MATH 218 LECTURE NOTES (SPRING 2022)

JOEY ZOU

Contents

Date: June 2, 2022.

Overview

The course will focus on various aspects of linear parabolic and hyperbolic equations, with a particular emphasis on the construction and properties of parametrices and solution operators to such equations. Towards the end of the quarter, I'll plan on giving an introduction to microlocal analysis, with a particular focus on applications to parabolic and hyperbolic problems.

Schedule

(may be subject to change depending on student interest)

- Weeks 1,2: Review of distribution theory and Fourier transform
- Weeks 3-5: Study of parabolic equations, with particular focus on parabolic regularity and the structure of the heat kernel
- Weeks 6,7: Study of hyperbolic equations: linear wave equation and the geometric optics ansatz
- Week 8: Introduction to microlocal analysis and parametrices for differential operators
- Week 9: Construction of parametrices for parabolic operators
- Week 10: Construction of parametrices for hyperbolic operators

1. Lecture 1 (03/29): Distribution Theory: Preliminaries

Distributions are generalizations of functions that work particularly well with differentiation, convolution, Fourier transforms, etc.. We review the theory of distributions in this lecture.

The reference for this section is Hörmander's *The Analysis of Partial Differential Operators I* [Hör90].

1.0. Conventions.

We consider functions (either $\mathbb{R}$-valued or $\mathbb{C}$-valued; usually the difference is not significant) defined on an open subset U of $\mathbb{R}^n$. A *multi-index* is an n-tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ of nonnegative integers, with the corresponding differential operator $\partial^\alpha := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \dots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}$ on U. The sum $|\alpha| = \alpha_1 + \cdots + \alpha_n$ is the *order* of the multi-index α/the differential operator ∂^α. We denote $C^k(U)$ the space of functions f on U which has partial derivatives of order up to k, with $\partial^\alpha f$ continuous on U for all α with $|\alpha| \leq k$. We denote $C^\infty(U) = \cap_k C^k(U)$. When needed, we'll define $C^k(\overline{U})$ and $C^\infty(\overline{U})$ analogously, with the additional requirement that the partial derivatives be *continuous up to the boundary.*

For $U \subset \mathbb{R}^n$ open, we denote

$$C_c^\infty(U) := \{u \in C^\infty(\mathbb{R}^n) \,:\, \operatorname{supp} u \text{ is a compact subset of } U\}.$$

1.1. Definitions and properties.

Definition 1.1. Let $U \subset \mathbb{R}^n$ be open. A *distribution* on U is a linear functional $u : C_c^\infty(U) \to \mathbb{C}$ which is continuous with respect to the topology on $C_c^\infty(U)$ (defined below). The space of all distributions on U is denoted[1] by $\mathcal{D}'(U)$.

For $u \in \mathcal{D}'(U)$ and $\phi \in C_c^\infty(U)$, we'll write the application of u against ϕ as $u(\phi)$ or (u, ϕ) or $(u, \phi)_U$ (when emphasizing the domains on which the distributions live).

The topology on $C_c^\infty(U)$ is a bit involved to describe[2], but the practical interpretation of the topology is as follows: a sequence $\{\phi_n\}$ in $C_c^\infty(U)$ *converges to* $\phi \in C_c^\infty(U)$ if:

- There exists a compact set $K \subset U$ such that[3] $\operatorname{supp} \phi_n \subset K$ for all n, and
- All derivatives of ϕ_n converge uniformly to the corresponding derivative of ϕ on U, i.e. for all multi-indices α we have

$$\sup_{x \in U} |\partial^\alpha \phi_n - \partial^\alpha \phi| \xrightarrow{n \to \infty} 0.$$

---

[1]The notation comes from Laurent Schwartz's work where the space of "test functions", i.e. $C_c^\infty(U)$, was denoted by $\mathcal{D}(U)$.

[2]For the curious: it can be described as an inductive limit topology on $C_c^\infty(K)$ (for all $K \subset U$ compact), with the topology on $C_c^\infty(K)$ given by the seminorms $\phi \mapsto \|\partial^\alpha \phi\|_{L^\infty(K)}$ for all multi-indices α. This is apparently called the "LF topology" (for *l*imit of *F*réchet spaces).

