac{1}{x+i\varepsilon} = \operatorname{PV}\frac{1}{x} - \pi i\delta_0.$$

What is $\partial_x \operatorname{PV}\frac{1}{x}$? We can calculate that

$$-\lim_{\varepsilon\to 0}\frac{1}{(x-i\varepsilon)} = \left(\operatorname{PV}\frac{1}{x}\right)' + \pi i\delta_0',$$

and repeat this idea to find the derivatives of $\operatorname{PV}\frac{1}{x}$.

13.1.2 $1/|x|$ as a distribution

What is $\frac{1}{|x|}$ as a distribution?

$$\begin{aligned}\lim_{\varepsilon\to 0}\int_{[-1,1]\setminus[-\varepsilon,\varepsilon)}\frac{1}{|x|}\varphi(x)\,dx &= \int\frac{1}{|x|}(\varphi(x)-\varphi(0))\,dx + \varphi(0)\int\frac{1}{|x|}\,dx\\ &\to \int_{-1}^{1}\frac{1}{|x|}(\varphi(x)-\varphi(0))\,dx + 2\varphi(0)|\log\varepsilon|.\end{aligned}$$

But this does not converge as $\varepsilon\to 0$. So we can try to **renormalize**, calculating the integral when we subtract out the divergent term:

$$\frac{1}{|x|}(\varphi) := \lim_{\varepsilon\to 0}\int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]}\frac{1}{|x|}(\varphi(x)-\varphi(0))\,dx - 2\varphi(0)|\log\varepsilon|$$

However, this breaks the homogeneity.

13.2 Properties of convolution

Definition 13.1. Let $\varphi,\psi\in\mathcal{D}$. The **convolution** is the function

$$(\varphi*\psi)(x) = \int\varphi(y)\psi(x-y)\,dy.$$

Observe that this is smooth in x. What about the support?

Proposition 13.1.

$$\operatorname{supp}\varphi*\psi\subseteq\operatorname{supp}\varphi+\operatorname{supp}\psi$$

Proof. If we want to know the support, call $K=\operatorname{supp}\varphi$ and $K_1=\operatorname{supp}\psi$. If $(\varphi*\psi)(x)\neq 0$, then we must have $x\in K+K_1$. □

So we can think about convolution as a function

$$* : \mathcal{D} \times \mathcal{D} \to \mathcal{D}.$$

Proposition 13.2 (commutativity of convolution)**.**

$$\varphi * \psi = \psi * \varphi.$$

Proof. Make the change of variables $z = x - y$ in the integral. □

Proposition 13.3 (associativity of convolution)**.**

$$\varphi * (\psi * \zeta) = (\varphi * \psi) * \zeta.$$

So $(\mathcal{D}, +, *)$ is a commutative algebra. We have another commutative algebra structure on $\mathcal{D}$, $(\mathcal{D}, +, \cdot)$. We will later see that these structures are not unrelated; they are mirror images of each other.

With multiplication, we have the Leibniz rule:

$$\partial(\psi\varphi) = \partial\psi \cdot \varphi + \psi \cdot \partial\varphi.$$

We don't exactly have a Leibniz rule for convolution:

Proposition 13.4.

$$\partial(\psi * \varphi) = \psi * \partial\varphi = \varphi * \partial\psi.$$

Proposition 13.5. *If $\varphi \in L^1$ and $\psi \in L^\infty$, then*

$$\|\varphi * \psi\|_{L^\infty} \leq \|\varphi\|_{L^1} \|\psi\|_{L^\infty}.$$

Proof.

$$\begin{aligned} |(\varphi * \psi)(x)| &\leq \int |\varphi| \cdot \sup |\psi| \\ &= \|\varphi\|_{L^1} \|\psi\|_{L^\infty}. \end{aligned}$$

□

When you think of convolution, you want to think of two things: regularity and support. If $\varphi \in \mathcal{D}$ and $\psi \in \mathcal{E}$, then we lose information about the support, so $\varphi * \psi \in \mathcal{E}$. So $\mathcal{D} * \mathcal{E} \to \mathcal{E}$. On the other hand, if we take a derivative of the convolution, we just need to be able to take a derivative of one of the factors. Here is the takeaway:

- For the support of the convolution, we need the support of both factors.
- For regularity, we need the regularity of just one factor!

We can think of convolutions as distributions: If $\varphi \in \mathcal{E}$ and $\psi \in \mathcal{D}$,

$$\varphi * \psi(x) = \varphi(\psi(x - \cdot)).$$

This right hand side is well-defined even if $\varphi \in \mathcal{D}'$. So we see that

$$\mathcal{D}' * \mathcal{D} \to \mathcal{E}.$$

Similarly, we have

$$\mathcal{E}' * \mathcal{D} \to \mathcal{D}.$$

What about $\mathcal{E}' * \mathcal{E}'$? If $u, v, \varphi \in \mathcal{D}$, then

$$(u * v)(\varphi) = \iint u(y)v(x - y)\, dy\varphi(x)\, dx$$

Change variables using $z = x - y$ so $\varphi(x) = \varphi(z + y)$.

$$\begin{aligned} &= \iint u(y)v(z)\varphi(z + y)\, dy\, dz \\ &= \int u(y) \underbrace{\int v(z)\varphi(z + y)\, dz}_{v(\varphi(y+\cdot))}\, dy \\ &= u(v(\varphi(y + \cdot))). \end{aligned}$$

This conclusion makes sense even if $u, v \in \mathcal{E}'$. We can make this precise if we can approximate elements of $\mathcal{E}'$ by elements in $\mathcal{E}$. So we get

$$\mathcal{E}' * \mathcal{E}' \to \mathcal{E}'.$$

However, $\mathcal{D}' * \mathcal{D}'$ is undefined.

13.3 Fundamental solutions to PDEs

Now suppose we have the PDE

$$P(\partial)u = f,$$

where P is linear with constant coefficients and f is a distribution. The simplest f we can consider is δ_0, which gives us the equation

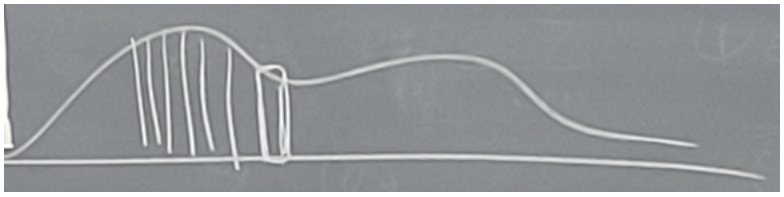
$$P(\partial)K = \delta_0$$

The next simplest f we can consider is δ_{x_0}. So we get

$$P(\partial)K(\cdot - x_0) = \delta_{x_0}$$

by invariance with respect to translations.

Can we write a general function as a superposition of δ functions? If we have a Riemann integral, we can approximate it by a sum of pieces which look like Dirac masses.



So can we make sense of something that looks like

$$f = \int f(x_0)\delta_{x_0}\, dx_0?$$

We can define this by applying f to a test function:

$$\varphi(\varphi) = \int f(x_0) \underbrace{\delta_{x_0}(\varphi)}_{=\varphi(x_0)}\, dx_0.$$

So if we can deal with a Dirac masses, we can deal with a lienar combination of Dirac masses and hence any function as a superposition of Dirac masses. So the solution should looks like

$$u(x) = \int f(x_0)K(x - x_0)\, dx_0.$$

This was some intuition[11], but here are some definitions.

Definition 13.2. K is a **fundamental solution** of $P(\partial)$ if

$$P(\partial)K = \delta_0.$$

Proposition 13.6. *The function $u = K * f$ solves the equation*

$$P(\partial)u = f.$$

Proof.

$$\begin{aligned} P(\partial)u &= P(\partial)(K * f) \\ &= P(\partial K) * f \\ &= \delta_0 * f. \end{aligned}$$

We are done if $f * \delta_0 = f$. If $f \in \mathcal{D}$, then

$$f * \delta_0(x) = \delta_0(f(x - \cdot)) = f(x).$$

The same works for $f \in \mathcal{D}'$. □

[11] Or maybe confusion!

In this proof, we saw that δ_0 is the identity with respect to $*$. For multiplication, $\mathbf{1}$ is the identity. The constant $\mathbf{1}$ function has support on all of $\mathbb{R}^n$, but it has regularity; conversely, δ_0 has 1 point as it support but no regularity. You can think of these as opposites.

Example 13.1. With our notation, the fundamental theorem of calculus looks like this:

Theorem 13.1. *If $\partial_x u = f$ in $\mathbb{R}$, then*

$$u = \int f(x)\,dx + C.$$

If we specify that $u(-\infty) = 0$, then

$$u(x) = \int_{-\infty}^{x} f(y)\,dy.$$

We want to interpret this as a convolution. First, let's compute the fundamental solution:

$$\partial_x K = \delta_0, \qquad K(-\infty) = 0.$$

This tells us that

$$K = H(x)$$

is the Heaviside function. By our proposition, $u = K * f$. We can write this as

$$u(x) = \int H(x-y) f(y)\,dy$$

For $H(x-y)$ to give 1 and not 0, we need $x - y > 0$.

$$= \int_{-\infty}^{x} f(y)\,dy.$$

Is the fundamental solution K unique? In general, if K is a constant solution, then $K + C$ is a fundamental solution for any constant C. If we ask for $K = 0$ at $-\infty$, we get $K = H$. But if we ask for $K = 0$ at $+\infty$, we get $K = H - 1$. If we ask for K to be odd, we get $K = H - 1/2$.

14 Fundamental Solutions for PDEs in 2 Dimensions and for the Laplacian

14.1 Fundamental solutions in 1 and 2 dimensions

Last time, we discussed fundamental solutions for partial differential equations. Suppose we have a differential operator in 1 dimension

$$P(\partial)K = \delta_0.$$

Solve the homogeneous equation and look for the fundamental solution

$$K(x) = \begin{cases} u_1^{\text{hom}}(x) & x < 0 \\ u_2^{\text{hom}}(x) & x > 0. \end{cases}$$

Plug this in into $P(\partial)K = \delta_0$ and get a linear system for the constants. As an exercise, try to solve the equation with the operator $P(\partial) = \partial^2 - 1$.[12]

What about in 2 dimensions? In complex analysis, one way to specify whether a function is holomorphic is via the Cauchy-Riemann equations. If our coordinates are (x, y), then let $z = x + iy$.

Definition 14.1. A function $f : \mathbb{R}^2 \to \mathbb{C}$ is **holomorphic** if

$$(\partial_x + i\partial_y)f = 0.$$

If we write $f = u + iv$, we can express this as equations for the real and imaginary parts:

$$\begin{cases} \partial_x u - \partial_y v = 0 \\ \partial_y u + \partial_x v = 0. \end{cases}$$

These are the **Cauchy-Riemann equations**. From the perspective of PDEs, this is just one equation.

Denote the operator

$$\overline{\partial} = \partial_x + i\partial_y.$$

Sometimes people will use this notation to denote $1/2$ this quantity. Complex differentiation is given by the operator

$$\partial = \partial_x - i\partial_y.$$

Our goal is to find the fundamental solution for $\overline{\partial}$.

Looking at $\overline{\partial}K = \delta_0$, notice that δ_0 is homogeneous of order -2 and $\overline{\partial}$ reduces order of homogeneity by 1. So we should look for a K which is homogeneous of order -1. Away

[12] Last week, this was a midterm question for Professor Tataru's undergraduate class.

from $z = 0$, $\overline{\partial} K = 0$, so K is holomorphic. So we should look for K of the form $K = \frac{c}{z}$, where c is a constant. This is locally integrable, unlike in 1 dimension. So we can define

$$K(\phi) = c \int_{\mathbb{R}^2} \frac{\phi(z)}{z}\, dx\, dy,$$

where we can use $dz\, d\overline{z}$ instead of $dx\, dy$. If K is a fundamental solution, $\overline{\partial} K = \delta_0$, so $\partial K(\phi) = \phi(0)$, which gives $K(-\overline{\partial}\phi) = \phi(0)$. Here,

$$\begin{aligned} K(-\overline{\partial}\phi) &= -c \iint_{\mathbb{R}^2} \frac{(\partial_x + i\partial_y)\phi(z)}{z}\, dx\, dy \\ &= \lim_{\varepsilon \to 0} -c \iint_{\mathbb{R}^2 \setminus B_\varepsilon} (\partial_x + i\partial_y)\phi \cdot \frac{1}{z}\, dx\, dy \end{aligned}$$

We want to use integration by parts. Using Green's theorem,

$$= \lim_{\varepsilon \to 0} c \iint_{\mathbb{R}^2 \setminus B_\varepsilon} \phi \cdot \underbrace{(\partial_x + i\partial_y)\frac{1}{z}}_{=0}\, dx\, dy - c \int_{\partial B_\varepsilon} (\nu_x + i\nu_y)\phi \cdot \frac{1}{z}\, ds,$$

where ν is the inner normal vector to the boundary of B_ε. In particular, $\nu = -\frac{(x,y)}{|z|}$.

$$\begin{aligned} &= \lim_{\varepsilon \to 0} c \int_{\partial B_\varepsilon} \frac{z}{|z|}\phi \cdot \frac{1}{z}\, ds \\ &= \lim_{\varepsilon \to 0} \frac{c}{\varepsilon} \int_{\partial B_\varepsilon} \phi(z)\, dz \\ &= 2\pi c \phi(0). \end{aligned}$$

We want $2\pi c = 1$, so we should pick $c = \frac{1}{2\pi}$. Thus, our fundamental solution is

$$K(z) = \frac{1}{2\pi z}.$$

Remark 14.1. We can rewrite this line integral in a complex fashion, as

$$\int \frac{\phi(z)}{z}\, dz = 2\pi i \phi(0),$$

by the residue theorem. So we have recovered the residue theorem. In essence, the residue theorem is the analogue of the fundamental theorem of calculus for 2 dimensions.

14.2 Fundamental solution for the Laplacian

Our next exercise is to find the fundamental solution to $P(\partial) = -\Delta$, where

$$\Delta = \partial_1^2 + \cdots + \partial_n^2.$$

Since δ_0 is homogeneous of order $-n$, and $P(\partial)$ will decrease the order of homogeneity by 2, K should be homogeneous of order $2-n$. To look for a candidate for a solution, we should look at the symmetries of Δ, in particular invariance with respect to rotations.

If $y = Ax$ is a linear change of variables, then $\frac{\partial}{\partial x_i} = A_{i,j}\frac{\partial}{\partial y_j}$. Then $\Delta = A_{i,j}A_{i,k}\frac{\partial}{\partial y_j}\frac{\partial}{\partial y_k}$. Here, we are using Einstein summation notation, in which the sum is implicit but unwritten. Do we get back Δ in y? The answer is yes, if

$$A_{i,j}A_{i,k} = I_n \iff A^\top A = I.$$

That is, we want A to be **orthogonal**. Recall that if A is orthogonal,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax\rangle \\ &= \langle x, A^\top Ax\rangle \\ &= \langle x, x\rangle \\ &= \|x\|^2.\end{aligned}$$

So we can look for K which is invariant with respect to rigid rotations, i.e. K is a spherically symmetric distribution.

Remark 14.2. We must be careful with this line of reasoning. We are just hoping that there exists some fundamental solution with this property. Not all fundamental solutions will have this property. For example, if we add x_1 to K, we will still have a fundamental solution, but it will not be radial.

We will guess

$$K = c_n \cdot \frac{1}{|x|^{n-2}},$$

where we will set the case $n = 2$ dimensions aside for now. Observe that

$$\begin{aligned}-\Delta K = \delta_0 &\iff -\Delta K(\phi) = \phi(0) \\ &\iff K(-\Delta\phi) = \phi(0) \\ &\iff \int_{\mathbb{R}^n} -\Delta\phi\frac{1}{|x|^{n-2}}\,dx = \phi(0).\end{aligned}$$

As before, write this integral as

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^n\setminus B_\varepsilon} -\Delta\phi\cdot\frac{1}{|x|^{n-2}}.$$

We want to integrate by parts. Here is Green's theorem in this setting:

Theorem 14.1 (Green's theorem for the Laplacian)**.**

$$\int_\Omega \Delta u\cdot v - u\cdot\Delta v\,dx = \int_{\partial\Omega}\frac{\partial u}{\partial\nu}v - u\frac{\partial v}{\partial\nu}\,d\sigma.$$

Proof.

$$\begin{aligned}\int_\Omega \Delta u \cdot v \, dx &= \int_\Omega \partial_j \partial_j u \cdot v \\ &= -\int_\Omega \partial_j u \cdot \partial_j v \, dx + \int_{\partial\Omega} \nu_j \partial_j u \cdot v \, d\sigma,\end{aligned}$$

where σ is surface measure on $\partial\Omega$. Observe (for the future) that $\partial_j u \cdot \partial_j v = \nabla u \cdot \nabla v$ in the first term and $\nu_j \partial_j u = \nu \cdot \nabla u := \frac{\partial u}{\partial \nu}$ is the **normal derivative** in the second term.

$$= \int_\Omega u \cdot \underbrace{\partial_j \partial_j}_{\Delta} v + \int_{\partial\Omega} \frac{\partial u}{\partial \nu} v - u \frac{\partial v}{\partial \nu} \, d\sigma. \qquad \square$$

Returning to our computation, we want

$$\phi(0) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^n \setminus B_\varepsilon} \phi \left(-\Delta \frac{1}{|x|^{n-2}} \right) dx - \int_{\partial B_\varepsilon} \frac{\partial \phi}{\partial \nu} \cdot \frac{1}{|x|^{n-2}} - \phi \frac{\partial}{\partial \nu} \frac{1}{|x|^{n-2}} \, d\sigma$$

The first integral goes away because $-\Delta \frac{1}{|x|^{n-2}} = 0$. We can see this via a formula for the Laplacian on radial functions: $\Delta F(r) = (\partial_r^2 + \frac{n-1}{r} \partial_r) F(r)$. This is the chain rule, switching to polar coordinates in n dimensions.

The second integral is

$$\int_{\partial B_\varepsilon} \frac{\partial \phi}{\partial \nu} \cdot \frac{1}{|x|^{n-2}} \, dA = O(\varepsilon) \to 0,$$

as $\frac{\partial \phi}{\partial \nu}$ is bounded, $\frac{1}{|x|^{n-2}} = \varepsilon^{2-n}$, and dA has order ε^{n-1}.

The third integral is

$$\begin{aligned}\int_{\partial B_\varepsilon} \phi \cdot \frac{\partial}{\partial \nu} \frac{1}{|x|^{n-2}} \, d\nu &= \int_{\partial B_\varepsilon} \phi \cdot (n-2) \frac{1}{|x|^{n-1}} \, d\sigma \\ &\approx \phi(0) \cdot \frac{n-2}{\varepsilon^{n-1}} \varepsilon^{n-1} a_n, \end{aligned}$$

where a_n is the area of the unit sphere.

$$= (n-2) a_n \phi(0).$$

So we need

$$c = c_n = \frac{1}{(n-2)a_n}.$$

Theorem 14.2. *If $n \geq 3$, then the fundamental solution for $-\Delta$ is*

$$K(x) = \frac{1}{(n-2)a_n} \cdot \frac{1}{|x|^{n-2}},$$

where a_n is the area of the unit sphere.