[3]From the convergence requirement below, we see *a posteriori* that $\operatorname{supp} \phi \subset K$ as well.

Thus, we say that a linear functional $u : C_c^\infty(U) \to \mathbb{C}$ is[4] *continuous* if, whenever $\phi_n \to \phi$ with respect to the convergence defined above in $C_c^\infty(U)$, we also have $u(\phi_n) \to u(\phi)$ in $\mathbb{C}$ as well.

Lemma 1.2. *A linear functional $u : C_c^\infty(U) \to \mathbb{C}$ is continuous if and only if, for every compact subset $K \subset U$, there exists $k \in \mathbb{N}_{\geq 0}$ and $C > 0$ such that*

$$|u(\phi)| \leq C \sum_{|\alpha|\leq k} \sup_K |\partial^\alpha \phi| \text{ for all } \phi \in C_c^\infty(U) \text{ with supp } \phi \subset K.$$

Proof. The "if" part is straightforward to verify. For the "only if" part, suppose for some compact $K \subset U$ that no such C and k were to exist to satisfy the above inequality. Then, for every $k \in \mathbb{N}_{\geq 0}$, we could find $\phi \in C_c^\infty(U)$ supported in K such that $u(\phi)$ is arbitrarily large compared to $\sum_{|\alpha|\leq k} \sup_K |\partial^\alpha \phi|$. In particular, for each k we can find ϕ_k supported in K such that $u(\phi_k) = 1$, but $\sum_{|\alpha|\leq k} \sup_K |\partial^\alpha \phi_k| \leq 1/k$. Note that this implies that $\sup_K |\partial^\alpha \phi_k| \leq 1/k$ whenever $k \geq |\alpha|$. In particular, we see that for each fixed α we would have $\sup_U |\partial^\alpha \phi_k - 0| \xrightarrow{k\to\infty} 0$, and hence $\phi_k \to 0$ in the topology of $C_c^\infty(U)$. But then we should have $u(\phi_k) \to u(0) = 0$ as well due to the continuity of u, which contradicts the assumption that $u(\phi_k) = 1$ for each k. □

Thus, we could have alternatively *defined* a distribution u as a linear functional satisfying

$$|u(\phi)| \leq C \sum_{|\alpha|\leq k} \sup_K |\partial^\alpha \phi| \text{ for all } \phi \in C_c^\infty(U) \text{ with supp } \phi \subset K$$

for some constants C and k for each compact $K \subset U$ (note that C and k in general depend on K).

Example 1.3. The following are distributions (the continuity part is left as an exercise; the important aspect is viewing these as linear functionals):

- Any $f \in L^1_{loc}(U)$ can be identified with a distribution

$$T_f(\phi) := \int_U f(x)\phi(x)\,dx.$$

(Note that the RHS makes sense for any $\phi \in C_c^\infty(U)$, since $f \in L^1(K)$ for any compact subset $K \subset U$; in particular this is the case for $K = \text{supp } \phi$.) In such cases, we'll refer to the distribution as f as well, and we'll say that a distribution u is in $L^1_{loc}(U)$ (or L^p, continuous, C^∞, etc.) if it can be identified with a function in $L^1_{loc}(U)$ via the above identification.

Remark 1. This is slightly different than the complex inner product $\langle f, \phi\rangle = \int_U f(x)\overline{\phi}(x)\,dx$–note that $\langle \cdot, \cdot \rangle$ is $\mathbb{C}$-anti-linear in the second variable, i.e. $\langle f, \alpha\phi\rangle = \overline{\alpha}\langle f, \phi\rangle$ for $\alpha \in \mathbb{C}$, whereas distributions are $\mathbb{C}$-linear.

[4]Strictly speaking, this should be called *sequentially continuous*, as the topology on $C_c^\infty(U)$ is not metrizable.

- For any $x_0 \in U$ and multi-index α, the functional
$$\phi \mapsto \partial^\alpha \phi(x_0)$$
is a distribution. The case $\alpha = 0$ is called the *Dirac delta* at x_0, denoted δ_{x_0} (the Dirac delta at the origin 0 is often just denoted δ).
- In $\mathbb{R}$, the distributions $(x \pm i0)^{-1}$ are defined by
$$\phi \mapsto \lim_{\epsilon \to 0^+} \int_{\mathbb{R}} \frac{\phi(x)}{x \pm i\epsilon}\, dx.$$
The limit on the RHS does indeed exist for any $\phi \in C_c^\infty(\mathbb{R})$, and the limits may be different depending on the sign $\pm$. In fact, we have
$$(x - i0)^{-1} - (x + i0)^{-1} = 2\pi i \delta_0.$$

The space of distributions $\mathcal{D}'(U)$ also has a topology, given by the weak-* topology viewing it as the dual space to $C_c^\infty(U)$.