Returning to the 2 dimensional case, we want $K = K(r)$, and outside $K = 0$, we want

$$(\partial_r^2 + \frac{1}{r}\partial_r)K = 0.$$

We can write this as

$$(\partial_r + \frac{1}{r})\underbrace{(\partial_r K)}_{L} = 0.$$

This tells us that

$$\frac{L'}{L} = -\frac{1}{r},$$

so

$$\log L = -\log r + c,$$

which we can write as

$$L = c \cdot \frac{1}{r}.$$

Substituting back in for K, we have $\partial_r K = \frac{c}{r}$, which tells us that

$$K = c \ln r + d,$$

where d is a constant that we can choose to fit our problem.

What is the constant c? Instead of a computation, we'll do some carefully selected handwaving. Note that

$$\frac{\partial}{\partial \nu} \log r = -\frac{1}{r},$$

so there is no $n - 2$. We get the last line of the higher-dimensional computation, but without the $n - 2$:

$$c = \frac{1}{a_2} = \frac{1}{2\pi}.$$

So

$$K(x) = \frac{1}{2\pi} \ln r,$$

where we can add a constant if we wish.

Remark 14.3. If we think of the Laplacian in 2 dimensions as $\Delta = \partial\overline{\partial}$, then the fundamental solutions follow

$$K_{-\Delta} = K_{\partial} * K_{\overline{\partial}} = \frac{1}{z} * \frac{1}{\overline{z}}.$$

We get a divergent integral, but with a proper renormalization, we can make sense of this.

15 Introduction to the Fourier Transform

15.1 Motivation: diagonalization for differential operators

We would like to have a better way to think about fundamental solutions to PDEs. Here is an analogy for the Fourier transform. Suppsoe we have a symmetric matrix in $\mathbb{R}^n$. Then A is diagonalizable, with orthonormal eigenvectors $u_1, \dots, u_n$. If you want to better represent your matrix, you can change coordinates to this basis, or you can express an arbitrary vector with $u = c_1 u_1 + \cdots c_n u_n$, where $c_j = u \cdot u_j$. If you have two (or a family of) commuting matrices, you can find an orthonormal basis of eigenvectors for both (or all) matrices simultaneously.

If we have PDEs with constant coefficients, then the operators $P(\partial), Q(\partial), \dots$ are all commuting operators. Can we find a common eigenbasis of functions? Here are some candidates for eigenfunctions $e^{ix\cdot\xi}$, where the i is there to make sure that these don't blow up at ∞. Then

$$P(\partial)e^{ix\cdot\xi} = P(i\xi)e^{ix\cdot\xi},$$

so these exponentials naively serve as eigenfunctions for these operators with eigenvalues $P(i\xi)$. Here, we don't always have real eigenvalues, but we have complex eigenvalues.

Here are some issues:

- Are these functions orthogonal? Consider the Hilbert space $L^2(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \mid \int_{\mathbb{R}^n} |u|^2\, dx < \infty$. If we consider the $L^2(\mathbb{R}^n)$ inner product, $u \cdot v = \int_{\mathbb{R}^n} u(x)v(x)\, dx$ (with v replaced by $\overline{v}$ for complex functions), are these orthonormal? In fact, $e^{ix\cdot\xi} \notin L^2$, so we cannot properly analyze

$$\int_{\mathbb{R}^n} e^{ix\cdot\xi_1} e^{-ix\xi_2}\, dx.$$

- For our diagonaiization, we have uncountably many eigenvectors. $L^2(\mathbb{R}^n)$ is a separable Hilbert space with a countable orthonormal basis. So we have too many functions.

However, we can think of $e^{ix\cdot\xi}$ as *generalized eigenfunctions*. We can still ask the question: Given $f \in L^2(\mathbb{R}^n)$, can we write it as a superposition as $e^{ix\cdot\xi}$? That is, can we write

$$f(x) = \int e^{ix\cdot\xi} c(\xi)\, d\xi?$$

If we disregard the above issues, can we still recover an identity like $c_j = u \cdot u_j$ as before? We may want to try

$$c(\xi) = \int f(x) e^{-ix\xi}\, dx.$$

But since we have trouble normalizing the eigenfunctions, should there be a normalization constant in front?

If we can achieve such a representation, then we get a lot out of it:

$$P(\partial)f = \int e^{ix\cdot\xi} c(\xi) P(i\xi)\, d\xi.$$

So the map $f \mapsto P(\partial)f$ just acts diagonally on this basis: $c(\xi) \mapsto P(i\xi)\cdot c(\xi)$.

15.2 Properties of the Fourier transform

We will use the notation $D_j = \frac{1}{i}\partial_j$, so that $D_j e^{ix\cdot\xi} = \xi_j e^{ix\cdot\xi}$. So we will think of $P(D)$ instead of $P(\partial)$. In this notation, $P(D)e^{ix\cdot\xi} = P(\xi)e^{ix\cdot\xi}$, and we call $P(\xi)$ the **symbol** of P.

Example 15.1. If $P(x,D) = \sum_\alpha c_\alpha(x) D^\alpha$, then the symbol is $P(x,\xi) = \sum_\alpha c_\alpha(x)\xi^\alpha$.

Definition 15.1. The **Fourier transform** of a function f is

$$(\mathcal{F}f)(\xi) = \widehat{f}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\, dx.$$

Our goal is to show that

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \widehat{f}(\xi)\, d\xi.$$

For what f is $\widehat{f}$ well-defined? The integral is absolutely convergent if $f \in L^1$, i.e. $\int |f| < \infty$. We will not use L^1 functions much in our context. If we have $f \in L^1$, then

$$|\widehat{f}(\xi)| \le \frac{1}{(2\pi)^{n/2}} \|f\|_{L^1},$$

which we can write as

$$\|\widehat{f}\|_{L^\infty} \le \frac{1}{(2pi)^{n/2}} \|f\|_{L^1}.$$

The problem is that we want to be able to undo the Fourier transform, and for L^∞ functions, the Fourier transform is not well-defined.

What about the Fourier transform on test functions? If $f \in \mathcal{D}$, then $\widehat{f} \in \mathcal{E}$, so there is no compact support. But if we have $f \in \mathcal{E}$, then $\widehat{f}$ does not exist, since the integral may not converge. It seems that $\mathcal{D}$ is too small, and $\mathcal{E}$ is too large. What should be our intermediate space where $\mathcal{F}$ acts? We will use the Schwartz space $\mathcal{S}$.[13] For $u \in \mathcal{S}$, we want the derivatives to not only be bounded but have decay at infinity.

[13] This is not the same as Schwarz from the Cauchy-Schwarz inequality. Professor Tataru got to meet Schwartz once.

Definition 15.2. The **Schwartz space** is the space of $C^\infty(\mathbb{R}^n)$ functions which are **rapidly decreasing**, in the sense that

$$|x^\alpha \partial^\beta u| \le c_{\alpha,\beta}$$

for all $\alpha, \beta \in \mathbb{N}^n$.

The Schwartz space $\mathcal{S}$ is a locally convex space with seminorms

$$p_{\alpha,\beta}(u) = \|x^\alpha \partial^\beta u\|_{L^\infty}.$$

Theorem 15.1. *The Fourier transform is $\mathcal{F} : \mathcal{S} \to \mathcal{S}$, and the inverse $\mathcal{F}^{-1} : \mathcal{S} \to \mathcal{S}$.*

We have not proven that $(\mathcal{F}^{-1} f)(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \widehat{f}\, dx$ gives the inverse, but we will call it the inverse for now. How do we prove this theorem?

Observe that in the expression $x^\alpha \partial^\beta$, the order of x^α and ∂^β does not matter. How do ∂, x interact with the Fourier transform?

Proposition 15.1. *For $f \in \mathcal{S}$, $\partial_j \widehat{f} = -i\widehat{x_j f}$.*

Proof.

$$\begin{aligned}\partial_j \widehat{f}(\xi) &= \frac{1}{(2\pi)^{n/2}} \int e^{-ix\cdot\xi} f(x)(-ix_j)\, dx \\ &= -i\widehat{x_j f}.\end{aligned}$$

□

Proposition 15.2. *For $f \in \mathcal{S}$, $\xi \widehat{f} = -i\widehat{\partial_x f}$.*

Proof.

$$\xi_j \widehat{f}(\xi) = \frac{1}{(2\pi)^{n/2}} \int e^{-ix\cdot\xi} f(x)\xi_j\, dx$$

Use integration by parts.

$$\begin{aligned}&= \frac{1}{(2\pi)^{n/2}} \int i\frac{\partial}{\partial x_j}(e^{-ix\cdot\xi}) f(x)\, dx \\ &= \frac{1}{(2\pi)^{n/2}} \int -i(e^{-ix\cdot\xi}) f(x)\, dx.\end{aligned}$$

□

So multiplication by x on the physical side is differentiation on the Fourier side, and multiplication by ξ on the Fourier side is differentiation on the physical side.

Proof. If $f \in \mathcal{S}$, then (using $\beta = 0$ and $|\alpha| \le N$ for $N > n$)

$$|f(x)| \le \frac{c_N}{(1+|x|)^N} \in L^1.$$

So $\|\widehat{f}\|_{L^\infty} \le c\|f\|_{L^1}$.

Together, our propositions give us

$$\xi^\alpha \partial_\xi^\beta \widehat{f} = (-i)^{|\alpha|+|\beta|} \widehat{\partial_x^\alpha x^\beta f}.$$

Here, we have

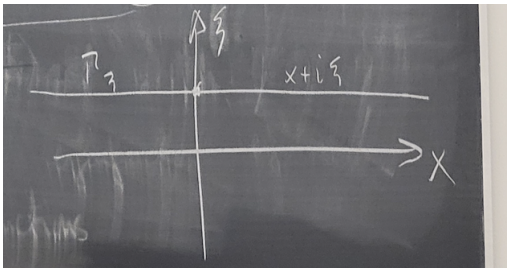
$$\|f^\alpha \partial_\xi^\beta \widehat{f}\|_{L^\infty} \le \|\partial_x^\alpha x^\beta f\|_{L^1}.$$

If $f \in \mathcal{S}$, then $\partial_x^\alpha x^\beta f \in \mathcal{S} \subseteq L^1$. So the right hand side is finite, controlled by finitely many of our Schwartz seminorms. □

Example 15.2 (Fourier transform of a Gaussian)**.** Suppose $f(x) = e^{-x^2/2}$. What is $\widehat{f}$?

$$\begin{aligned}\widehat{f}(\xi) &= \frac{1}{(2\pi)^{n/2}} e^{-x^2/2} e^{-ix\xi}\, dx \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\xi^2/2} \int e^{-(x+i\xi)^2/2}\, dx\end{aligned}$$

How do we deal with this integral? If we write $z = x = i\xi$, we are doing a complex integral on the curve Γ_ξ:



So we get

$$\begin{aligned}\widehat{f}(\xi) &= \frac{1}{(2\pi)^{n/2}} \int_{\Gamma_\xi} e^{-z^2/2}\, dz \\ &= e^{-\xi^2/2} \frac{1}{(2\pi)^{n/2}} \int_{\Gamma_0} e^{-z^2/2}\, dz \\ &= e^{-\xi^2/2} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-x^2/2}\, dx\end{aligned}$$

We can recall $\int e^{-x^2}\, dx = \sqrt{\pi}$, so a change of variables gives

$$= e^{-\xi^2/2}.$$

So we have seen that

$$\mathcal{F}(e^{-x^2/2}) = e^{-\xi^2/2}.$$

In general, what is $\mathcal{F}(e^{-\lambda x^2/2})$? Here is how the Fourier transform behaves under scaling:

Proposition 15.3. *For $f \in \mathcal{S}$,*

$$\widehat{f(\mu\cdot)} = \frac{1}{\mu^n}\widehat{f}(\cdot/\mu).$$

Proof.

$$\mathcal{F}f(\mu x) = \int e^{-ix\cdot\xi} f(\mu x)\, dx$$

Make the change of variables $y = \mu x$.

$$= \frac{1}{\mu^n}\int e^{-iy\cdot\xi/\mu} f(y)\, dy$$
$$= \frac{1}{\mu^n}\widehat{f}(\xi/\mu).$$ □

Remark 15.1. You might call $f(\mu x)$ an L^∞ scaling, whereas $\frac{1}{\mu^n}\widehat{f}(\xi/\mu)$ is an L^1 scaling.

Example 15.3. Setting $\mu = \sqrt{\lambda}$,

$$\mathcal{F}(e^{-\lambda x^2/2}) = \frac{1}{\lambda^{n/2}} e^{-\xi^2/(2\lambda)}.$$

We will work towards the following Fourier inversion theorem:

Theorem 15.2. *$\mathcal{F}^{-1}\mathcal{F} = \mathcal{F}\mathcal{F}^{-1} = I$ in $\mathcal{S}$.*

Remark 15.2. You can think of $\mathcal{F}\mathcal{F}^{-1}$ as the complex conjugate of $\mathcal{F}^{-1}\mathcal{F}$.

16 Fourier Inversion, Plancherel's Theorem, and Temperate Distributions

16.1 Fourier inversion

Last time, we introduced the Fourier transform

$$\mathcal{F}u(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} u(x)\, dx.$$

We had an "inverse"

$$\mathcal{F}^{-1}v(x) = \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} v(\xi)\, d\xi.$$

Both $\mathcal{F}$ and $\mathcal{F}'$ are functions from $\mathcal{S} \to \mathcal{S}$, where $\mathcal{S} = \{\varphi : |x^\alpha \partial^\beta \varphi| \le c_{\alpha,\beta}\}$ is the Schwartz space.

Theorem 16.1. $\mathcal{F}^{-1}\mathcal{F} = \mathrm{Id}$ *on* $\mathcal{S}$.

Proof. Let's first try a brute-force approach and see what happens.

$$\begin{aligned}
\mathcal{F}^{-1}\mathcal{F}u &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \widehat{u}(\xi)\, d\xi \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \int_{\mathbb{R}^n} e^{-i\xi\cdot y} u(y)\, dy\, d\xi \\
&\overset{?}{=} \frac{1}{(2\pi)^n} \iint e^{i(x-y)\cdot\xi}\, d\xi\, dy
\end{aligned}$$

We know $\widehat{u}$ has rapid decay, so the first integral is well-defined. But it is not clear how we can integrate here. The $d\xi$ integral should evaluate to be $\delta_{x=y}$ in some way. Here is what we actually do:

$$= \lim_{\varepsilon\to 0} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} e^{-\frac{\varepsilon}{2}\xi^2} \int_{\mathbb{R}^n} u(y)\, dy$$

Now we can legitimately apply Fubini's theorem.

$$\begin{aligned}
&= \lim_{\varepsilon\to 0} \frac{1}{(2\pi)^n} \iint u(y) e^{i(x-y)\cdot\xi} e^{-\frac{\varepsilon}{2}\xi^2}\, d\xi\, dy \\
&= \lim_{\varepsilon\to 0} \int u(y) e^{-\frac{(x-y)^2}{2\varepsilon}} \varepsilon^{-n/2}\, dy \\
&= \lim_{\varepsilon\to 0} \int u * \varphi_\varepsilon \\
&= u,
\end{aligned}$$

where

$$\varphi_\varepsilon(y) = \frac{1}{(2\pi)^n} e^{-\frac{y^2}{2\varepsilon}} \frac{1}{\varepsilon^{n/2}} \xrightarrow{\varepsilon\to 0} \delta_0.$$

□

16.2 Isometry properties of $\mathcal{F}$ on L^2

Now let's shift our attention to L^2, with inner product $\langle u, v\rangle = \int u\overline{v}\,dx$.

Proposition 16.1. *The Fourier transform is unitary on L^2. That is,*

$$\mathcal{F}^* = \mathcal{F}^{-1}, \qquad (\mathcal{F}^{-1})^* = \mathcal{F}.$$

Proof.

$$\begin{aligned}
\langle \mathcal{F}, uv\rangle &= \iint e^{-ix\xi}u(x)\,dx\,\overline{v}(\xi)\,d\xi \\
&= \iint e^{-ix\cdot\xi}\overline{v}(\xi)\,d\xi\,u(x)\,dx \\
&= \iint \overline{e^{ix\xi}v(\xi)}\,d\xi\,u(x)\,dx \\
&= \langle u, \mathcal{F}^{-1}v\rangle.
\end{aligned}$$

□

This has the following consequence:

Theorem 16.2. *$\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is an L^2-isometry.*

Proof. If we set $u = v$, we get

$$\|u\|_{L^2}^2 = \int |u|^2\,dx = \|\mathcal{F}u\|_{L^2}^2.$$

□

We can use this to extend $\mathcal{F}$ to $L^2(\mathbb{R}^n)$ by density. If $u \in L^2$, find $u_n \in \mathcal{S}$ such that $u_n \to u$ in L^2. Then u_n is Cuachy in L^2, so $\mathcal{F}u_n$ is Cauchy in L^2. So $\lim_{n\to\infty}\mathcal{F}u_n =: \mathcal{F}u$.

Remark 16.1. The Hahn-Banach theorem says that we can extend operators that are densely defined, but in general, there is no guarantee of uniqueness.

However, it is not immediately clear that we can do this approximation of elements of L^2 by elements in $\mathcal{S}$.

Proposition 16.2. *If $u \in L^2$, then there exist $u_n \in \mathcal{D}$ such that $u_n \to u$ in L^2.*

This says that $\mathcal{D}$ is dense in L^2.

Proof. Step 1: Approximate u by compactly supported functions $u = \lim_{n\to\infty} u_n := u\mathbb{1}_{\{|x|\le n\}}$.

Step 2: Regularize $u = \lim_{\varepsilon\to 0} u * \varepsilon$. Here, $\varphi \in \mathcal{D}$ iwth $|int\varphi = 2$, and $\varphi_\varepsilon = \varepsilon^{-n}\varphi(x/\varepsilon)$, so $\varphi_\varepsilon \to \delta_0$ as $\varepsilon \to 0$. So $u * \varphi_\varepsilon \to u$ in $\mathcal{D}'$ if $u \in \mathcal{D}'$ and in L^2 if $u \in L^2$. □

So we get the following theorem:

Theorem 16.3 (Plancherel)**.** *$\mathcal{F} : L^2 \to L^2$ is an isometry.*

16.3 Temperate distributions

Can we extend $\mathcal{F}$ to any larger spaces? First, we will talk about the Fourier transform as a map $\mathcal{F} : \mathcal{S}' \to \mathcal{S}'$.

Definition 16.1. $\mathcal{S}'$, the space of **temperate distributions**, is the space of distributions which exend to continuous linear functionals on $\mathcal{S}$.

$u \in \mathcal{S}'$ if there is a constant c such that for $R\varphi \in \mathcal{S}$,

$$|u(\varphi)| \le c \sum_{\text{finite}} p_{\alpha,\beta}(\varphi), \qquad p_{\alpha,\beta}(\varphi) = \sup |x^\alpha \partial^\beta \varphi|.$$

.

Heer is how we extend $\mathcal{F}$ and $\mathcal{F}'$ to $\mathcal{S}'$: For $u, v \in \mathcal{S}$,

$$\langle \mathcal{F}u, v\rangle = \langle u, \mathcal{F}^{-1}v\rangle,$$

so we have $\mathcal{F}u(\overline{v}) = u(\overline{\mathcal{F}^{-1}}v)$. Replacing v by $\overline{v}$ give $\mathcal{F}u(v) = u(\mathcal{F}v)$, where $u \in \mathcal{S}'$ and $\mathcal{F}v \in \mathcal{S}$. So we can define

$$\mathcal{F}u = u(\mathcal{F}v)$$

for $u \in \mathcal{S}', v \in \mathcal{S}$.

$\mathcal{S} \subseteq \mathcal{E}$, so $\mathcal{E}' \subseteq \mathcal{S}'$. If $u \in \mathcal{E}'$ (is compactly supported), then

$$\mathcal{F}u(\xi) = u\left(\frac{1}{(2\pi)^{n/2}} e^{-x\xi}\right).$$

So we see that $\mathcal{F} : \mathcal{E}' \to \mathcal{E}$. The moral here is that "$\mathcal{F}$ interchanges decay and regularity."

16.4 Examples of temperate distributions

When is a function a temperate distribution? If $u \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$,

$$u(\varphi) := \int u(x)\varphi(x)\,dx,$$

where $\varpi(x)$ is rapidly decreasing. So if $|u(x)| \le c(1+|x|^N)$, then the integral is convergent.

Example 16.1. All rational functions are temperate distributions.

You should not get the idea that these are all the temperate distributions.

Example 16.2. Consider

$$u(x) = e^x \cos e^x.$$

Think of $u = \frac{\partial}{\partial x} \sin e^x = \partial_x f$. Then

$$u(\varphi) = -f(\partial_x \varphi),$$

where $\partial_x \varphi \in \mathcal{S}$ if $\varphi \in \mathcal{S}$. So a temperate distribution may not have much decay if it has enough oscillation, and there is a delicate balance between the two.

Here, if we have $x, \partial : \mathcal{S} \to \mathcal{S}$, we have extended $x, \partial : \mathcal{S}' \to \mathcal{S}'$.

16.5 The Fourier transforms of δ_0 and H

What is $\widehat{\delta_0}$?

$$\widehat{\delta_0}(\xi) = \delta_0\left(\frac{1}{(2\pi)^{n/2}} e^{ix\cdot\xi}\right) = \frac{1}{(2\pi)^{n/2}}.$$

Remark 16.2. People will often change the normalization constant in the Fourier transform to get $\widehat{\delta_0} = 1$. So people will also replace $e^{ix\cdot\xi}$ with $e^{2\pi i x\cdot\xi}$. This is useful if you want to deal with Fourier series or if you want to make a distinction between the $\mathbb{R}^n$ of the input and the $\mathbb{R}^n$ of the ourput. These are actually the same space because $\mathbb{R}^n$ is the cotangent space fo $\mathbb{R}^n$. For more general spaces, the Fourier transform will not have the same input and output domain. We will not need to worry about this for our PDEs.

In 1 dimension, we have $\partial_x H = \delta_0$. Then

$$\mathcal{F}(\partial_x H) = \mathcal{F}(\delta_0),$$

which tells us that $-i\xi\mathcal{F}(H) = \frac{1}{(2\pi)^{n/2}}$. So we get that

$$\widehat{H} = \frac{i}{(2\pi)^{n/2}} \cdot \frac{1}{\xi}.$$

Take u compactly ussported in $[0, \infty)$. Then

$$\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx.$$

Switch to complex numbers $\xi + i\zeta$. This integral ecomes

$$\int e^{-ix\xi + x\zeta} u(x)\, dx.$$

If $\zeta < 0$, we have exponential decay for $x > 0$. So $\widehat{u}(\xi)$ extends to a holomorphic function in $\{\operatorname{Im} z \leq 0\}$.

In this picture, we can think of

$$\widehat{H} = \frac{i}{(2\pi)^{n/2}} \cdot \frac{1}{\xi - i0}.$$

We can also look at

$$\widehat{H - 1} = \frac{i}{(2\pi)^{n/2}} \cdot \frac{1}{\xi + i0}.$$

So if we take the average, we get

$$\widehat{H - \frac{1}{2}} = \frac{i}{(2\pi)^{n/2}} \operatorname{PV} \frac{1}{\xi}.$$

17 Using the Fourier Transform to Find Fundamental Solutions

17.1 The Paley-Wiener theorem and the Fourier transform of even and odd functions

We have been looking at the Fourier transform

$$\widehat{u}(\xi) = \frac{1}{(2\pi)^{n/2}} \int e^{-ix\cdot\xi} u(x)\, dx.$$

We initially defined $\mathcal{F} : \mathcal{S} \to \mathcal{S}$, but we can also define it $L^2 \to L^2$ (with the isometry property) and $\mathcal{S}' \to \mathcal{S}'$. We have also seen that $\mathcal{F} : L^1 \to L^\infty$.

Last time, we also saw that

$$\widehat{H} = \frac{i}{x - i0}.$$

If $u \in \mathcal{S}'$ with $\operatorname{supp} u \subseteq [0, \infty)$, then $\widehat{u}$ has a holomorphic extension to $\{\operatorname{Im} z \leq 0\}$. If u is a measure, then $\widehat{u}$ is bounded in $\{\operatorname{Im} z \leq 0\}$. This leads us to the following property. First, let's generalize this statement.

Suppose $\operatorname{supp} u \subseteq [a, \infty)$. Then

$$\widehat{u}(\xi + i\zeta) = \int e^{ix\xi + x\zeta} u(x)\, dx,$$

so

$$|\widehat{u}(\xi + i\zeta)| \leq e^{a\zeta}.$$

The best we can hope for is a bound of the form $e^{a\zeta}|\xi|^N$.

Theorem 17.1 (Paley-Wiener)**.** *$u \in \mathcal{S}'$ has $\operatorname{supp} u \subseteq [a, \infty)$ if and only if $\widehat{u}$ has a holomorphic extension to the lower half-plane such that*

$$|\widehat{u}(z)| \leq e^{-a \operatorname{Im} z} |z|^N.$$

Remark 17.1. There is a Paley-Wiener theorem in higher dimensions. If $\operatorname{supp} u \subseteq K$ for some compact K, then $\widehat{u}(\xi)$ is defined for $\xi \in \mathbb{C}^n$. Instead of getting the support of u as K in the other direction, we get the convex hull of K.

We can also think of the $e^- ix \cdot \xi$ in the Fourier transform as $\cos(-x \cdot \xi) + i \sin(-x \cdot \xi)$.

- If u is real and even, hen $\widehat{u}$ is real and even.
- If u is real and odd, then $\widehat{u}$ is imaginary and odd.
- If u is imaginary and even, then $\widehat{u}$ is imaginary and even.
- If u is imaginary and odd, then $\widehat{u}$ is real and odd.

17.2 Using the Fourier transform to find fundamental solutions

Suppose we have a constant coefficient partial differential operator $P(\partial)$, and we want to compute a fundamental solution $P(\partial)K = \delta_0$. Let $D = \frac{1}{i}\partial$. Taking the Fourier transform gives

$$P(\xi)\widehat{K} = \frac{1}{(2\pi)^{n/2}}\mathbf{1}.$$

This tells us that

$$\widehat{K} = \frac{1}{(2\pi)^{n/2}}P(\xi).$$

So we can invert the Fourier transform to get K:

$$K = \frac{1}{(2\pi)^{n/2}}\mathcal{F}^{-1}\left(\frac{1}{P(\xi)}\right).$$

Here are some issues.

- $p(\xi)$ may have zeros.
- If p has zeroes, then $\frac{1}{p}$ is not uniquely determined as a distribution.
- This procedure only gives fundamental solutions which are temperate distributions.

The easy case is when $p(\xi) \neq 0$ for any $\xi \in \mathbb{R}^n$. Then $\frac{1}{p} \in \mathcal{S}'$, so this computation is justified.

Example 17.1. Suppose $P = -\partial_x^2 + 1 = D_x^2 + 1$. Then $P(\xi) = (1 + \xi^2)$. So we compute

$$K(x) = \mathcal{F}^{-1}\left(\frac{1}{1+\xi^2}\right).$$

This $K(x)$ is real and even. We are looking at

$$\int_{\mathbb{R}} \frac{1}{\xi^2} e^{ix\xi}\, d\xi.$$

This integrand has a pole at i and a pole at $-i$. However, we can expend this using partial fractions:

$$\frac{1}{1+\xi^2} = \frac{i}{2}\frac{1}{\xi+i} - \frac{i}{2}\frac{1}{\xi-i},$$

where the first term is holomorphic if $\operatorname{Im}\zeta > 0$ and the second is holomorphic if $\operatorname{Im}\zeta < 0$. So the Paley-Wiener theorem tells us that the first one will have an inverse Fourier transform supported in $(-\infty, 0]$, and the second one will have an inverse Fourier transform supported in $[0, \infty)$.

If $x < 0$, we can use complex analysis to say

$$\int_{\mathbb{R}} \frac{1}{\xi + i} e^{ix\xi}\, d\xi = \text{Residue at } i = e^x.$$

A similar computation for $x > 0$ suggests that we should get

$$\int_{\mathbb{R}} \frac{1}{\xi^2} e^{ix\xi}\, d\xi = ce^{-|x|}.$$

In general, if K is a fundamental solution, then so will be $K + K_0$, where K_0 solves the homogeneous equation $P(\partial)K_0 = 0$. In this case, our general solution is $K = ce^{|x|} + c_1e^x + c_2e^{-x}$. We did not get these latter two terms before because they are not temperate distributions.

Example 17.2. If $P = -\Delta + 1$, then $P(\xi) = \xi^2 + 1$ in $\mathbb{R}^n$. Then

$$K = \mathcal{F}^{-1}\left(\frac{1}{1 + \xi^2}\right)$$

gives the unique temperate fundamental solution. Note that $e^{ix\cdot\xi}$ is a solution iff $1+\xi^2 = 0$. In 3 dimensions, this is $K(x) = e^{-|x|}\frac{1}{|x|}$.

Example 17.3. Let $P = -\Delta$, so $P(\xi) = \xi^2$. Then $K = \frac{1}{\xi^2}$ is locally integrable in $\mathbb{R}^n$ if $n \geq 3$. So if $n \geq 3$, we get that $K \in \mathcal{S}'$ is a homogeneous temperate distribution. Since $\frac{1}{\xi^2}$ is homogeneous of order -2, $K = \mathcal{F}^{-1}(\frac{1}{\xi^2})$ will be homogeneous of order $2 - n$.

Proposition 17.1. *If u is homogeneous of order s, then $\widehat{u}$ is homogeneous of order $-n-s$.*

The example to keep in mind to make sure your numbers are right is $\widehat{\delta} = \frac{1}{(2\pi)^{n/2}}$. The Dirac mass is homogeneous of order $-n$, whereas this constant function is homogeneous of order 0.

Example 17.4. If $P = -\Delta$ with $n = 2$, perform the same computation as before, but interpret $\frac{1}{\xi^2}$ as a distribution:

$$\frac{1}{|\xi|^2}(\varphi) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^2 \setminus B(0,\varepsilon)} \frac{\varphi(\xi)}{|\xi|^2}\, d\xi - \varphi(0) \ln \varepsilon,$$

so we pay a price of log, which makes us lose the homogeneity property.

Example 17.5. Suppose $P(\xi) = A\xi \cdot \xi$, where A is a positive deifnite matrix. This is a second order, elliptic, constant coefficient PDE with $P = a^{i,j}\partial_i\partial_j$. We can transform $A \to \text{Id}$ by a linear fransformation. Let $x = By$, so $x \cdot \xi = By \cdot \xi = y \cdot B^\top \xi$. If we carry out the computation, we end up with

$$K = \frac{1}{(A^{-1}x \cdot x)^{(n-2)/2}}.$$

Hormander's book extensively discusses how the Fourier transform behaves under linear changes of coordinates.

17.3 Fundamental solution of the heat equation

Recall the heat equation

$$(\partial_t - \Delta)u = f.$$

We think of u as the temperature of an infinite solid and f as describing the heat sources. This is also called the *diffusion equation*, since we can, for example, interpret $u(t,x)$ as a local concentration of salt in the water of an ocean. In probability theory, the heat equation has connections to Brownian motion, where we let a particle move randomly at every time, independently of the movement at other times.

Our Fourier variables will be ξ (corresponding to x) and τ (corresponding to t). We can write our operator as[14]

$$\partial_t - \Delta = iD_t + D_x^2,$$

so

$$P(\xi,\tau) = iT + \xi^2,$$

which vanishes only at $\tau = 0, \xi = 0$. Is $\frac{1}{i\tau+\xi^2} \in L^1_{\text{loc}}$? Yes! The $1/\tau$ increases the local integrability of this expression, so we will not need to make a distinction between the cases $n = 2$ and $n \geq 3$. We want to calculate

$$\mathcal{F}^{-1}\left(\frac{1}{i\tau+\xi^2}\right).$$

First integrate in τ: We have a pole at $\tau = i\xi^2$. This pole is in the upper half plane, so $\mathcal{F}_\tau^{-1}(\frac{1}{i\tau+\xi^2})$ is supported where $t > 0$. This says that the evolution of heat is well-defined in the future, rather than in the past. We conclude that

$$\mathcal{F}_\tau^{-1}\left(\frac{1}{i\tau+\xi^2}\right) = ce^{-t\xi^2}\mathbb{1}_{\{t\geq 0\}}.$$

for some constant c. Then we can calculate

$$\mathcal{F}^{-1}\left(\frac{1}{i\tau+\xi^2}\right) = \frac{1}{(4\pi t)^{n/2}}e^{-\frac{x^2}{4t}}\mathbb{1}_{\{t\geq 0\}}.$$

Here is another approach. We can try to solve

$$\begin{cases}(\partial_t - \Delta)u = 0\\ u(0) = \delta_0\end{cases}$$

Take the Fourier transform in x to get

$$\begin{cases}(\partial_t + \xi^2)\widehat{u} = 0\\ \widehat{u}(0) = \frac{1}{(2\pi)^{n/2}}.\end{cases}$$

[14] Warning: Evans' book means something different with the D notation.

This gives

$$\widehat{u} = \frac{1}{(2\pi)^{n/2}} e^{-t\xi^2}.$$

So we get the same result.

For $t > 0$, we can consider

$$\begin{cases} (\partial_t - \Delta)u = 0 \\ u(0) = u_0. \end{cases}$$

Extend u to

$$\widetilde{u} = \begin{cases} u & t > 0 \\ 0 & y < 0. \end{cases}$$

Then

$$(\partial_t - \Delta)\widetilde{u} = u_0(x)\delta_{t=0}.$$

Here, $u_0 = \delta_{x=0}$, so $u_0\delta_{t=0} = \delta_{(0,0)}$.

18 The Schrödinger Equation, the Uncertainty Principle, and Oscillatory Integrals

18.1 Fundamental solution of the Schrödinger equation

Recall the heat equation

$$(\partial_t - \Delta)u = f \qquad \text{in } \mathbb{R}_t \times \mathbb{R}^n_x.$$

This has fundamental solution

$$K(t,x) = \frac{1}{(4\pi t)^{n/2}} e^{-x^2/(4t)} \mathbb{1}_{\{t \geq 0\}}.$$

This is the unique temperate distribution for the heat equation.

We also have the Schrödinger equation

$$(i\partial_t + \Delta)u = f \qquad \text{in } \mathbb{R} \times \mathbb{R}^n.$$

Unlike the heat equation, this equation fundamentally has complex-valued solutions. This is the fundamental PDE in quantum mechanics, where $u(t)$ is interpreted as the state of a particle at time t in a probabilistic sense as follows: $\|u\|_{L^2} = 1$, and $|u|^2$ is viewed as a probability distribution. In particular,

$$\mathbb{P}(p \in E) = \int_E |u|^2 \, dx,$$

where p can be the position of a particle. In this picture, the Fourier transform also plays a role. Here, $|\widehat{u}|^2$ is the probability density of the velocity of the particle. Plancherel's theorem tells us that $\|\widehat{u}\|_{L^2} = 1$, as well.

Let $P(\tau, \xi) = \tau - \xi^2$. Then the fundamental solution to the Schrödinger equation should be $K = \mathcal{F}^{-1}(\frac{1}{\tau - \xi^2})$. The issue is that $\tau - \xi^2$ has an entire parabola worth of zeroes. How do we think of $\frac{1}{\tau - \xi^2}$ as a distribution? If we just view this as a distribution in the variable τ, this is like the distribution $\frac{1}{x}$, which gives a few different ways to think of it:

$$\frac{1}{-\tau - \xi - i0}, \qquad \frac{1}{-\tau - \xi^2 + i0}, \qquad \operatorname{PV} \frac{1}{-\tau - \xi^2}.$$

Note that these first two solutions indicate that the Schrödinger equation, unlike the heat equation, can be run backwards in time. How do we pick one of these options? We might want to look for a solution that looks like it's moving forward in time: $\operatorname{supp} K \subseteq \{t \geq 0\}$. This implies that $\widehat{K}$ should have a holomorphic extension in the lower half-plane. Then our forward fundamental solution is

$$K(t,x) = \mathcal{F}^{-1}\left(\frac{1}{-\tau - \xi^2 - i0}\right).$$

First, we will take the Fourier transform with respect to τ. That $\xi = 0$, this gives $H(t)$. Recall that $\mathcal{F}^{-1}\delta_0 = 1$, and $\mathcal{F}^{-1}\delta_{\xi_0} = e^{ix\xi_0}$. This is a general rule for the Fourier transform of the translation of a distirbution, so when $\xi \neq 0$, we get $K(t,\xi) = H(t)e^{-i\xi^2 t}$.

Alternatively, take only a spatial Fourier transform of the Schrödinger equation

$$\begin{cases} (i\partial_t + \Delta)u = 0 \\ u(0) = u_0 = \delta_0(u) \end{cases}$$

to get

$$\begin{cases} (i\partial_t + \xi^2)\widehat{u}(\xi) = 0 \\ \widehat{u}(0) = 1. \end{cases}$$

This gives $\widehat{u}(\xi) = e^{it\xi^2}$, so $u = \mathcal{F}^{-1}(e^{-it\xi^2})$. Recall that $\mathcal{F}(e^{-\xi^2/2} = e^{-x^2/2}$ and more generally that $\mathcal{F}e^{-\lambda\xi^2/2} = \frac{1}{\lambda^{n/2}}e^{-x^2/(2\lambda)}$ for $\lambda \in \mathbb{R}^+$.

Extend this to complex λ. For what complex λ is $\frac{1}{\lambda^{n/2}}e^{-x^2/(2\lambda)}$ a temperate distribution? This is the right half plane $\{\lambda : \operatorname{Re}\lambda \geq 0\}$. For $\operatorname{Re}\lambda > 0$, the function $e^{-\lambda\xi^2/2}$ is analytic with values in $\mathcal{S}$. This tells us that its Fourier transform is analytic for $\operatorname{Re}\lambda > 0$ and we can uniquely extend it to an analytic function on $\{\operatorname{Re}\lambda > 0\}$. What about when $\operatorname{Re}\lambda = 0$? As $\lambda = it + \varepsilon \to it$, $e^{-(it+\varepsilon)\xi/2} \to e^{-it\xi^2/2}$ in $\mathcal{S}$, i.e. in the topology of temperate distributions. So the Fourier transforms converge in the same sense. Thus, we get fundamental solution

$$K(t,x) = \frac{1}{(4\pi it)^{n/2}}e^{ix^2/(4t)}\mathbb{1}_{\{t\geq 0\}}.$$

Remark 18.1. Note that $\widehat{u}(t,\xi) = e^{it\xi}\widehat{u}_0(\xi)$, which means that

$$|\widehat{u}(t,\xi)| = |u_0(\xi)| \implies \|\widehat{u}(t)\|_{L^2} = \|u_0\|_{L^2}.$$

So $|\widehat{u}|$ remains a probability distribution for all time $t \geq 0$.