Definition 1.4. We say that a sequence u_n in $\mathcal{D}'(U)$ *converges* to $u \in \mathcal{D}'(U)$ *as distributions* if, for every $\phi \in C_c^\infty(U)$, we have
$$u_n(\phi) \to u(\phi) \quad (\text{in } \mathbb{C}).$$

In practice, convergence in $\mathcal{D}'(U)$ is weaker than most kinds[5] of convergence that can be considered. For example, if u_n has more structure, e.g. belongs to $C_c^\infty(U)$, $C^\infty(U)$, even $L^1_{loc}(U)$, and it converges, e.g. uniformly or even in L^1_{loc} (i.e. in L^1 on every compact set), then it also converges as distributions.

If $V \subset U$ is open, then there is a continuous inclusion $\iota : C_c^\infty(V) \hookrightarrow C_c^\infty(U)$.

Definition 1.5. If $V \subset U$ is open, and $u \in \mathcal{D}'(U)$, the *restriction* of u to V is the distribution on V defined by
$$u|_V(\phi) := u(\iota\phi).$$

Definition 1.6. The *support* of a distribution $u \in \mathcal{D}'(U)$ is the set
$$\operatorname{supp} u := \{x \in U \,:\, u|_V \text{ is } \textbf{not} \text{ identically zero for any neighborhood } V \ni x\}.$$

Example 1.7. If f is a continuous function on U, then the support of f, viewing f as a distribution, is the same as the support of f, viewed as a function. That is,
$$\operatorname{supp} f = \overline{\{x \in U \,:\, f(x) \neq 0\}}$$
(here the closure is taken with respect to U).

[5]The only kind of convergence I can think of that is not stronger than convergence in $\mathcal{D}'(U)$ is pointwise convergence. But even then, even a very weak bound on the sequence u_n, combined with pointwise convergence of u_n, is usually enough to give convergence in $\mathcal{D}'(U)$, due to the extremely nice properties of $C_c^\infty(U)$.

1.2. **Operations on distributions: differentiation, multiplication.** One essential operation that can be applied to distributions is differentiation. The idea is motivated by the integration by parts identity

$$\int_U \partial^\alpha f(x)\phi(x)\,dx = (-1)^{|\alpha|}\int_U f(x)\partial^\alpha \phi(x)\,dx$$

which holds for all $f \in C^\infty(U)$ and $\phi \in C_c^\infty(U)$ (the compact support of ϕ guarantees the lack of "boundary terms" that would normally arise from integration by parts). We thus *define* differentiation via this feature:

Definition 1.8. For $1 \le j \le n$ and $u \in \mathcal{D}'(U)$, the *partial derivative* $\partial_j u$ is the distribution on U defined by

$$\partial_j u(\phi) := -u(\partial_j \phi) \text{ for all } \phi \in C_c^\infty(U).$$

For a multi-index α, the derivative $\partial^\alpha u$ is defined by iterating partial derivatives, in the same manner as iterating partial derivatives for smooth functions.

Note that $\phi \in C_c^\infty(U) \implies \partial_j \phi \in C_c^\infty(U)$, so the right-hand side in the above definition does indeed make sense.

Example 1.9. For $a > -1$, define

$$\chi_+^a(x) = \begin{cases} \frac{x^a}{\Gamma(a+1)} & x > 0 \\ 0 & x \le 0 \end{cases}.$$

Then $\chi_+^a \in L^1_{loc}(\mathbb{R})$, and $\frac{d}{dx}\chi_+^a = \chi_+^{a-1}$ for $a > 0$. Moreover, $\chi_+^0(x)$ is the so-called "Heaviside function" $H(x)$, and $\frac{d}{dx}\chi_+^0 = \delta_0$.