18.2 The uncertainty principle

Cane we closely predict both position and velocity? Can we have $\operatorname{supp} u \subseteq I$ and $\operatorname{supp}\widehat{u} \subseteq J$ for compactly supported intervals I, J? The answer is no. If $\operatorname{supp} u$ is compact, then $\widehat{u}$ is analytic. So u must be 0.

Let's try to localize our particle at $x = 0$, $\xi = 0$. Let

$$(\delta x)^2 = \int |u|^2(x) \cdot x^2 \, dx$$

be the mean square deviation from 0. We can do the same for velocity to get

$$(\delta\xi)^2 = \int |\widehat{u}|^2(\xi) \cdot \xi^2 \, d\xi.$$

Is there a function $u \in L^2$ with $\|u\|_{L^2} = 1$ such that δx and $\delta\xi$ are simultaneously small? This is not possible. Observe that

$$\delta x = \|x \cdot u\|_{L^2},$$

while Plancherel's theorem tells us that

$$\delta\xi = \|\xi \cdot \widehat{u}\|_{L^2} = \|\partial_x u\|_{L^2}$$

We can compute the inner product

$$\operatorname{Re} \int xu \cdot \overline{\partial_x u}\, dx = \int x \cdot \frac{1}{2} \underbrace{\partial_x |u|^2}_{u\partial_x\overline{u} + \overline{u}\partial_x u}\, dx$$

Now integrate by parts to get

$$\begin{aligned} &= -\int \frac{n}{2}|u|^2\, dx \\ &= -\frac{n}{2}\|u\|_{L^2}^2. \end{aligned}$$

So we conclude that

$$\begin{aligned} \|u\|_{L^2}^2 &= -2n \operatorname{Re}\langle xu, \partial_x u\rangle_{L^2} \\ &\leq 2n\|xu\|_{L^2}\|\partial_x u\|_{L^2}. \end{aligned}$$

So we get the following:

Theorem 18.1 (Uncertainty principle)**.**

$$\delta x \cdot \delta\xi \geq \frac{1}{2n}$$

This says that we cannot know the position of an electron without sacrificing information about its velocity. In physics, people write the Schrödinger equation as $i\partial_t + c\Delta u = f$, where c is a constant involving $\hbar$, **Planck's constant**. This gives the following physically normalized version of the uncertainty principle:

$$\delta x \cdot \delta\xi \geq \frac{\hbar}{2n}$$

18.3 Oscillatory integrals and the KdV equation

We have seen the integral $\int e^{it\xi^2} e^{ix\cdot\xi}$. Can we compute the more general integral $\int e^{i\lambda\varphi(\xi)}\, d\xi$, where φ is a **phase function**? How does this integral behave as $\lambda \to \infty$? Let us make the following observation in 1 dimension.

Proposition 18.1. *If $\varphi' \neq 0$, then for any N,*

$$\int e^{i\lambda\varphi(\xi)} a(\xi)\, d\xi = o(\lambda^{-N}).$$

This is called an **oscillatory integral**.

Proof. Suppose $\varphi' \neq 0$. Then localize to a compact set with a function a and integrate by parts:

$$\begin{aligned}\int e^{i\lambda\varphi(\xi)} a(\xi) &= \int \varphi' e^{i\lambda\varphi} \cdot \frac{a}{\varphi'}\, d\xi \\ &= \frac{i}{\lambda} \int e^{i\lambda\varphi(\xi)} \partial_\xi \left(\frac{a}{\varphi'}\right) d\xi,\end{aligned}$$

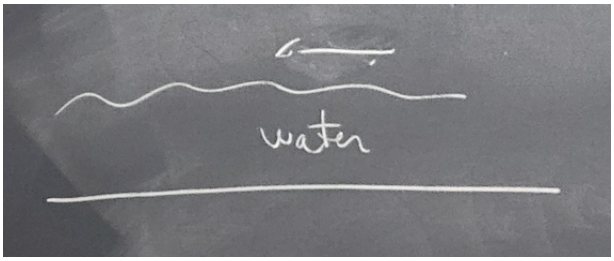
so we have gained a factor of $1/\lambda$. Now repeat this. □

The conclusion is that the main contribution comes from the critical points of φ. The study of oscillatory integrals via their critical points is called the **method of stationary phase**.[15] From the perspective of PDEs, we want to use oscillatory integrals to compute asymptotic expansions of fundamental solutions which are not explicit.

Example 18.1 (KdV equation[16])**.** The KdV equation is

$$(\partial_t + \partial_x^3)u = 0.$$

It describes unidirectional waves in a canal.



If you want to make this a linear equation, we can consider the case where this equals $6uu_x$. Let's compute a fundamental solution. We want to compute the inverse Fourier transform of $\frac{1}{\tau - \xi^3}$. For a forward fundamental solution, we want

$$K = \mathcal{F}^{-1}\left(\frac{1}{\tau - \xi^3 - i0}\right).$$

[15] This is an important topic in harmonic analysis, and people have spent their whole careers studying oscillatory integrals.

[16] This is short for Korteweg-de Vries.

We have

$$K(t, \xi) = e^{it\xi^3}$$

If we take the Fourier transform in time, we get $K(t, \xi) = e^{it\xi^3}$. So now we want to take the integral

$$\int e^{i(t\xi^3 + x\xi)}\, d\xi$$

The solution will not be an algebraic function; instead, it will be something we label as a "special function," the **Airy function**. In particular, $\mathcal{F}^{-1}(e^{i\xi^3}) = \mathrm{Ai}(x)$.

Let's try to compute the asymptotic behavior. The phase is $\varphi(\xi) = t\xi^3 + x\xi$. The critical points are when

$$3t\xi^2 + x = 0 \implies \xi^2 = -\frac{x}{3t}.$$

This has roots only when $x < 0$, which is why this equation only gives waves in 1 direction. We get two critical points:

$$\xi^1 = \sqrt{-\frac{x}{3t}}, \qquad \xi^2 = \sqrt{-\frac{x}{3t}}$$

At each critical point, replace the cubic polynomial with a quadratic polynomial which is the Taylor series of the polynomial, and take the Fourier transform like with our analysis of the Schrödinger equation.

19 The KdV Equation and the Wave Equation

19.1 Fundamental solution of the KdV equation

Last time, we were discussing the KdV equation

$$(\partial_t + \partial_x^3)u = f.$$

We saw that the fundamental solution was given by

$$\widehat{K}(t, \xi) = e^{it\xi^3}.$$

Taking the inverse Fourier transform in x gives

$$K(t, x) = \int e^{i(t\xi^3 - x\xi)}\, d\xi.$$

This problem admits a type of scaling. If we want

$$(\partial_t + \partial_x^3)u = 0,$$

then we can make a change of variables $u(x, t) \mapsto u(\lambda x, \lambda^3 t)$. If we want to get rid of the time variable, we can set $\xi = t^{-1/3}\xi$, so the integral becomes

$$\begin{aligned} K(t, x) &= t^{-1/3} \int e^{i\eta^3 + x/t^{1/3}\eta}\, d\eta \\ &= t^{-1/3} K(1, x/t^{1/3}) \\ &= t^{-1/3}\, \mathrm{Ai}(x/t^{1/3}), \end{aligned}$$

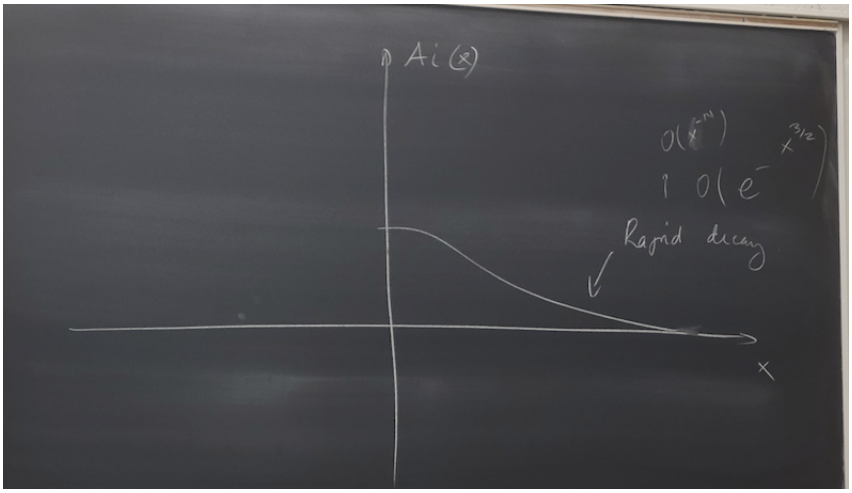
where

$$\mathrm{Ai}(x) = \mathcal{F}^{-1}(e^{i\xi^3}) = \int e^{i(\xi^3 + x\xi)}\, d\xi.$$

In this integral, we have the phase function $\varphi(\xi) = \xi^3 + x\xi$. The critical points, with $\varphi_\xi = 0$, are $\xi_{1,2} \pm \sqrt{-x/3}$ with $x < 0$.

Let's draw a picture of the Airy function; this is real-valued because the equation is real, so the real and imaginary parts of any solution should also be solutions. At $+\infty$, we have no stationary points, so we expect rapid decay. This decay is $O(e^{-x^{3/2}})$, which one can prove by changing the contour in the integral (to some other integral over a contour

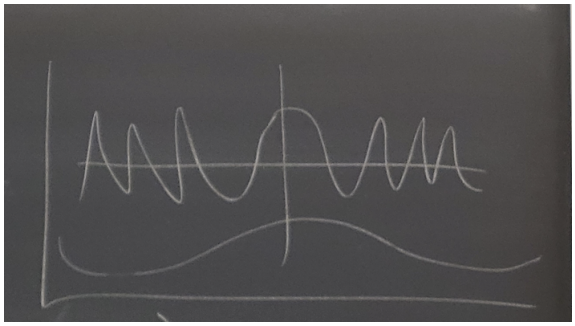
in the complex plane).



Choose $\xi = \sqrt{-x/3}$, look at the contribution around ξ_1, and take $2\,\mathrm{Re}$.

$$\varphi(\xi) = \varphi(\xi_1) + \frac{1}{2}\varphi''(\xi_1)(\xi - \xi_1)^2 + \underbrace{O((\xi - \xi_1)^3)}_{\text{discard}}$$

We are multiplying two functions, a Gaussian and a function with oscillation.



Recall that $\mathcal{F}^{-1}(e^{i\lambda\xi^2/2}) = -\frac{1}{(i\lambda)^{n/2}} e^{ix^2/(2\lambda)}$. Now obsesrve that the Fouerier transform lets us figure out the integral of a function: $\widehat{u}(0) = \frac{1}{(2\pi)^{n/2}} = \int u(x)\,dx$. So can calculate this integral:

$$\int e^{i(\varphi(\xi_1) + \frac{1}{2}\varphi'(\xi_1)(\xi - \xi_1)^2)}\,d\xi = e^{i\varphi(\xi_1)} \frac{1}{(i\varphi''(\xi_1))^{1/2}}.$$

Now write

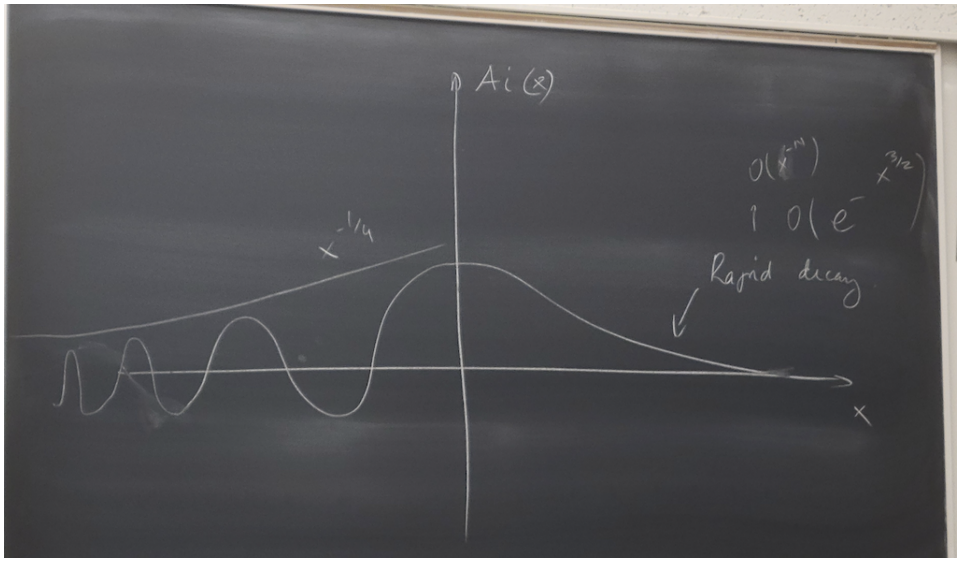
$$\varphi(\xi_1) = \xi_1(\xi_1^2 + x) = \frac{2}{3} x\sqrt{-x/2} = c(-x)^{3/2},$$

$$\varphi''(\xi) = 6\xi = c(-x)^{1/2}$$

In total, we get something of the form

$$e^{ic(-x)^{3/2}}(-x)^{-1/4}.$$

This left term oscillates faster and taster, while the right term has a decay. So we can improve our picture of the Airy function:



The homework says that $\mathrm{Ai}''(x) = x\,\mathrm{Ai}(x)$ (up to some constants/signs). There are two solutions to this equation; why are we only getting the Airy function? This is because using the Fourier transform only solves for temperate solutions. The other solution will look like the Airy function for negative x but has exponential (specifically $e^{+x^{3/2}}$) growth as $x \to \infty$. The Airy function has nice properties, and it actually extends to a holomorphic function.[17]

19.2 Analysis of the wave equation

Definition 19.1. The **d'Allembertian** is the partial differential operator

$$\Box = \partial_t^2 - \Delta_x.$$

Definition 19.2. The **wave equation** is the equation

$$\begin{cases} \Box u = f \\ u(t=0) = u_0 \\ \partial_t u(t=0) = u_1 \end{cases}$$

[17] Professor Tataru really likes the Airy function. He used to put it on exams, until one time when he put it on a calculus exam. That didn't go so well.

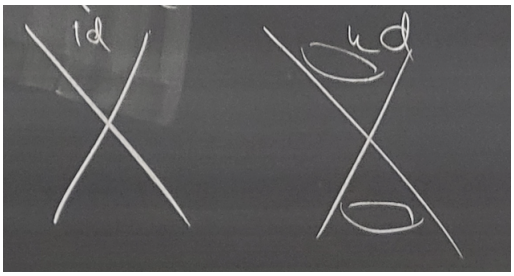
This is an evolution equation which is 2nd order in t.

What does the wave equation model? In 1 dimension, this modes an elastic string. In 2 dimensions, it models an elastic drum, and in 3 dimensions, it models an elastic solid. The wave equation also models, sound, light, and electromagnetism.

Our goal is to find the fundamental solution. The symbol for the equation is $P(\tau, \xi) = -\tau^2 + \xi^2$. Then we get

$$K(t,x) - \mathcal{F}^{-1}\left(\frac{1}{-\tau^2+\xi^2}\right).$$

The zero set of P, (the **characteristic set**) contains the points where $\tau^2 = \xi^2$. In 1 dimension, this looks like an X, but in n dimensions, this looks like 2 cones.



Like we have seen before, this is not uniquely defined as a distribution. We want to pick a forward fundamental solution, so we will look at this as a function of τ and think of this as a function which is holomorphic in the lower half plane:

$$K(t,x) = \mathcal{F}^{-1}\left(\frac{1}{-(\tau - i0)^2 + \xi^2}\right).$$

We will take the Fourier transform first in τ and then in ξ. First, expand the fraction into partial fractions:

$$\frac{1}{-\tau^2+\xi^2} = \frac{A}{\tau - |\xi|} + \frac{B}{\tau + |\xi|},$$
$$A = -\frac{1}{2|\xi|}, \qquad B = \frac{1}{2|\xi|}.$$

So

$$\begin{aligned}\mathcal{F}^{-1}\left(\frac{1}{-(\tau - i0)^2+\xi^2}\right) &= -H(t)\frac{e^{it|\xi|} - e^{-it|\xi|}}{2|\xi|} \\ &= -i\frac{\sin(t|\xi|)}{|\xi|}.\end{aligned}$$

This is hard to compute the Fourier transform directly in n dimensions, but the 1-dimensional computation is easier: We are looking at

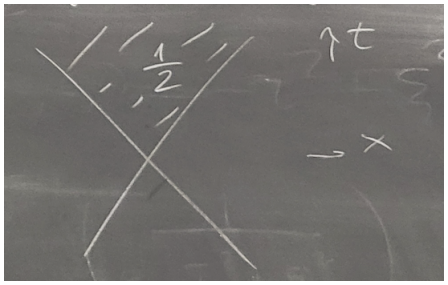
$$-i\frac{\sin(t\xi)}{\xi} = \frac{1}{2}\frac{e^{it\xi} - e^{-it\xi}}{\xi - i0}.$$

Taking the inverse Fourier transform, we note that multiplying phase factors just translate the Fourier transform. We get

$$\frac{1}{2}(H(x+t) - H(x-t)) = \frac{1}{2}\mathbb{1}_{[-t,t]}.$$

Theorem 19.1. *The fundamental solution to the wave equation in 1 dimension is*

$$K(t,x) = \begin{cases} 1/2 & x > 0, -t \le x \le t \\ 0 & \textit{otherwise.} \end{cases}$$



How should we approach this for $n \ge 2$? The distribution $\frac{1}{-(\tau - i0)^2 + \xi^2}$ is homogeneous of order -2, so K will be homogeneous of order $(-n-1) - (-2) = -n+1$. We could try to replace x by $r = |x|$, making an *ansatz* that the solution is radial, but this is not very nice because we still have a PDE in 2-dimensions. This is easier to solve in 3 dimensions, so we can add a dimension and then pretend it doesn't exist after we solve the equation; this is probably how it was done in the early 1900s.

Instead, let's look at the symmetries of $\Box$. This is translation-invariant and invariant under rigid rotations in x. The latter suggests that we could look for more general linear transformations which $\Box$ is invariant under. We will change our notation from (t,x) to $(x_0, \dots, x_n)$, where $t = x_0$ and $x = (x_1, \dots, x_n)$. Here's how to make sure we won't get confused about which variables are time. A notational convention which goes back to Einstein says we write x_α for $\alpha = 0, \dots, n$ and x_j for $j = 1, \dots, n$. Now apply a change of variables to get $y = Ax$. Then

$$\frac{\partial}{\partial x_k} = \frac{\partial y_j}{\partial x_k}\frac{\partial}{\partial y_j} = a_{j,k}\frac{\partial}{\partial y_j}.$$

Here, we are using Einstein's summation convention, where we omit the sum over j.

Change convention so that $\Box = -\partial_t^2 + \Delta_x$ and write $\Box = m^{\alpha,\beta}\partial_\alpha\partial_\beta$, where[18]

$$M = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

If we switch to the coordinates of $y = Ax$, we have

$$\begin{aligned} \Box &= m^{\alpha,\beta} a_{\alpha,\gamma}\partial_{y,\gamma} a_{\beta\delta}\partial_{y\delta} \\ &= b_{\gamma,\delta} \cdot \partial_{y,\gamma}\partial_{y,\delta}, \end{aligned}$$

Here, $\beta_{\gamma,\delta} = a_{\alpha,\gamma} m^{\alpha,\beta} a_{\beta,\delta}$. This new matrix is $A^\top MA$. Then A is a symmetry of $\Box$ iff $M = A^\top MA$.

Next time, we will find what the symmetries are.

[18] The letter m comes from Minkowski. instead of Euclidean space, we can think of this as a Minkowski space

20 Fundamental Solutions to the Wave Equation

20.1 Lorentz invariance of fundamental solution

Last time, we were solving the wave equation

where

$$\begin{aligned}\Box &= \partial_t^2 - \Delta_x \\ &= m^{\alpha,\beta}\partial_\alpha\partial_\beta\end{aligned}$$

in coordinates $t = x_0$ and $\partial_t = \partial_0$. The matrix M is given by

$$M = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

Last time, we determined that a fundamental solution is homogeneous of order $1-n$ and must move forward in time. We looked at a symmetries of the equation when we make a linear change of coordinates $x = Ay$. We saw that such a linear change of coordinates leaves $\Box$ unchanged if and only if

$$A^\top M A = M.$$

This is a group, called the **Lorentz**[19] **group**; if M were the identity matrix, this would be the group or orthogonal matrices. What are the generators for this group?

1. Rigid rotations: $A = \begin{bmatrix} 1 & 0 \\ 0 & O \end{bmatrix}$, where O is an $n \times n$ orthogonal matrix. These were the symmetries corresponding to the Laplacian.

2. Look at 1+1 dimensions and leave the rest unchanged: Since we can apply rotations to the last n dimensions, we only need to mix the time dimension and the first space dimension. Observe that

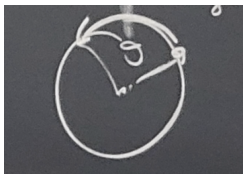
$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

   If -1 were 1, we would get rotations:

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \text{rotation by angle } \theta.$$

[19] This is not to be confused with Lorenz, another mathematician who also had a hand in some things we will be discussing today. To make matters worse, they even wrote a paper together!

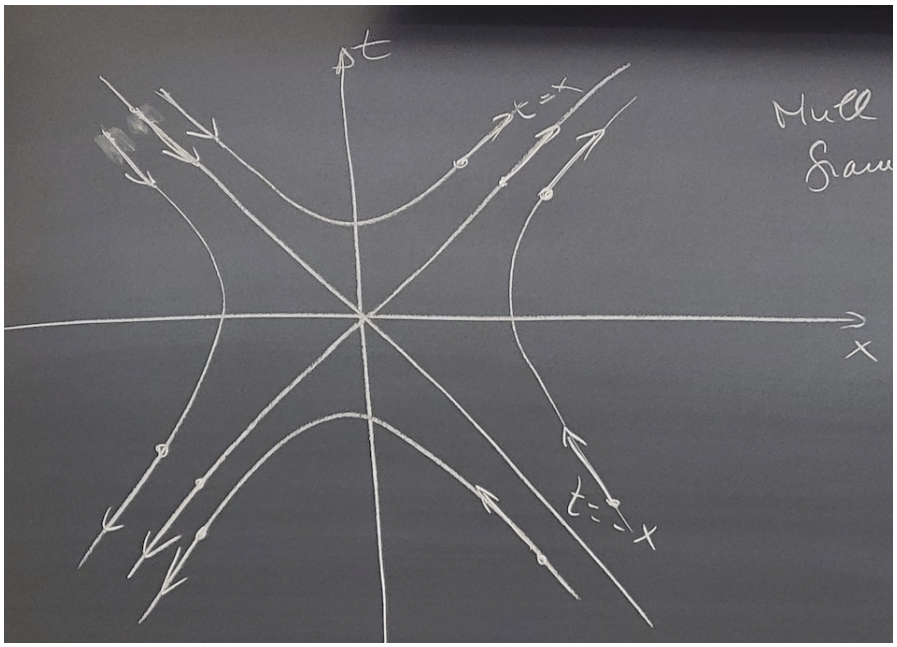
This keeps $t^2 + x^2$ unchanged; this is like rotating around a point in a circle by angle θ.



With the -1, we get

$$A = \begin{bmatrix} \cosh\varphi & \sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{bmatrix} = \text{hyperbolic rotation by angle } \varphi.$$

Such matrices keep $t^2 - x^2$ unchanged. Rather than circles, here's what the level sets look like:



Here, we dilate the $t = x$ direction and shrink the $t = -x$ direction. This suggests that we make a change of variables $u = t + x$ and $v = -x$. Then $\partial_t^2 - \partial_x^2 = 4\partial_u\partial_v$. Then the transformation $u \mapsto \lambda u$, $v \mapsto \lambda^{-1}v$ preserves the operator in this **null frame**.

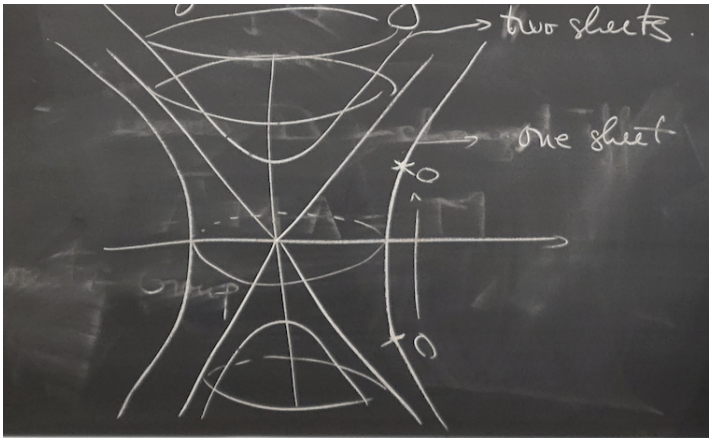
Theorem 20.1. *The Lorentz group is generated by rigid spatial rotations and 1-d hyperbolic rotations.*[20]

[20] Hyperbolic rotations are sometimes referred to as *Lorentz boosts*. These hyperbolic rotations are what happens in special relativity when you switch between observers in different reference frames.

We say that the solution to the wave equation is **Lorentz invariant**.

20.2 Calculation of fundamental solutions

We now know that the fundamental solution of the wave equation should be a "function" of $t^2 - x^2$. Here is what the picture should look like in higher dimensions.



The level sets should be forward and backward cones and hyperboloids. We get 1-sheeted and 2-sheeted hyperboloids. On the 1-sheeted hyperboloids, the forward in time points are connected to the backwards in time points, which must give 0 for our forward time solution. So these must be 0. Thus, $K = K(t^2 - x^2) = K(y)$ must be supported in the forward cone $\{t^2 - x^2 \geq 0\}$.[21] We want a homogeneous distribution of y which is $\frac{1-n}{2}$ homogeneous (since we are now working with the squares of t, x) and supported in $y \geq 0$.

- In 1 dimension, we want a homogeneous distribution of order 0, supported where $y > 0$. So $K(y) = cH(y)$, and we saw earlier that this constant is $c = 1/2$.
- In 2-dimensions, we want a homogeneous distribution of order $-1/2$, supported where $y > 0$. So we get

$$K(y) = \begin{cases} c_2 \frac{1}{\sqrt{y}} & y > 0 \\ 0 & y \leq 0. \end{cases}$$

  So we get

$$K(t, x) = c_2 \frac{1}{\sqrt{(t^2 - x^2)_+}} \mathbb{1}_{t \geq 0}.$$

[21] A backward time solution would still be 0 on the sides. It would just be supported on the backward cone.

- In 3-dimensions, we cannot get a function which is homogeneous of order -1. The two distributions that span the space of homogeneous distributions of order -1 are δ_0 and $\mathrm{PV}\,\frac{1}{y}$. The latter is supported everywhere, so we take $K(y) = \delta_{y=0}$.

$$K(t,x) = c_3 \delta_{t^2-x^2=0} \mathbb{1}_{t\geq 0}.$$

- In 4 dimensions, we need homogeneity of order $-3/2$. However, $\frac{1}{y_+^{3/2}} \notin L^1_{\mathrm{loc}}$. Define

$$\frac{1}{y_+^{3/2}} := -2\partial_y \frac{1}{y_+^{1/2}}.$$

  This is a distribution, not a function. We can repeat this differentiation procedure to get a solution for all even dimensions.

- In 5 dimensions, we can get a solution which is homogeneous of order 2 by differentiating $\delta_{y=0}$. We can keep differentiating to get solutions in all odd odd dimensions.

20.3 Determination of constants for fundamental solutions

Here is a formal computation: If $\Box u = f$, let's see how $\int u\,dx$ behaves as a function of time.

$$\frac{d}{dt}\int u\,dx = \int u_t\,dx.$$

$$\frac{d^2}{dt^2}\int u\,dx = \int u_{tt}\,dx = \int \nabla u + f\,dx = \int f\,dx,$$

since we can get rid of the Laplacian using integration by parts. If $f = \delta_0$ and $u = K$, then $u = 0$ for $t < 0$, so

$$I(t) = \int u\,dx = 0 \qquad \text{for all } t < 0.$$

Additionally, we get

$$I''(t) = \delta_{t=0}.$$

This tells us that

$$I(t) = t\mathbb{1}_{t\geq 0},$$

so

$$\int K(t,x)\,dx = t.$$

- In 2 dimensions, we have

$$K(t,x) = \frac{c_2}{\sqrt{(t^2-x^2)_+}},$$

so the equation

$$t = c_2 \int \frac{1}{\sqrt{(t^2 - x^2)_+}}\, dx$$

holds for all t. if we set $t = 1$, then we get

$$1 = c_2 \int_{B(0,1)} \frac{1}{\sqrt{1 - r^2}} r\, dr\, d\theta = c_2 2\pi \left[-\sqrt{1 - r^2} \right]_0^1,$$

which tells us that

$$c_2 = \frac{1}{2\pi}.$$

- In 3 dimensions, we want to find c_3. What is $\delta_{t^2 - x^2}$?

$$\delta_0 = \frac{1}{2\pi i} \left(\frac{1}{y - i0} - \frac{1}{y + i0} \right),$$

so we can write

$$\delta_{t^2 - x^2} = \frac{1}{2\pi i} \left(\frac{1}{t^2 - x^2 + i0} - \frac{1}{t^2 - x^2 - i0} \right).$$

Note that

$$\frac{t^2 - x^2}{=} \frac{1}{t - |x|} \frac{1}{t + |x|},$$

where the left term vanishes on the cone, and $t + |x|$ is $2t$ on the cone. so we can write

$$\delta_{t^2 - x^2 = 0} = \underbrace{\delta_{t = |x|}}_{\text{surface measure on } |x| = t} \cdot \frac{1}{2t}.$$

If we have a surface $\Sigma = \{\phi = 0\}$, this is like normalizing to make $|\nabla \phi| = 1$.

The computation becomes

$$K(t, x) = c_3 \frac{1}{t} \delta_{|x| = t}.$$

$$t = \int \frac{c_3}{t} \delta_{|x| = t}\, dx = \frac{c_3}{t} \underbrace{\text{Area}(\{|x| = t\})}_{= 4\pi t^2}.$$

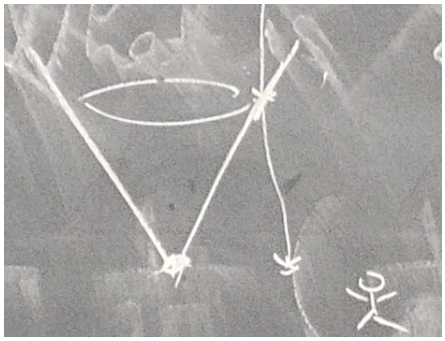
so we get

$$c_3 = \frac{1}{4\pi}.$$

20.4 Physical interpretation of solutions to the wave equation

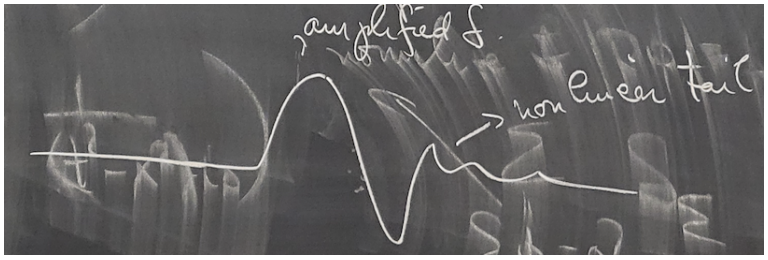
Here are two key properties of the wave equation:

1. All forward solutions are supported on the forward cone. This is referred as the **finite speed of propagation**. This says that waves move with speed ≤ 1. If we normalize the equation with physical constants to get $c^2\partial_t^2 - \Delta_x$, where c is the speed of light, then this says that no waves move faster than the speed of light. An observer at position x only observes the wave at the time at which the cone hits the observer's timeline:



2. Consider 3 dimensions, where the fundamental solution K is supported exactly on the cone. Here, waves hit the observer just once, and we don't see them again. This is called the **Huygens principle**.

Remark 20.1. The equations of physics are nonlinear; this linear PDE is just the best linear approximation. The finite speed of propagation remains, but Huygen's principle does not hold in general. When scientists observed gravitational waves recently, they observed both a Dirac mass and a nonlinear tail.



20.5 Next steps: Fourier series

Our next goal is to learn about the connection between the Fourier transform and Fourier series. The Fourier transform $\widehat{u}$ of $u : \mathbb{R}^n \to \mathbb{C}$ is given by

$$u(x) = \int \widehat{u}(\xi) e^{ix\cdot\xi}\, d\xi.$$

In calculus, you may have encountered Fourier series:

Definition 20.1. If $u : [0, 2\pi] \to \mathbb{C}$, then the **Fourier series** for u is given by

$$u(x) = \sum_n c_n e^{inx} = \sum_n c_n(\cos(nx) + i\sin(nx)).$$

Not all PDEs can be solved; we will see more about this next time.

21 Fourier Transforms of Periodic Functions and Local Solvability of Partial Differential Operators

21.1 Fourier transforms of periodic functions

A function f is periodic if

$$f(x) = f(x + a)$$

for some a and for all x.

Definition 21.1. $f \in \mathcal{D}'$ is **periodic** of period a if

$$f(\phi) = f(\phi(\cdot + a))$$

Suppose f is periodic; what can we say about $\widehat{f}$? Recall that for functions,

$$\widehat{f}(\cdot + a) = e^{ia\cdot\xi}\widehat{f}.$$

Using the periodic condition, write this as the multiplication

$$\widehat{f}(1 - e^{ia\xi}) = 0.$$

Note that $1 - e^{ia\xi} \neq 0$ unless $\xi = \frac{2\pi n}{a}$. Then $\operatorname{supp} \widehat{f} \subseteq \frac{2\pi n}{a}\mathbb{Z}$. As an analogy look at the condition $xf = 0 \implies f = c\delta_0$; here, we have zeros at many points. So we conclude that

$$\widehat{f} = \sum_n c_n \delta_{\frac{2\pi n}{a}}.$$

Theorem 21.1. *The coefficients c_n are the Fourier coefficients for f in the interval $[0, a]$, and*

$$f(x) = \sum_n c_n e^{\frac{2\pi i}{a}n}.$$

Here, we have ignored the factors of 2π.

Remark 21.1. We can multiply f by $e^{-\frac{2\pi i}{a}m}$ and integrate from 0 to a to get

$$c_n = \int f(x) e^{-\frac{2\pi i}{a}mx}.$$

Example 21.1. The simplest periodic distribution is

$$f_a = \sum_n \delta_{na}.$$

Then

$$\widehat{f_a} = \sum_n c_n \delta_{\frac{2\pi}{a}n}.$$

If we write

$$f_a(1 - e^{\frac{2\pi i x}{a}}) = 0,$$

then we get

$$\widehat{f_a} = \widehat{f_a}(\cdot + \tfrac{2\pi}{a}).$$

Thus, all the c_ns are the same. So

$$\widehat{f_a} = c_a \sum_n \delta_{\frac{2\pi n}{a}} = c_a f_{\frac{2\pi}{a}}.$$

What is c_a? Apply this to a Schwarz function: $\widehat{f}(\phi) = f(\widehat{\phi})$ by definition, so

$$c_a \sum_{n\in\mathbb{Z}} \phi(\tfrac{2\pi n}{a}) = \sum_{m\in\mathbb{Z}} \widehat{\phi}(ma).$$

This is called the **Poisson summation formula**.

Now what happens if we replace ϕ by $\phi e^{ix\cdot\xi_0}$? Then $\widehat{(}\xi)$ becomes $\widehat{\phi}(\xi - \xi_0)$. The Poisson summation formula gives

$$c_a \sum_n \phi(\tfrac{2\pi n}{a}) e^{i\frac{2\pi n}{a}\xi_0} = \sum \widehat{\phi}(ma - \xi_0).$$

The dependence of ξ_0 on the left hand side is simple. Integrate to get

$$\underbrace{\int_0^a \sum_n \phi(\tfrac{2\pi n}{a}) e^{i\frac{2\pi n}{a}\xi_0}\, d\xi_0}_{=ac_a\phi(0)} = \int_0^a \sum_m \widehat{\phi}(ma - \xi_0)\, d\xi_0$$

$$= \int \widehat{\phi}(\xi)\, d\xi$$
$$= \widehat{\phi}(1)$$
$$= \phi(\tfrac{1}{\sqrt{2\pi}}\delta_0)$$
$$= \frac{1}{\sqrt{2\pi}}\phi(0).$$

Accounting for the constants we ignored before, we get

$$c_a = \frac{1}{2\pi a}.$$

Remark 21.2. We can use the Poisson summation formula to compute all sorts of series. Recall that $\mathcal{F}(\frac{1}{1+x}) = ce^{-|\xi|}$ (perhaps omitting constants). Choose $a = 2\pi$. The Poisson summation formula tells us that

$$\sum_{m\in\mathbb{Z}} \frac{1}{n^2+1} = \sum_m e^{-2\pi|m|} = \frac{2}{1 - e^{2\pi}} - 1,$$

where we have ignored the constants.

21.2 Local solvability of partial differential operators

Let $P(D)$ be our partial differential operator with constant coefficients.

Definition 21.2. $P(D)$ is **solvable** if for each f, the equation $P(D)u = f$ admits at least one solution.

If $f \in \mathcal{D}'$, then $u \in \mathcal{D}'$. If $f \in \mathcal{S}$, then $u \in \mathcal{S}$. In general, the regularity of f and u will be related, so when we say $P(D)$ is solvable, we specify a class of functions f.

Definition 21.3. $P(D)$ is **locally solvable** if for each $f \in \mathcal{E}'$, there exists a solution $u \in \mathcal{D}'$ in a neighborhood of the support of f.