Similarly, if $\rho \in C^\infty(U)$, then we have the identity

$$\int_U (\rho(x)u(x))\phi(x)\,dx = \int_U u(x)(\rho(x)\phi(x))\,dx$$

for any function u just by rearranging terms; moreover note that $\rho\phi \in C_c^\infty(U)$ if $\phi \in C_c^\infty(U)$. Thus, we can define:

Definition 1.10. For $\rho \in C^\infty(U)$ and $u \in \mathcal{D}'(U)$, the *product* ρu is the distribution on U defined by

$$(\rho u)(\phi) := u(\rho\phi).$$

Remark 2. Differentiation and multiplication by $\rho \in C^\infty(U)$, as defined above, are in fact *continuous* linear operators $\mathcal{D}'(U) \to \mathcal{D}'(U)$, such that the restriction of these operators to $C_c^\infty(U) \subset \mathcal{D}'(U)$ give the same results as the usual differentiation and multiplication on $C_c^\infty(U)$. The reason why we insist on only[6] multiplying by functions in $C^\infty(U)$ is that it turns out there does **not** exist a *continuous* operator $\mathcal{D}'(\mathbb{R}^n) \times \mathcal{D}'(\mathbb{R}^n) \to \mathcal{D}'(\mathbb{R}^n)$ which extends the notion of pointwise multiplication defined initially on $C_c^\infty(\mathbb{R}^n) \times C_c^\infty(\mathbb{R}^n)$.

[6] At least *a priori* it is not clear that multiplication by other functions will work. It turns out that we can multiply by slightly less regular functions under mild assumptions, but that this notion does not extend to *all* distributions.

2. Lecture 2 (03/31): Composition, homogeneous distributions, and convolution

2.1. **Composition with smooth maps and homogeneous distributions.** In some cases, it is possible to define composition of distributions. We'll focus on the case of composing a distribution in $\mathcal{D}'(\mathbb{R}^m)$ with a map $f : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$, with $m \le n$.

We suppose first that there exist auxiliary functions $y_{m+1}(x), y_{m+2}(x), \ldots, y_n(x)$ such that the function

$$g : U \to \mathbb{R}^n, \quad g(x) = (f_1(x), \ldots, f_m(x), y_{m+1}(x), \ldots, y_n(x))$$

has a C^∞ inverse $h : g(U) \to U$. For $y \in \mathbb{R}^n$, write $y = (y', y'') \in \mathbb{R}^m \times \mathbb{R}^{n-m}$. Note that

$$y = g(h(y)) \implies y' = f(h(y))$$

since the first m components of $g(x)$ are the m components of $f(x)$. Then, if $\phi \in C_c^\infty(U)$, we see that for $u \in C^0(\mathbb{R}^m)$ we have[7]

$$\begin{aligned}
\int_U u(f(x))\phi(x)\,dx &= \int_U u(f(x))\phi(x)\,dx \\
&= \int_{g(U)} u(f(h(y)))\phi(h(y))\,|\det Dh(y)|\,dy \\
&= \int_{\mathbb{R}^n} u(y')\phi(h(y))|\det Dh(y)|\,dy \\
&= \int_{\mathbb{R}^m} u(y')\tilde{\phi}(y')\,dy'
\end{aligned}$$

where

$$\tilde{\phi}(y') = \int_{\mathbb{R}^{n-m}} \phi(h(y', y''))|\det Dh(y', y'')|\,dy''.$$

Thus, if we can construct auxiliary functions $y_{m+1}, \ldots, y_n$ such that $(f, y_{m+1}, \ldots, y_n)$ has a C^∞ inverse, then we can (uniquely) extend the notion of composition of a continuous function $u : \mathbb{R}^m \to \mathbb{R}$ with $f : U \to \mathbb{R}^m$ to distributions on U by defining

$$(u \circ f, \phi)_U := (u, \tilde{\phi})_{\mathbb{R}^m}$$

where $\tilde{\phi}$ is defined as above.