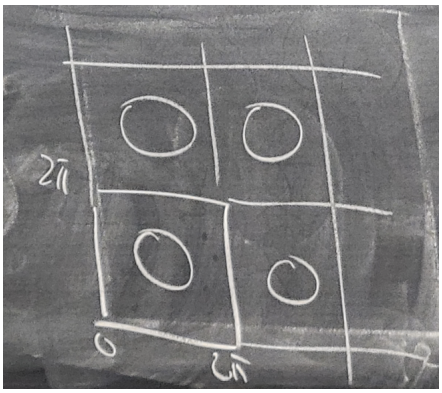
If $u \in \mathcal{E}'$, then $P(\xi)\widehat{u}(\xi) = \widehat{f}(\xi)$ for $\xi \in \mathbb{C}^n$. Here is a narrower version, which we may regard as the "real definition" of local solvability:

Definition 21.4. $P(D)$ is **locally solvable** if for each x_0, there is an $\varepsilon > 0$ such that if $\operatorname{supp} f \subseteq B(x_0, \varepsilon)$, then a solution exists.

For today, we will deal with the first, more relaxed definition.

Theorem 21.2. *Every constant coefficient partial differential operator is locally solvable (in the relaxed sense).*

Proof. Suppose f is supported in $B \subseteq [0, 2\pi]^n$. Take $\widetilde{f}$ to be the periodic extension of f, and look for a periodic solution $\widetilde{u}$ to $P(D)\widetilde{u} = \widetilde{f}$.



What does this periodization do? Originally, $P(D)u = f$ gives $P(\xi)\widehat{u} = \widehat{f}$, so $\widehat{u} = \frac{1}{P(\xi)}\widehat{f}$. However, this has issues because $P(\xi)$ can have issues. In the periodic case, we know

$$\widehat{\widetilde{f}}(\xi) = \sum_{m \in \mathbb{Z}^n} f_m \delta_m,$$

$$\widehat{\widetilde{u}}(\xi) = \sum_{m\in\mathbb{Z}} u_m \delta_m.$$

We need $P(m)u_m = f_m$, which gives

$$u_m = \frac{f_m}{P(m)}, \qquad m \in \mathbb{Z}^n.$$

The advantage is that we only $P(m) \neq 0$ on lattice points $m \in \mathbb{Z}^n$. However, the Fourier transform is defined for temperate distributions, so we need about on $\frac{f_m}{P_m}$. More precisely, we need a bound

$$|P(m)| \geq (1 + |m|)^{-N}$$

What if P has zeroes on the lattice points? Make the change of notation $f \mapsto fe^{ix\cdot\xi} = g$, so $u \mapsto ue^{ix\xi} = v$. We can ask this question for the phase-shifted variables. To study our equation, we need to expand

$$P(D)u = P(D)(ve^{-ix\cdot\xi}).$$

To use the Leibniz rule, note that,

$$\begin{aligned} D_j(ve^{-ix\xi}) &= Dve^{-ix\cdot\xi} + vD_je^{-ix\cdot\xi} \\ &= e^{-ix\cdot\xi}(D_jv - v\xi_j) \\ &= e^{-ix\cdot\xi}(D_j - \xi_j)v, \end{aligned}$$

We can write this as $e^{ix\cdot\xi}D_je^{-ix\xi} = D_j - \xi_j$, which we may think of as a **conjugation**. Referring to our equation, we get

$$\begin{aligned} P(D)u &= P(D)(ve^{ix\ cdot\xi} \\ &= e^{-ix\cdot\xi}p(D-\xi)v \\ &= f, \end{aligned}$$

which tells us that we have replaced $P(D)u = f$ with

$$P(D-\xi)v = g.$$

So we only need to solve the new periodic problem is to define

$$v_m = \frac{g_m}{P(m-\xi)}, \qquad m \in \mathbb{Z}.$$

Now we only need to find some $\xi \in [0,1]^n$ such that

$$|P(m-\xi)| \geq (1+|m|)^{-N} \qquad \forall m.$$

The following lemma tells us we can find such a ξ.

Lemma 21.1. *If δ is small enough, then*

$$\int \frac{1}{(P(\eta))^{\delta}} \frac{1}{(1+|\eta|)^{N}}\, d\eta < \infty.$$

Proof. In 1 dimension, use partial fractions. Then reduce any number of dimensions to the 1-dimensional case. □

How does this help us? Write $\eta = m + \xi$ with $m \in \mathbb{Z}^n$ and $\xi \in [0,1]^n$. Then

$$\int_{\xi} \sum_{m} \frac{1}{P(m-\xi)|^{\delta}} \frac{1}{(1+|m|)^{N}}\, d\eta < \infty.$$

So for almost every ξ,

$$\sum_{m} \frac{1}{|P(m-\xi)|^{\delta}} \frac{1}{(1+|m|)^{N}} = M < \infty.$$

This tells us that

$$|P(m-\xi)| \geq M^{-1/d}(1+|m|)^{-N/\delta},$$

which is exactly the relation we want to have. □

22 Properties of Harmonic Functions

22.1 Elliptic regularity

Recall that if we have the Laplace equation

$$-\Delta u = f \qquad \text{in } \mathbb{R}^n,$$

then we have the fundamental solution

$$K(x) = \begin{cases} \frac{c_n}{|x|^{2-n}} = \frac{c_n}{|x|^{2-n|}} & n \geq 3 \\ \frac{1}{2\pi} \ln |x| & n = 2, \end{cases}$$

and we can get a solution $u = K * f$. However, there are a number of questions we have not answered, such as uniqueness of solutions.

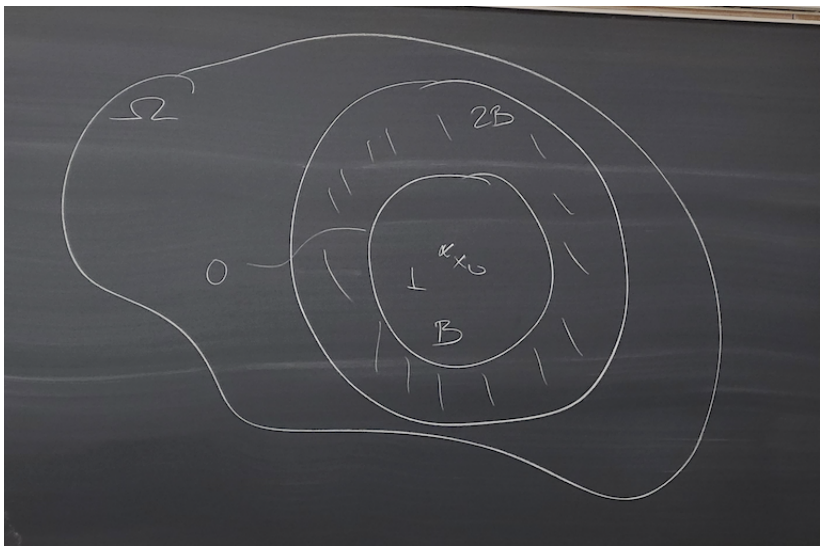
Definition 22.1. A function u such that $-\Delta u = 0$ is called **harmonic**.

Theorem 22.1 (Elliptic regularity)**.** *Harmonic functions are smooth.*

That is, if we have a local solution $u \in \mathcal{D}'$, we want to show that $u \in C^\infty$. Why should harmonic functions be smooth? This is because the fundamental solution K is smooth away from 0. Let's see how the reasoning goes.

Proof. Let Ω be the domain where u lives. Choose a point $x_0 \in \Omega$, and we want to show that u is smooth around x_0. Draw a ball B around x_0 and a larger ball $2B$ around B. To use the fundamental solution, chop off u by using a cutoff function

$$\chi(x) = \begin{cases} 1 & x \in B \\ \text{smooth} & x \in 2B \setminus B \\ 0 & x \in 2B^c \end{cases}$$



If we let $v = \chi u$, then

$$-\Delta u = \underbrace{-\chi \Delta u}_{=0} - u\Delta\chi - 2\nabla u \cdot \nabla\chi.$$

This gives us the new problem

$$-\Delta = f, \qquad f \in \mathcal{D}', \qquad \operatorname{supp} f \subseteq 2B \setminus B.$$

Then

$$\begin{aligned} v(x) &= (K * f)(x) \\ &= \int K(x-y) f(y)\, dy. \end{aligned}$$

Suppose we want a local solution in, say, $B/2$, where B has radius R. If $x \in B/2$ and $y \in 2B \setminus B$, then $|x - y| \geq r/2$. Now $K(z)$ is smooth where $|z| \geq r/2$, which means this convolution is smooth for $x \in B/2$. □

Remark 22.1. We didn't use much about the Laplace equation itself here. We only used the fact that K is smooth away from 0.

Remark 22.2. This is not all there is to elliptic regularity. K is analytic away from 0, which tells us that u is analytic.

Remark 22.3. More generally, we may want to make statements about what kind of regularity u has if f has a certain degree of regularity. This is what elliptic regularity really is, and this is only the tip of the iceberg.

22.2 The maximum principle

Definition 22.2. A function u such that $-\Delta u \leq 0$ is called **subharmonic**.

Definition 22.3. A function u such that $-\Delta u \geq 0$ is called **superharmonic**.

We will prove results for harmonic functions and claim that they hold for sub and superharmonic functions, as well.

Suppose $-\Delta u = 0$ in Ω. Where is the max/min of u? The first step to answering this question is to look at the **mean value property**.

Theorem 22.2 (Mean value property)**.** *Suppose $-\Delta u = 0$ in $B(x_0, a)$. Then*

$$\begin{aligned} u(x_0) &= \frac{1}{|B|} \int_B u(x)\, dx \\ &= \frac{1}{|\partial B|} \int_{\partial B} u(x) d\sigma, \end{aligned}$$

where σ is surface measure on the sphere ∂B.

Remark 22.4. If we assume u is subharmonic, i.e. $-\nabla u \leq 0$, then we get $\leq$ instead of equalities. The reverse inequality holds for superharmonic functions.

Lemma 22.1 (Green's theorem)**.** *Suppose $u : \Omega \to \mathbb{R}$. Then*

$$\int_\Omega \partial_j u \, dx = \int_{\partial\Omega} u \cdot \nu_j \, d\sigma,$$

where ν_j is the outward pointing normal to $\partial\Omega$. Equivalently,

$$\int \underbrace{\partial_j u_j}_{\operatorname{div} u} \, dx = \int_{\partial\Omega} u \cdot \nu \, d\sigma.$$

Here's how we can use this: Integrating by parts twice in the following integral keeps the sign the same and introducing 2 boundary terms:

$$\int -\Delta u \cdot v \, dx - \int_\Omega u \cdot (-\Delta n) \, dx = \int_{\partial\Omega} \underbrace{\partial_j u \nu_j}_{\frac{\partial u}{\partial v}} \cdot v - u \cdot \underbrace{\nu_j \partial_j v}_{\frac{\partial v}{\partial \nu}} \, d\sigma,$$

where these are normal derivatives. Now let's prove the mean value property:

Proof. Suppose $B = B(0, r)$, and apply Green's theorem with a well-chosen v. Looking at our equation, it would be nice if we could make $v = 0$ on the boundary. So we can try

$$v = K(|x|) - K(r).$$

We get

$$u(0) = c \int_{\partial B} u \, d\sigma.$$

This holds for all harmonic functions. If we set $u = 1$, then we get $c = \frac{1}{|\partial B|}$, so $u = \frac{1}{|\partial B|} \int_{\partial B} u$. □

Corollary 22.1. *If $u(x_0) = \max u$ for $x_0 \in B$, then u is constant in B.*

Remark 22.5. If u is subharmonic, the same holds. But if u is superharmonic, then we need to replace the maximum with the minimum in this property.

Theorem 22.3 (Strong maximal principle)**.** *Suppose $u \in C^2(\Omega) \cap C(\overline{\Omega})$ is harmonic. Then*

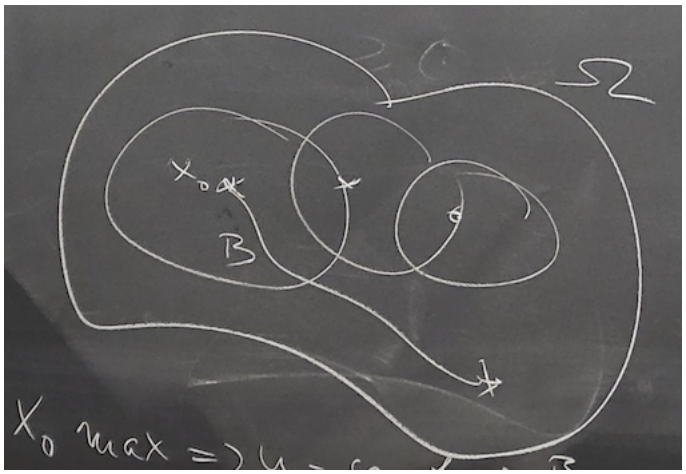
$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} u.$$

Moreover, if $\max u$ is attained inside Ω, then u is constant.

The hypotheses here are much stronger than they need to be.

Remark 22.6. If u is subharmonic, the same holds. But if u is superharmonic, then we need to replace the maximum with the minimum.

Proof. If $\max u$ is only attained on $\partial\Omega$, then we ar edone. What if $\max u$ is attained at $x_0 \in \Omega$? Here is a proof by picture. Put a ball around x_0. By the corollary, u is constant in B. Then the other points in this ball are maximum points, and we can get to any other point via a sequence of balls.



If you want to write down a proof, you can use path-connectedness, or you can use an argument like this: Let $A = \{x \in \Omega : u(x) = u(x_0)\}$. Since u is continuous, A is closed. But the corollary says that if $x_0 \in A$, then $B(X_0, r) \subseteq A$. So A is open. Thus, $A \subseteq \Omega$ is open and closed, and if Ω is connected, we get $A = \Omega$. □

The maximal principle is much more general than the proof we have given here. Here is a restatement of this property:

Corollary 22.2 (Comparison principle)**.** *Let u be subharmonic, i.e., $-\Delta u \leq 0$, and let v be subharmonic, i.e., $-\Delta v \geq 0$. If $u \leq v$ on $\partial\Omega$, then $u \leq v$ in $\overline{\partial}\Omega$.*

Proof. Apply the maximal principle to $u - v$. □

This comparison principle is the correct statement for nonlinear elliptic stuff and also for the Hamilton-Jacobi equations. There is a simpler proof of the maximum principle without the use of the fundamental solution where we drop the strong part.

Proof. Suppose first that $-\Delta u < 0$. Let x_0 be a maximum point inside Ω. Then $\nabla u(x_0) = 0$, and $Hu(x_0) \prec 0$, where $H = \frac{\partial^2 u}{\partial x_i \partial x_j}$ is the Hessian matrix. Observe that

$$\Delta u = \sum_j \partial_j \partial_j u = \operatorname{tr} Hu \leq 0.$$

Then $\Delta u(x_0) \leq 0$, so $-\Delta u(x_0) \geq 0$. But this contradicts our assumption that $-\Delta u < 0$.

Now if $-\Delta u \leq 0$, then we penalize u by replacing u by $u_\varepsilon = u + \varepsilon x^2$. Then

$$-\Delta u_\varepsilon = -\Delta u - 2u\varepsilon < 0.$$

This tells us that

$$\max_{\overline{\Omega}} u_\varepsilon = \max_{\partial\Omega} u_\varepsilon.$$

If we let $\varepsilon \to 0$, both sides converge uniformly to $\max_{\overline{\Omega}} u$ and $\max_{\partial\Omega} u$, respectively. □

22.3 Liouville's theorem

We have been looking at harmonic functions in a domain Ω. What if we are looking at harmonic functions in all of $\mathbb{R}^n$? If you allow exponential growth, then the sky is the limit as to what you can get. But what if we only want polynomial growth. Further yet, what if u is bounded?

Theorem 22.4 (Liouville)**.** *Let u be harmonic in $\mathbb{R}^n$. If u is bounded, then u is constant.*

Proof. If u is harmonic, so are its derivatives. Then

$$\begin{aligned} u(x_0) &\overset{\text{MVP}}{=} \fint_\Omega \partial_j u(x)\, dx \\ &= \frac{1}{|B_R|} \int_{\partial B_R} u \cdot \nu_j \, d\sigma(x). \end{aligned}$$

If $|u| \leq M$, we can estimate this by

$$\begin{aligned} |\partial_j u(x_0)| &\leq \underbrace{\frac{1}{B_R}}_{R^n} M \underbrace{|\partial B_R|}_{R^{n-1}} \\ &\lesssim \frac{M}{R} \\ &\xrightarrow{R\to\infty} 0. \end{aligned}$$

So $\nabla u(x_0) = 0$, which means that u is constant. □

Remark 22.7. If u is temperate, then $\widehat{u}|\xi|^2 = 0$, so $\widehat{u}$ is supported at 0. Then $\widehat{u} = \sum_\alpha c_\alpha \partial_0^{(\alpha)}$, which implies that u is a polynomial. Thus all temperate harmonic functions are polynomials. This also serves as a proof of Liouville's theorem, since the only bounded polynomials are constant.

22.4 Boundary value problems

Let $\Omega \subseteq \mathbb{R}^n$, and suppose that

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = g & \text{on } \partial\Omega. \end{cases}$$

This give us uniqueness: Suppose u_1, u_2 are solutions. If $u_1 - u_2 = v$, then v is harmonic. The maximum and minimum principles give

$$\max_{\Omega} v \leq \max_{\partial\Omega} v = 0,$$

$$\min_{\Omega} v \geq \min_{\partial\Omega} v = 0.$$

So $v = 0$.

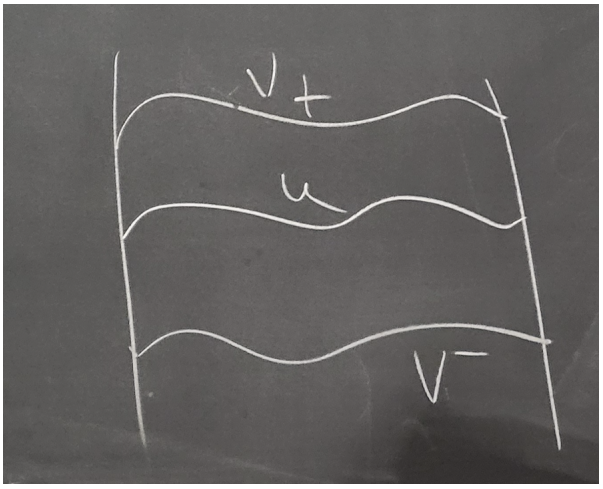
There is also a proof of existence using hte maximum principle. Consider a subsolution v^- satisfying

$$\begin{cases} -\Delta v^- \leq f \\ v \leq g \end{cases}$$

and a supersolution satisfying

$$\begin{cases} -\Delta v^+ \geq f \\ v \geq g \end{cases}$$

The maximum principle $v^* \geq v^-$. Taking the maximum over all supersolutions and subsolutions gives the largest subsolution and the smallest supersolution.



This is called **Perron's method**. We can also find a fundamental solution in Ω, called a **Green function**.

23 Boundary Value Problems for the Laplace Equation

23.1 The Dirichlet and Neumann problems

Last time, we were looking at the Laplace equation

$$-\Delta u = f \qquad \text{in } \mathbb{R}^n.$$

We saw a few ways to look at this:

- via the fundamental solution. This led to elliptic regularity.
- via the maximum principle. This gave us a way to prove uniqueness of solutions.
- via energy estimates. This is what we will discuss today.

When we look at the Laplace equation, we need some boundary behavior¿ The **Dirichlet problem** is to solve the Laplace equation with the following boundary condition.

$$\begin{cases} -\Delta u = f & \text{in } \Omega \subseteq \mathbb{R}^n \\ u = g & \text{on } \partial\Omega \end{cases}$$

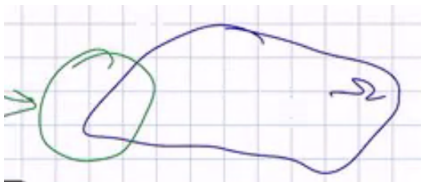
Alternatively, we can look at the **Neumann problem** with a boundary condition on the normal derivative of the solution.

$$\begin{cases} -\Delta u = f & \text{in } \Omega \subseteq \mathbb{R}^n \\ \frac{\partial u}{\partial \nu} = g & \text{on } \partial\Omega \end{cases}$$

Can we impose both the Dirichlet and Neumann boundary conditions? The answer is not always. The equation

$$\begin{cases} -\Delta u = f & \text{in } \Omega \subseteq \mathbb{R}^n \\ u = g & \text{on } \partial\Omega \\ \frac{\partial u}{\partial \nu} = g & \text{on } \partial\Omega \end{cases}$$

is an **overdetermined problem**. It makes sense to consider this locally.



This local problem will in general have uniqueness but not neccesarily existence. This leads to a type of problem called a **unique continuation problem**.

23.2 Uniqueness concerns for the Dirichlet and Neumann problems

Proposition 23.1 (Uniqueness for the Dirichlet problem)**.** *The solution to the Dirichlet problem*

$$\begin{cases} -\Delta u = f & \text{in } \Omega \subseteq \mathbb{R}^n \\ u = g & \text{on } \partial\Omega \end{cases}$$

is unique.

Proof. Suppose u_1, u_2 are solutions. Then $v = u_1 - u_2$ solves

$$\begin{cases} -\Delta v = 0 & \text{in } \Omega \subseteq \mathbb{R}^n \\ v = 0 & \text{on } \partial\Omega. \end{cases}$$

We want to show that $v = 0$. We have

$$\begin{aligned} 0 &= \int -\Delta v \cdot v \, dx \\ &= \int -\partial_j \partial_j v \cdot v \, dx \end{aligned}$$

Green's theorem gives

$$= \int \partial_j v \dot{\partial}_j v - \int_{\partial\Omega} \frac{\partial v}{\partial \nu} \cdot \underbrace{v}_{=0} \, d\sigma.$$

So we get

$$0 = \int_\Omega |\nabla v|^2 \, dx,$$

which tells us that $\nabla v = 0$ in Ω. Thus, v is constant, and the boundary condition tells us that $v = 0$. □

Remark 23.1. What about uniqueness of the Neumann problem?

$$\begin{cases} -\Delta u = f & \text{in } \Omega \subseteq \mathbb{R}^n \\ \frac{\partial u}{\partial \nu} = g & \text{on } \partial\Omega \end{cases}$$

By the same computation, we still get that

$$\int_\Omega |\nabla u|^2 = 0,$$

which tells us that u is constant. The boundary condition is satisfied by any constant, however. So solutions are unique modulo constants.

23.3 Existence using energy type estimates

If $f : \mathbb{R} \to \mathbb{R}$, then a minimum point x_0 for f must have $f'(x_0) = 0$. We can do the reverse. If we have an equation for a function, we can write it as the derivative of another function and interpret our equation as finding the minimizers (or critical points) for this function.

Looking at functions $u : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, associate the functional

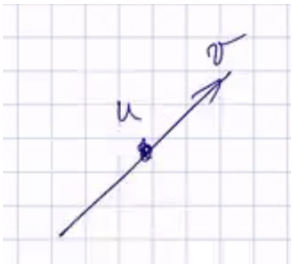
$$\mathcal{L}(u) = \int_\Omega |\nabla u|^2 - f \cdot u \, dx.$$

We will call this the **Lagrangian** of the problem. Our goal is to minimize $\mathcal{L}(u)$ over a good class of functions u; we can assume some nice regularity and our boundary condition. Let $\mathcal{A} = \{u : \Omega \to \mathbb{R} \mid u \in C^2, u = 0 \text{ on } \partial\Omega\}$. So we want

$$\min_{u \in \mathcal{A}} \mathcal{L}(u).$$

Does a minimum exist? We will not answer this today, but observe that $\mathcal{L}$ is *strictly convex* because it is the sum of a positive quadratic form and a linear term. If we have a minimum, then by strict convexity, the minimum will be unique.

We may also ask: What equation does a minimum satisfy? Suppose u is a minimum. Take $v \in \mathcal{D}(\Omega)$, and set $u_h = u + hv$.



Look at $\mathcal{L}(u_h)$ as a function of h. This has a minimum at $h = 0$, which tells us that

$$\frac{d}{dh}\mathcal{L}(u_h) = 0 \quad \text{at } h = 0.$$

Write

$$\mathcal{L}(u_h) = \int |\nabla(u + hv)|^2 - f \cdot (u + hv) \, dx,$$

so

$$\frac{\partial}{\partial h}\mathcal{L}(u_h)|_{h=0} = \int \nabla u \cdot \nabla v - f \cdot v \, dx.$$

Hence,

$$0 = \int \nabla u \cdot \nabla v - fc\dot{v} \, dx$$

for all $v \in \mathcal{D}(\Omega)$. Integration by parts gives us

$$\begin{aligned} &= \int_\Omega -\Delta u \cdot v - fv \, dx \\ &= \int_\Omega v(-\Delta u - f) \, dx. \end{aligned}$$

So we get

$$-\Delta = f \quad \text{in } \Omega.$$

And we can append our favorite boundary condition.

Remark 23.2. The regularity condition $u \in C^2$ is not the correct condition to use. Really, we want to use **Sobolev spaces**, which we have not discussed yet.

23.4 Green's functions for domains with boundary

Circle back to the fundamental solution and try to use it in a domain with boundary. We will look at how this doesn't work and how it can be fixed. In $\mathbb{R}^n$, we have the formal computation

$$\int -\Delta u \cdot K(x - x_0) \, dx = \int u \cdot -\underbrace{\Delta K}_{\delta_{x_0}}(x - x_0) \, dx.$$

If $-\Delta u = f$, then

$$\int f \cdot K(x - x_0) \, dx = u(x_0).$$

What about a domain with boundary?

$$\int_\Omega -\Delta u \cdot K(x - x_0) \, dx = \int u \cdot -K(x - x_0) \, dx + \int_{\partial\Omega} -\frac{\partial u}{\partial \nu} \cdot K(x - x_0) + u \cdot \frac{\partial}{\partial \nu} K(x - x_0) \, d\sigma.$$

If u solves the Dirichlet problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega \subseteq \mathbb{R}^n \\ u = g & \text{on } \partial\Omega, \end{cases}$$

then

$$u(x_0) = \int_\Omega f(x) K(x - x_0) \, dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu} K(x - x_0) - g \cdot \frac{\partial K}{\partial \nu}(x - x_0) \, d\sigma,$$

but we do not know what $\frac{\partial u}{\partial \nu}$ is. We do not have this information, and recall that if we do, then we have an overdetermined problem.

How do we fix this? Perhaps the fundamental solution is not the object we want to be looking at. Replace $K(x-x_0)$ by $G(x,x_0)$ to get

$$\int_\Omega -\Delta u \cdot G(x-x_0)\,dx = \int u \cdot -G(x-x_0)\,dx + \int_{\partial\Omega} -\frac{\partial u}{\partial \nu}\cdot G(x-x_0) + u\cdot \frac{\partial}{\partial \nu}K(x-x_0)\,d\sigma.$$

We would like the properties

$$\begin{cases} -\Delta_x G(x,x_0) = \delta_{x_0}, \\ G(x,x_0) = 0 & x\in\partial\Omega. \end{cases}$$

If we have these properties, then

$$u(x_0) = \int_\Omega f\cdot G(x,x_0)\,dx + \int_{\partial\Omega} g\frac{\partial}{\partial\nu}G(x,x_0)\,d\sigma.$$

If we had such a function G, then we could solve the Dirichlet problem. Call this function G the **Green function**[22] in Ω.

Remark 23.3. $G = G(x,x_0)$, not $G(x-x_0)$ because translation invariance is broken by our domain. If you translate the domain with boundary, you will not get the same domain.

To find such a function G, we would try

$$G(x,x_0) = K(x-x_0) + \psi(x,x_0)$$

and look for an equation for ψ. We need

$$\begin{cases} -\Delta_x \psi(x,x_0) = 0 & \text{in } \Omega \\ \psi(x,x_0) = -K(x-x_0) & x\in\partial\Omega. \end{cases}$$

$K(x-x_0)$ is smooth for $x\in\partial\Omega$ and $x_0\in\Omega$, so by elliptic regularity, ψ should be smooth. ψ can be found by solving a Dirichlet problem.

Remark 23.4. You may think this is leading us in a circle, but this is not the case: Here, we are solving a Dirichlet problem with a very special boundary data, $K(x-x_0)$. This may make the Green's function easier to find than solving the original equation otherwise.

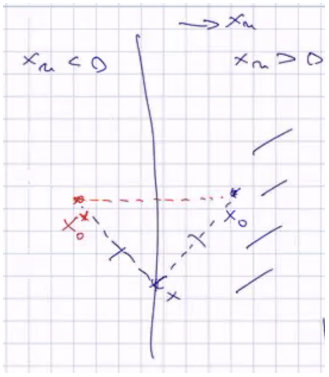
Remark 23.5. The Green function is symmetric:

$$G(x,y) = G(y,x).$$

Let's calculate some Green's functions.

[22] If we are solving the Neumann problem, we may get a different Green function

Example 23.1 (half-plane)**.** Let $\Omega = \{x_n > 0\}$. The idea is to calculate G using symmetries. Here is the picture we should have in mind:



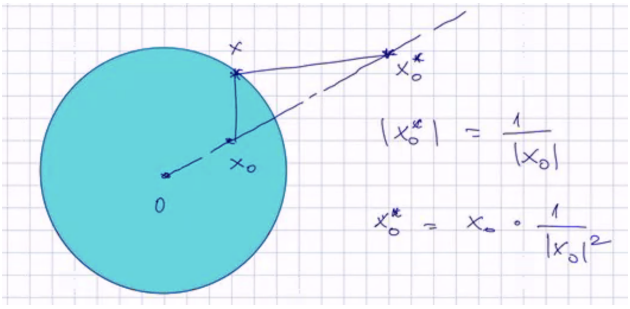
We have

$$G(x, x_0) = K(x - x_0) + \psi(x, x_0).$$

If we were to use x_0^*, then $K(x - x_0^*)$ is harmonic in Ω. Now also observe that for $x \in \partial\Omega$, $|x - x_0| = |x - x_0^*|$ for $x \in \partial\Omega$. Thus, the radial symmetry of K gives $K(x - x_0) = K(x - x_0^*)$. This implies that we can choose

$$G(x, x_0) = K(x - x_0) - K(x - x_0^*).$$

Example 23.2 (unit ball)**.** Let $\Omega = B(0, 1)$. Here, we can try repeating the same argument but with inversion about the boundary of the circle:



If we have point $x \in \partial\Omega$, then

$$|x - x_0| = |x - x_0^*| \cdot |x_0|.$$

So if we are in $\mathbb{R}^n$ for $n \geq 3$,

$$G(x, x_0) = K(x - x_0) - |x - 0|^{2-n} K(x - x_0^*)$$

The proportionality constant comes from the fact that the first term is like $|x - x_0|^{2-n}$, while the second term is like $|x - x_0^*|^{2-n}$.

These kinds of computations are available only in very specific domains, so the existence of Green's functions is more of a qualitative question than a computational one.

24 Boundary Value Problems for the Heat Equation

24.1 Properties of the heat equation

Consider the heat equation in $\mathbb{R}^+ \times \mathbb{R}^n$.

$$\begin{cases} (\partial_t - \Delta)u = f \\ u(0) = u_0 \end{cases}$$

We have already seen how to derive a solution via the fundamental solution:

$$u = f *_{x,t} K(t) + u_0 *_x K(t), \qquad K(t) = \frac{1}{(4\pi t)^{n/2}} e^{-x^2/(4t)} \mathbb{1}_{\{t \geq 0\}}.$$

This is the unique solution going forward in time which is a temperate distribution.

Here are some key properties for the homogeneous equation given by this fundamental solution: Consider the heat equation in $\mathbb{R}^+ \times \mathbb{R}^n$.

$$\begin{cases} (\partial_t - \Delta)u = 0 \\ u(0) = u_0 \end{cases}$$

- Infinite speed of propagation: Even if u_0 has compact support, the solution u immediately spreads to all of $\mathbb{R}^n$.
- Instant regularization:
$$u(t) = K(t) * u_0,$$
where $K(t)$ is smooth for $t > 0$. So u is smooth for $t > 0$.
- The fundamental solution has Gaussian decay at ∞: This means that any initial data u_0 with $|u_0| \leq e^{cx^2}$ will generate a local in time solution

24.2 The mean value property and the maximum principle

Now let's look at the heat equation in a domain $\Omega \subseteq \mathbb{R}^n$.

$$\begin{cases} (\partial_t - \Delta)u = f & \text{in } \Omega \times \mathbb{R}^+ \\ u(t=0) = u_0 & \text{in } \Omega \\ u(t,x) = g & \text{on } \partial\Omega \times \mathbb{R}^+ \end{cases}$$

The third equation is a **Dirichlet boundary condition**. We could replace it with a **Neumann boundary condition**

$$\frac{\partial u}{\partial \nu}(t,x) = g \qquad \text{on } \partial\Omega \times \mathbb{R}^+.$$

As with the Laplace equation, we use either one boundary condition or the other but not both.

Here are several ways to approach this:

- Via a maximum principle.
- Via energy estimates.
- Using Green's functions.

We first discuss the maximum principle. First, is there a mean value property for the heat equation? We would like to write something like

$$u(t_0, x_0) = \frac{1}{|D|} \int_D u(t, x)\, dx.$$

for some D. For the Laplace equation, we used a ball for D, but this should not be the case for the heat equation; unlike for the Laplace equation, balls are not level sets of the fundamental solution. We may also ask if we need any weights for the maximum principle.

Step 1: Green's theorem for the heat equation: Let u, v be such that v has compact support. Then

$$\iint (\partial_t - \Delta)u \cdot v\, dx\, dt = \iint (-\partial_t - \Delta)v \cdot u\, dx\, dt.$$

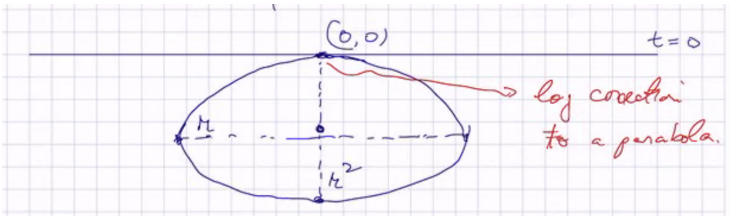
If we want to get $u(0,0)$ out of the right hand side, then we would need $(-\partial_t - \Delta)v = \delta_{(0,0)}$. Here, $-\partial_t - \Delta$ is the **adjoint heat operator**, which is a "backward heat operator" and gives a backward heat equation with a fundamental solution

$$K^{\text{back}}(x,t) = -\frac{1}{(4\pi|t|)^{n/2}} e^{x^2/(4t)} \mathbb{1}_{\{t \le 0\}}.$$

Define the parabolic balls

$$D_r(0,0) = \{|K^{\text{back}}(x,t)| \le r^{-n}\}.$$

What do these sets look like? If $x = 0$, then $K \asymp t^{-n/2}$, and $t^{-n/2} \ge r^{-n}$ iff $t \le r^2$. To figure out the sideways boundaries of these regions, take $t \approx \frac{1}{2}r^2$. Now change x so that $e^{x^2/(4t)} \gtrsim 1$. Then $|x| \le \sqrt{t} \approx r$. This looks like an ellipse, but near $(0,0)$, there is a logarithmic coorection to a parabola.



Our goal is to show that

$$u(0,0) = \int_{D_r(0,0)} \omega(t,x)u(t,x)\,dx$$

for some suitable positive weight ω (we want positive so we can think of this as an average). Look at our Green's theorem in $D_r(0,0)$, which gives boundary terms:

$$\iint_{D_r(0,0)} (\partial_t - \Delta)u \cdot v\,dx\,dt = \iint_{D_r(0,0)} (-\partial_t - \Delta)v \cdot u\,dx\,dt + \int_{\partial D_r(0,0)} \nu_t \cdot uv - \frac{\partial u}{\partial \nu} \cdot v + u \cdot \frac{\partial v}{\partial \nu}\,d\sigma.$$

For $v = K^{\text{back}}(t,x)$, this does not work because we get boundary terms. Instead, we can try

$$v = K^{\text{back}}(t,x) + r^{-n},$$

which makes $v = 0$ on $\partial D_r(0,0)$. This makes the first two boundary terms equal 0, but we would also like to make sure that $\frac{\partial v}{\partial \nu} = 0$ on $\partial D_r(0,0)$. This is the same as saying that $\nabla v = 0$ on ∂D_r. The way we can alter our fundamental solution to take advantage of this is

$$v = K^{\text{back}}(t,x) + r^{-n} + c\ln(-K^{\text{back}} \cdot r^n),$$

where c is chosen so that $\nabla v = 0$ on $\partial D_r(0,0)$. This choice gives us

$$\begin{aligned}\nabla v &= \nabla K + c\frac{\nabla K}{K} \\ &= \nabla K\left(1 + \frac{c}{K}\right),\end{aligned}$$

and since $K = -r^n$ on the boundary, we can pick $c = r^n$.