In general, if the differential Df of f is surjective everywhere on U, then this construction of auxiliary functions is possible locally, thanks to the Inverse Function Theorem. We can then take a locally finite partition of unity $1 = \sum \rho_k$, say with $\operatorname{supp} \rho_k \subset U_k$, with the property that for any compact subset $K \subset U$, only finitely many intersections $U_k \cap K$ are nonempty. Then, we can define

$$(u \circ f, \phi)_U := \sum_k (u|_{U_k}, \rho_k \phi)_{U_k},$$

[7] In the second line, we made the substitution $x = h(y)$, $dx = |\det Dh(y)|\,dy$. In the third line, we can switch the region of integration to $\mathbb{R}^n$, since $\phi(h(y))$ is nonzero only when $h(y) \in \operatorname{supp} \phi \subset U \implies y \in g(U)$.

where $(u|_{U_k}, \rho_k\phi)_{U_k}$ is defined in the special case discussed in the previous paragraph (note that the assumption that the partition is locally finite guarantees the above sum is a finite sum for any $\phi \in C_c^\infty(U)$). We summarize this as follows:

Theorem 2.1 (cf. Theorem 6.1.2 in [Hör90])**.** *Suppose $f : U \to \mathbb{R}^m$ is smooth, and $Df(x)$ is surjective for all $x \in U$. Then there is a unique continuous linear map $f^* : \mathcal{D}'(\mathbb{R}^m) \to \mathcal{D}'(U)$ such that $f^*u = u \circ f$ for all $u \in C^0(\mathbb{R}^m)$. Moreover, for any partial derivative ∂_j we have*

$$\partial_j(f^*u) = \sum_{k=1}^n (\partial_j f_k) \cdot f^*(\partial_k u) \quad \text{(i.e. } \partial_j(u \circ f) = \sum_{k=1}^n (\partial_k u \circ f)\partial_j f_k\text{)}.$$

Remark 3. If $V \subset \mathbb{R}^m$ is an open set containing the image of $f : U \to \mathbb{R}^m$, then the above theorem also holds replacing $\mathbb{R}^m$ by V.

Example 2.2. Consider $f : \mathbb{R}^{2n} \to \mathbb{R}^n$, $f(x, x') = x - x'$ for $(x, x') \in \mathbb{R}^n \times \mathbb{R}^n$. Then $g(x, x') = (x - x', x')$ admits a C^∞ inverse on $\mathbb{R}^{2n}$, namely $h(y', y'') = (y' + y'', y'')$ for $(y', y'') \in \mathbb{R}^n \times \mathbb{R}^n$ (then $|\det Dh| = 1$ everywhere). Then

$$(u \circ f, \phi)_{\mathbb{R}^{2n}} = (u, \tilde{\phi})_{\mathbb{R}^n}$$

where

$$\tilde{\phi}(y') = \int_{\mathbb{R}^n} \phi(y' + y'', y'')\, dy''.$$

In particular, the distribution $\delta(x - x')$ is the distribution satisfying

$$(\delta(x - x'), \phi)_{\mathbb{R}^{2n}} = (\delta, \tilde{\phi})_{\mathbb{R}^n} = \tilde{\phi}(0) = \int_{\mathbb{R}^n} \phi(0 + y'', y'')\, dy'' = \int_{\mathbb{R}^n} \phi(x, x)\, dx$$

for any $\phi \in C_c^\infty(\mathbb{R}^{2n})$.

Example 2.3. Suppose $U \subset \mathbb{R}^n$ is *conic*, meaning that $x \in U \implies tx \in U$ for all $t > 0$ (for example, $U = \mathbb{R}^n$ or $U = \mathbb{R}^n \backslash \{0\}$). For $t > 0$, the composition $u(tx)$ for $u \in \mathcal{D}'(U)$ is well-defined, and it satisfies

$$(u(tx), \phi) = t^{-n}(u, \phi(x/t)).$$

(This is the same result if u is a continuous function and $u(tx)$ is understood as a composition.)

Definition 2.4. We say that $u \in \mathcal{D}'(U)$ is *homogeneous* of degree $a \in \mathbb{R}$ if $u(tx) = t^a u(x)$ for all $t > 0$, where the composition $u(tx)$ is defined above. Equivalently,

$$(u, \phi(x/t)) = t^{n+a}(u, \phi) \quad \text{or} \quad (u, \phi(tx)) = t^{-n-a}(u, \phi).$$

Example 2.5. Consider the functions χ_+^a defined in Example 1.9 for $a > -1$. They define distributions which are homogeneous of degree a.