If $(\partial_t - \Delta)u = 0$, then we get

$$\iint D_t(-\partial - \Delta)v \cdot u\,dx\,dt = 0.$$

We can calculate

$$\begin{aligned}(-\partial_t - \Delta)v &= \delta_{(0,0)} + c(-\partial_t - \Delta)\ln(-r^n K^{\text{back}}) \\ &= \delta_{(0,0)} - c\frac{\partial_t K^{\text{back}}}{K^{\text{back}}} - c\nabla \cdot \frac{\nabla K^{\text{back}}}{K^{\text{back}}} \\ &= \delta_{(0,0)} - c\underbrace{\frac{(\partial_t - \Delta)K^{\text{back}}}{K^{\text{back}}}}_{=0} + c\frac{(\nabla K^{\text{back}})^2}{(K^{\text{back}})^2}\end{aligned}$$

$$= \delta_{(0,0)} + c(\nabla \ln K^{\text{back}})^2,$$

where this is a spatial gradient.

$$= \delta_{(0,0)} - r^{-n}\frac{x^2}{4t^2}.$$

We get:

Theorem 24.1 (Mean value property)**.** *If* $(\partial_t - \Delta)u = 0$ *in* $\Omega \times [0,T]$,

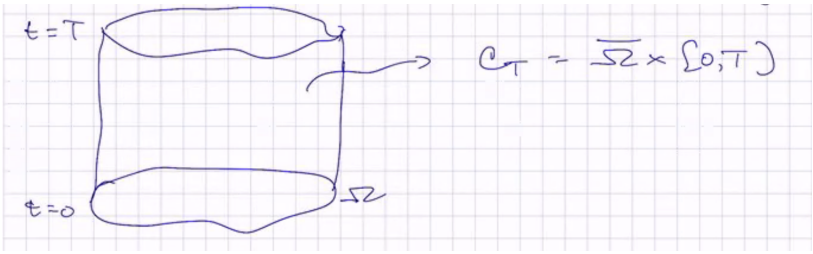
$$u(0,0) = r^{-n}\int_{D_r(0,0)} \frac{x^2}{4t^2} u(t,x)\,dx\,dt$$

Remark 24.1. How do we know this is an average? This holds for all solutions to the heat equation, so plug in a constant. This gives

$$r^{-n}\int_{D_r(0,0)} \frac{x^2}{4t}\,dx\,dt = 1.$$

So this is indeed a weighted average.

For our maximum principle, what is the boundary of our region $C_T = \overline{\Omega} \times [0,T]$?



If you consider causality, the $t = T$ boundary is determined by the rest, so it should not be considered. Write $\partial C_T = \overline{\Omega} \times \{0\} \cup \partial\Omega \times [0,T]$. The first part is the bottom, and the second part is the **lateral boundary**. Together, they make up the **parabolic boundary** of C_T.

Theorem 24.2 (Strong maximum principle)**.** *If* $(\partial_t - \Delta)u = 0$ *in* $\Omega \times [0,T]$*, then*

$$\max_{C_T} u = \max_{\partial C_T} u.$$

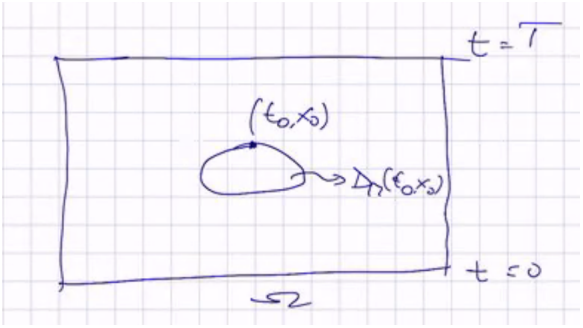
Further if $u(t_0,x_0) = \max u$ *for some* (t_0,x_0) *inside, then* u *is constant for* $t \le t_0$.

Proof. Take (t_0,x_0) to be a maximum inside. Then the mean value property gives

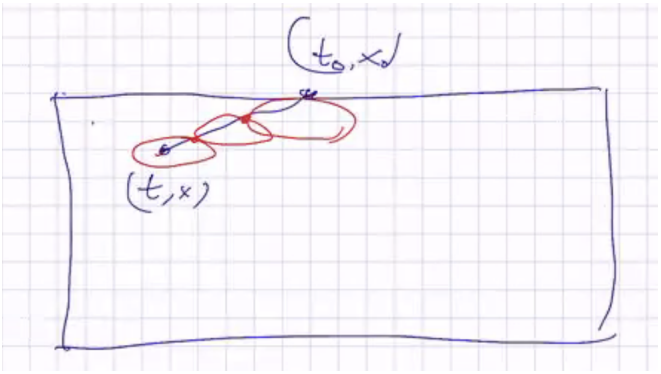
$$\max u = u(t_0,x_0)$$

$$
\begin{aligned}
&= r^{-n} \int \frac{(x-x_0)^2}{(t-t_-0)^2} u(t,x)\, dx\, dt \\
&\leq r^{-n} \int \frac{(x-x_0)^2}{(t-t_-0)^2} \max u\, dx\, dt \\
&= \max u.
\end{aligned}
$$

Equality must hold, so $u = \max u$ in $D_r(t_0, x_0)$.



How do we get the whole region $\{t \leq t_0\}$? Here is a picture:



□

Remark 24.2. Just like with the Laplace equation, we can talk about *subsolutions*

$$(\partial_t - \Delta)u \leq 0$$

and *supersolutions*

$$(\partial_t - \Delta)u \geq 0.$$

Using the mean value property with inequalities gives a maximum principle for subsolutions and a minimum principle for super solutions.

Theorem 24.3 (Comparison principle)**.** *Let u^- be a subsolution and u^+ be a supersolution for the heat equation. If $u^- \leq u^+$ on ∂C_T, then $u^- \leq u^+$ in C_T.*

Proof. $u^- - u^+$ is a supersolution. □

Here is a corollary of the maximum principle.

Corollary 24.1. *The solution to the Dirichlet problem is unique.*

Proof. Subtract two solutions to get $u = u_1 - u_2$. If

$$\begin{cases} (\partial_t - \Delta)u = 0 \\ u(0) = 0 \\ u(\partial\Omega) = 0, \end{cases}$$

then the maximum principle tells us that $u = 0$. □

24.3 Energy estimates

Consider the homogeneous Dirichlet problem

$$\begin{cases} (\partial_t - \Delta)u = 0 \quad \text{in } \Omega \times [0, T) \\ u(0) = u_0 \\ u(\partial\Omega) = 0, \end{cases}$$

and let

$$E(u(t)) = \int |u(t, x)|^2 \, dx.$$

Then we can compute

$$\begin{aligned} \frac{\partial}{\partial t} E(u(t)) &= 2 \int u \cdot u_t \, dx \\ &= 2 \int u \cdot \Delta u \, dx \\ &= -2 \int |\nabla u|^2 \, dx \\ &\leq 0, \end{aligned}$$

which tells us that E is nonincreasing in time $E(t) \leq E(0)$. So if $u_0 = 0$, then $E(t) = 0$, which gives $u(t) = 0$.

We can also look at the relation

$$\|u(0)\|_{L^2}^2 = \|u(T)\|_{L^2}^2 + \int_0^t |\nabla u|_{L^2}^2 \, dx.$$

If we start with $u(0) \in L^2$, we get $\nabla u(t) \in L^2$ for a.e. t. We can think of this as a **parabolic regularizing effect**.

25 Initial Value Problems and Energy Estimates for the Wave Equation

25.1 Initial value problems for the wave equation

Today, we will be looking at the wave equation

$$\begin{cases} \Box u = f & \text{in } \mathbb{M}^{n+1} = \mathbb{R} \times \mathbb{R}^n \\ u(0) = u_0 \\ \partial_t u(0) = u_1, \end{cases}$$

where

$$\Box = \partial_t^2 - \Delta_x = -m^{\alpha\beta}\partial_\alpha \partial_\beta,$$

$$m = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

We have seen that the fundamental solution (forward in time) is

$$K(t,x) = \begin{cases} \frac{1}{2}\mathbb{1}_{\{t>|x|\}} & n = 1 \\ c_n (t^2 - x^2)_+^{(1-n)/2} & n \geq 2 \text{ even} \\ c_n \delta_{t^2-x^2}^{(\frac{n-1}{2})} & n \geq 2 \text{ odd} \end{cases}$$

The solution for the inhomogeneous problem is $u = K * f$ (as if the Cauchy data equals 0 at $-\infty$). The solution for the homogeneous problem ($f = 0, u_0, u_1 \neq 0$) is a bit more tricky. Let

$$\widetilde{u} = \begin{cases} u & t > 0 \\ 0 & t < 0. \end{cases}$$

Let's find an equation for $\widetilde{u}$.

$$\Delta \widetilde{u} = \begin{cases} \Delta u & y \geq 0 \\ 0 & t < 0, \end{cases}$$

$$\partial_t \widetilde{u} = \begin{cases} \partial_t u & y > 0 \\ 0 & t < 0, \end{cases} + \delta_{t=0} \cdot u_0.$$

The second time derivative is then

$$\partial_t \widetilde{u} = \begin{cases} \partial_t^2 u & y > 0 \\ 0 & t < 0, \end{cases} + \delta_{t=0} \cdot u_1 + \delta'_{t=0} \cdot u_0.$$

This gives us

$$\Box \widetilde{u} = \delta_{t=0} u_1 + \delta'_{t=0} u_0,$$

which implies that

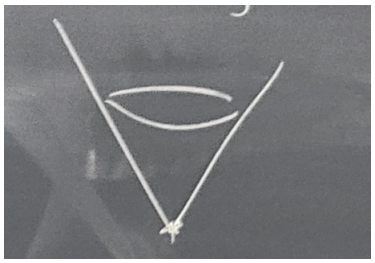
$$\widetilde{u} = K *_{t,x} (\delta_{t=0} u_1 + \delta'_{t=0} u_0).$$

Taking the convolution first in t gives

$$\widetilde{u}(t) = K(t) *_x u_1 + \partial_t K(t) *_x u_0.$$

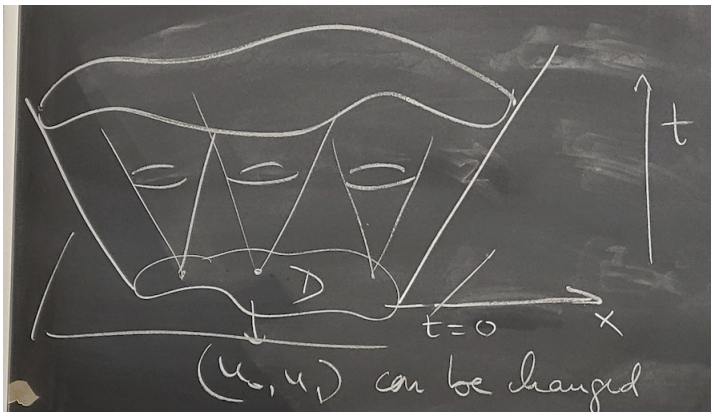
Here are some properties of the wave equation:

- Finite speed of propagation: The solution only exists inside the positive cone.



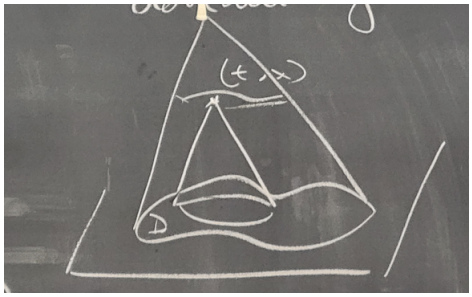
- Huygens principle: When $n \geq 3$ is odd, the fundamental solution is supported exactly on the cone.

Suppose now that we have some region with initial data (u_0, u_1) which can be changed. Where does the solution change? At each point, we have an upward cone, and we take the union of these cones.

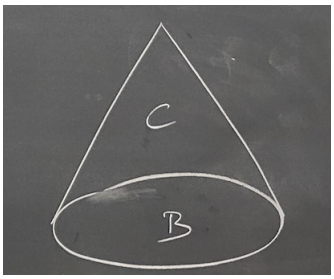


The **domain of influence**.is $\Omega = \cup_{x \in D}\{(0, x) + K\}$, where K is the forward cone. Now suppose we only have initial data (u_0, u_1) in the domain D. Where can we find the solution?

If we look at a point (t, x), then $u(t, x)$ depends on u_0, u_1 in $B(x, |t|)$. The value $u(x, t)$ is uniquely determined if $B(x, |t|) \subseteq D$. The union, $\{(t, x) : B(x, |t|) \subseteq D\}$ is called the **domain of uniqueness** for D.

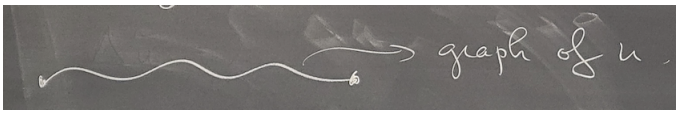


Example 25.1. When the base domain B is a ball, then the domain of uniqueness C is exactly a cone:



25.2 Energy estimates for the wave equation

Here's how energy estimates work for the wave equation. When we say energy, we want to think a quantity which is conserved. Suppose we have a vibrating string.



We can think of the energy as potential energy P, expressed in terms of "how extended is the string." This can be measured by some average of the slope of the string:

$$P = \int |\partial_x u|^2 \, dx.$$

The second part of the energy should be the kinertic energy, which measures the velocity of the string:

$$\int |\partial_t u|^2 \, dx.$$

If we were physicists, we would have constants in front, but we are mathematicians, so we will set some constants equal to 1. We can write the total energy as

$$E(u(t)) = \frac{1}{2}\int |\partial_t u|^2 + |\nabla_x u|^2 \, dx.$$

Theorem 25.1. *If $\Box u = 0$, then $E(u(t))$ is constant.*

Corollary 25.1. *The wave equation has at most 1 solution.*

Proof. The naive proof of this theorem is to take

$$\begin{aligned}\frac{d}{dt}E &= \int \partial_t u \partial_t^2 u + \sum_{j=1}^{n} \partial_j u \cdot \partial_t \partial_j u \, dx \, dt \\ &= \int \partial_t u \sum_{j=1}^{n} \partial_j \partial_j u + \sum_{j=1}^{n} \partial_j u \cdot \partial_t \partial_j u \, dx \, dt \\ &= 0\end{aligned}$$

by Green's theorem, assuming that $u = 0$ at ∞. □

Why should we improve on this? We have seen that "conservation laws" imply features of our problem. If we have

$$\partial_t \underbrace{u}_{\text{density}} + \partial_x \underbrace{F(u)}_{\text{flux}} = 0,$$

we can integrate to get

$$\int \partial_t u \, dx = -\int \partial_x F(u) \, dx,$$

which tells us that

$$\partial_t \int u = 0.$$

For the wave equation, we have the **energy density**

$$e(t,x) = \frac{1}{2}|\partial_t u|^2 + |\nabla_x u|^2,$$

so that

$$E = \int e.$$

Note that this doesn't go the other way around; there may be many densities that integrate to the same total energy E. We can look at

$$\partial_t e(t,x) = \partial_t u \cdot \partial_t^2 u + \partial_j u \cdot \partial_t \partial_j u$$

$$= \partial_t u \cdot \Box u + \partial_t u \cdot \partial_j^2 u + \partial_j u \partial_t \partial_j u$$
$$= \partial_j (\underbrace{\partial_t u \cdot \partial_j u}_{\text{energy flux}}) + \partial_t u \cdot \Box u$$

This leads us to another proof of the energy estimates for the wave equation:

Proof. Start with $\Box u = 0$, and get $\Box u \cdot \partial u = 0$. Then integrate by parts. □

Let's see what happens when we take

$$\begin{aligned}\Box u \cdot \partial_k u &= (\partial_t^2 u - \partial_j \partial_j u) \cdot \partial_k u \\ &= \partial_t(\partial_t u \cdot \partial_k u) - \partial_t u \cdot \partial_t \partial_k u - \partial_j(\partial_j u \cdot \partial_k u) + \partial_j u \cdot \partial_j \partial_k u\end{aligned}$$

We can think of the first and third terms as derivatives.

$$= \partial_t(\underbrace{\partial_t u \cdot \partial_k u}_{\text{density}}) - \underbrace{\frac{1}{2}\partial_k(\partial_t u)^2 - \partial_j(\partial_j u \partial_k u) + \frac{1}{2}\partial_k(\partial_j u)^2}_{\text{divergence of a flux}}.$$

We get a new, conserved quantity, the **momentum**

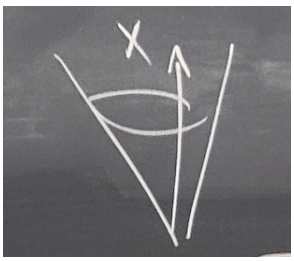
$$P_k = \int \partial_t u \cdot \partial_k u \, dx.$$

This tells you in what direction the energy is moving around. Conservation says that if the energy is moving around in one direction, it will be moving in that same direction forever.

More generally, consider

$$\Box u \cdot Xu, \quad \text{where} \quad X = \sum X^\alpha \partial_\alpha.$$

This gives a conserved quantity E_X, which is positive definite if the vector field X is **forward time-like**.



Remark 25.1. We can put the energy and the momentum into one conserved quantity, called the **energy-momentum tensor**,

$$T^{\alpha\beta}(u) = \partial^\alpha u \partial^\beta u - \frac{1}{2} m^{\alpha\beta} \partial^\gamma u \partial_j u,$$

where

$$\partial^\alpha = m^{\alpha\beta}\partial_\beta u.$$

For the wave equation, this looks like

$$\partial^0 u = -\partial_0 u, \qquad \partial^j u = \partial_j u.$$

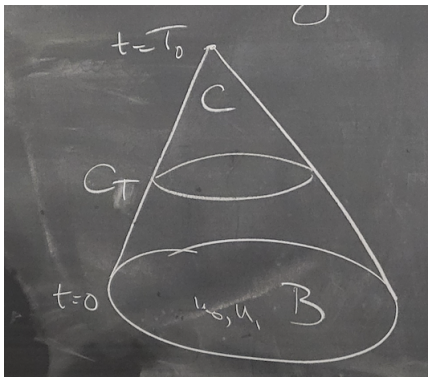
This is a divergence free tensor:

$$\partial_\alpha T^{\alpha\beta} = 0 \quad \forall\beta \quad \text{if} \quad \Box u = 0.$$

If $\beta = 0$, this is the energy, and if $\beta = j \neq 0$, this is the momentum P_j.

25.3 Finite speed of propagation via energy estimates

The finite speed of propagation is a robust phenomenon. We can show this by providing a proof which does not rely on the fundamental solution and only requires energy estimates. If we have a ball B for our inital data, and a cone C, we want to show that (u_0, u_1) in B uniquely determines the solution in C.



This is the same as saying that if $(u_0, u_1) = (0, 0)$ in B, then $u = 0$ in C. Suppose we want to show that $u = 0$ in the slice C_T of the cone. We saw the following density flux relation for the energy:

$$\partial_t e(t, x) = \partial(\partial_t u \cdot \partial_u).$$

Integrate over $C_{[0,T]}$, the section of the cone up to time T.

$$\int_{C_T} e - \int_{C_0} + \int_{\partial C_{[0,T]}} e \cdot \nu_t - p_j \nu_x = 0$$

Moving the middle term to the right hand side, this tells us that

$$\text{Energy}(t = 0) = \text{Energy}(t = T) + \text{Flux(boundary)}.$$

The former term is the part that is left in the cone, and the latter term is the part that goes out. If the energy at time $t = 0$ is 0, then these two terms must both equal 0.

The remaining objective is to show that the Flux term is nonnegative. What does it mean that the slope of the cone is -1? This means that the outward pointing normal is $\nu = (1, \omega)$ with $|\omega| = 1$. Then

$$\begin{aligned} e \cdot \nu_t - p_j \cdot \nu_j &= \frac{1}{2}(u_t^2 + |\nabla_x u|^2) - \partial_t u \cdot \partial_j \cdot \omega_j \\ &\overset{?}{\geq} 0. \end{aligned}$$

We can use Cauchy-Schwarz twice to say

$$\begin{aligned} |\partial_t u \cdot \partial_j u \cdot \omega_j| &\leq |\partial_t u| \cdot |\partial u| \\ &\leq \frac{1}{2}(|\partial_t u|^2 + |\partial u|^2). \end{aligned}$$