Example 2.6. The Dirac delta δ on $\mathbb{R}^n$ is homogeneous of degree $-n$. Indeed, $(\delta, \phi(tx)) = (\delta, \phi)$ for any $t > 0$ since both sides evaluate to $\phi(0)$, so $(\delta, \phi(tx)) = t^{-n-a}(\delta, \phi)$ for $-n - a = 0$, i.e. $a = -n$.

Example 2.7. Consider $f : \mathbb{R}^{n+1}_{t,x} \to \mathbb{R}$ be given by $f(t,x) = t^2 - |x|^2$ for $(t,x) \in \mathbb{R}\times\mathbb{R}^n$. Note that Df is non-vanishing on $\mathbb{R}^{n+1}\backslash\{0\}$, so $u(t^2-|x|^2)$ is well-defined as a distribution on $\mathbb{R}^{n+1}\backslash\{0\}$. Moreover, if u is homogeneous of degree a, then $u(t^2-|x|^2)$ is homogeneous of degree $2a$.

Note that given a distribution u on $\mathbb{R}^n\backslash\{0\}$, it is not always possible to extend it to a distribution $\tilde{u}$ on $\mathbb{R}^n$ (i.e. we cannot always find $\tilde{u} \in \mathcal{D}'(\mathbb{R}^n)$ such that $\tilde{u}|_{\mathbb{R}^n\backslash\{0\}} = u$). However, for *homogeneous* distributions this is possible:

Theorem 2.8 (cf. Theorems 3.2.3 and 3.2.4 in [Hör90])**.** *Suppose $u \in \mathcal{D}'(\mathbb{R}^n\backslash\{0\})$ is homogeneous of degree a. Then there exists $\tilde{u} \in \mathcal{D}'(\mathbb{R}^n)$ such that $\tilde{u}|_{\mathbb{R}^n\backslash\{0\}} = u$. Moreover, if either $a > -n$ or $a \notin \mathbb{Z}$, then such an extension is unique and also homogeneous of degree a.*

2.2. **Convolution.** Recall that the convolution of two functions $f, g : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)\,dy = (f, g(x-\cdot))$$

where $g(x-\cdot)$ is the function $y \mapsto g(x-y)$. This allows us to easily define convolution when f is a distribution and g is smooth *with compact support*:

Definition 2.9. Let $u \in \mathcal{D}'(\mathbb{R}^n)$ and $\phi \in C_c^\infty(\mathbb{R}^n)$. The convolution of u and ϕ is the *function* $u * \phi : \mathbb{R}^n \to \mathbb{R}$ defined by

$$(u * \phi)(x) := (u, \phi(x-\cdot)).$$

(Note that $\phi(x-\cdot) \in C_c^\infty(\mathbb{R}^n)$ for any $x \in \mathbb{R}^n$ if $\phi \in C_c^\infty(\mathbb{R}^n)$.)

Some properties of convolution as defined above are as follows (proofs are in Section 4.1 of [Hör90]):

Theorem 2.10. *Let $u \in \mathcal{D}'(\mathbb{R}^n)$ and $\phi, \psi \in C_c^\infty(\mathbb{R}^n)$. Then:*

- *$u * \phi \in C^\infty(\mathbb{R}^n)$, with*[8] *supp $(u*\phi) \subset$ supp u + supp ϕ.*
- *For any multi-index α we have*

$$\partial^\alpha(u * \phi) = (\partial^\alpha u) * \phi = u * (\partial^\alpha \phi).$$

- *We have*[9]*$(u * \phi) * \psi = u * (\phi * \psi)$.*
- *$\phi * \psi = \psi * \phi$ (viewing $C_c^\infty(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n)$).*

Next time: We'll define the composition of distributions (in limited contexts) by having it satisfy the associativity condition

$$(u_1 * u_2) * \phi = u_1 * (u_2 * \phi).$$

We'll also discuss Schwartz kernels.

[8]For two subsets $A, B \subset \mathbb{R}^n$, we define $A + B = \{a + b : a \in A, b + B\}$.

[9]Note in this equation that $u * \phi \in C^\infty(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n)$, so we can again convolve it on the right with another function in $C_c^\infty(\mathbb{R}^n)$, while $\phi * \psi \in C_c^\infty(\mathbb{R}^n)$, so we can again convolve it on the left with a distribution.

3. Lecture 3 (04/05): Duality, Convolutions, and Schwartz Kernel

3.1. **A comment on defining operations by duality.** For all of the operations on distributions defined in the first two lectures, we can summarize the definitions in a common way. Suppose we have a continuous linear map $L : C_c^\infty(X) \to \mathcal{D}'(Y)$.

Definition 3.1. The *adjoint* of $L : C_c^\infty(X) \to \mathcal{D}'(Y)$ is the operator ${}^tL : C_c^\infty(Y) \to \mathcal{D}'(X)$ defined by

$$({}^tL\psi, \phi)_X = (L\phi, \psi)_Y \quad \text{for } \phi \in C_c^\infty(X), \psi \in C_c^\infty(Y).$$

Note then that ${}^t({}^tL) = L$.

Example 3.2. Some examples:

- For $L = \partial_j$, we have ${}^tL = -\partial_j$.
- For $L = \rho$ (i.e. multiplication by $\rho \in C^\infty$), we have ${}^tL = \rho$ (i.e. the same operator).
- If $f : X \to Y$ is invertible with smooth inverse and $L = f^*$ (i.e. $L\phi = \phi \circ f$), we have ${}^tL = |\det f^{-1}|(f^{-1})^*$ (i.e. ${}^tL\psi = (\psi \circ f^{-1})|\det f^{-1}|$).

Suppose now that ${}^tL$ now maps $C_c^\infty(Y)$ not just into $\mathcal{D}'(X)$, but rather into $C_c^\infty(X)$. Then, we can extend L to an operator $\tilde{L} : \mathcal{D}'(X) \to \mathcal{D}'(Y)$, defined by

$$(\tilde{L}u, \psi)_Y := (u, {}^t L\psi)_X.$$

Then, $\tilde{L}$ is continuous, and moreover it agrees with L on $C_c^\infty(X)$. This is indeed how all of the operations defined so far (aside from convolution with C_c^∞) have been defined.

If ${}^tL$ does not map $C_c^\infty(Y)$ into $C_c^\infty(X)$, but rather just into $C^\infty(X)$, we can still extend L, albeit to a slightly different space.

Definition 3.3. The space $\mathcal{E}'(X)$ is the space of distributions $u \in \mathcal{D}'(X)$ such that $\operatorname{supp} u$ is compact.

Theorem 3.4. *$\mathcal{E}'(X)$ is isomorphic, as topological vector spaces, to the dual space of $C^\infty(X)$, where the topology of $C^\infty(X)$ is that given by the seminorms*

$$\phi \mapsto \sum_{|\alpha| \le k} \sup_K |\partial^\alpha \phi|, \quad k \in \mathbb{N}_{\ge 0}, K \subset X \text{ compact.}$$

Thus, if ${}^tL$ maps $C_c^\infty(Y)$ into $C^\infty(X)$, then L can be extended as a map $\mathcal{E}'(X) \to \mathcal{D}'(Y)$. Similarly, if ${}^tL$ maps into $\mathcal{E}'(X)$, then L can be extended as a map $C^\infty(X) \to \mathcal{D}'(Y)$.

3.2. **Convolutions, continued.** To define the convolution of two distributions u_1 and u_2, we could try to have it satisfy the "associativity" property above, i.e. for $\phi \in C_c^\infty(\mathbb{R}^n)$ we would want $u_1 * u_2$ to satisfy

$$(u_1 * u_2) * \phi = u_1 * (u_2 * \phi).$$

There are two issues with doing so:

- First, the above condition gives a condition we'd like to be satisfied involving the *convolution* of our mystery distribution $u_1 * u_2$ against ϕ; *a priori* it's not clear how that defines the pairing $(u_1 * u_2, \phi)$.
- For the right-hand side to make sense, we want to somehow arrange for $u_2 * \phi$ to have compact support.

The second issue can be addressed by taking u_2 to have compact support. For the first issue, it turns out that knowledge of how a distribution convolves (i.e. knowledge of the operator $C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$, $\phi \mapsto u * \phi$) is enough information to determine the distribution itself. Indeed, just note that for any $\phi \in C_c^\infty(\mathbb{R}^n)$, if $\check{\phi}(x) = \phi(-x)$, then

$$(u * \check{\phi})(0) = (u, \phi).$$

However, there is another necessary condition that a convolution operator must satisfy. Namely, if for $h \in \mathbb{R}^n$ we let $\tau_h : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$, $\tau_h\phi(x) = \phi(x - h)$, then $\tau_h(u * \phi) = u * \tau_h\phi$ (this follows basically from the definition). It turns out that this is essentially sufficient as well:

Theorem 3.5 (cf. Theorem 4.2.1 of [Hör90])**.** *If U is a continuous*[10] *linear map from $C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$, and $U \circ \tau_h = \tau_h \circ U$ for all $h \in \mathbb{R}^n$, then there exists a unique $u \in \mathcal{D}'(\mathbb{R}^n)$ such that $U\phi = u * \phi$ for all $\phi \in C_c^\infty(\mathbb{R}^n)$.*

Proof. If such a u were to exist, it must satisfy $(u, \phi) = (u * \check{\phi})(0) = U(\check{\phi})(0)$; hence *define* $u \in \mathcal{D}'(\mathbb{R}^n)$ by $(u, \phi) := U(\check{\phi})(0)$. This is a distribution, i.e. is continuous, due to the continuity assumptions in the hypothesis. It remains to verify that $U\phi = u * \phi$ for all $\phi \in C_c^\infty(\mathbb{R}^n)$. This is where the assumption of commuting with translations comes in: just note that

$$(U\phi)(-h) = \tau_h(U\phi)(0) = U(\tau_h\phi)(0) = (u, \check{\tau_h\phi}) = (u*\tau_h\phi)(0) = \tau_h(u*\phi)(0) = (u*\phi)(-h)$$

for each $h \in \mathbb{R}^n$. Thus, $U\phi = u * \phi$ as functions in $C^\infty(\mathbb{R}^n)$. □

As such, if $u_1, u_2 \in \mathcal{D}'(\mathbb{R}^n)$ with supp u_2 compact, we see that

$$U\phi := u_1 * (u_2 * \phi)$$

satisfies the assumptions in the theorem. Hence, we can make a definition:

Definition 3.6. Suppose $u_1, u_2 \in \mathcal{D}'(\mathbb{R}^n)$ with supp u_2 compact. The *convolution* $u_1 * u_2$ is the unique distribution $u \in \mathcal{D}'(\mathbb{R}^n)$ satisfying

$$u * \phi = u_1 * (u_2 * \phi) \text{ for all } \phi \in C_c^\infty(\mathbb{R}^n).$$

Example 3.7. The Dirac delta δ_0 satisfies $\delta_0 * \phi = \phi$ for all $\phi \in C_c^\infty(\mathbb{R}^n)$. As a consequence, for all distributions $u \in \mathcal{D}'(\mathbb{R}^n)$ we also have $u * \delta_0 = u$.

By leveraging facts about convolution of functions, we can state:

Theorem 3.8. *Let $u_1, u_2 \in \mathcal{D}'(\mathbb{R}^n)$ with supp u_2 compact. Then:*

[10]The topology on $C^\infty(\mathbb{R}^n)$ is the seminorm topology induced by the seminorms $\phi \to \sum_{|\alpha| \leq k} \sup_K |\partial^\alpha \phi|$ over all $k \in \mathbb{N}_{\geq 0}$ and K compact, i.e. a sequence of smooth functions converges if and only if it converges with respect to each of the preceding seminorms.

- *If supp u_1 is also compact, then $u_1 * u_2 = u_2 * u_1$.*
- *We have supp $(u_1 * u_2) \subset$ supp u_1 + supp u_2.*
- *If $u_3 \in \mathcal{D}'(\mathbb{R}^n)$ has compact support, then $u_1 * (u_2 * u_3) = (u_1 * u_2) * u_3$.*
- *We have $\partial^\alpha(u_1 * u_2) = (\partial^\alpha u_1) * u_2 = u_1 * (\partial^\alpha u_2)$. In particular, if $P = \sum a_\alpha \partial^\alpha$ is a **constant-coefficient** differential operator, then $P(u_1 * u_2) = (Pu_1) * u_2 = u_1 * (Pu_2)$.*

An application of the last fact is the following: suppose P is a constant-coefficient differential operator, and u_2 is compactly supported and satisfies $Pu_2 = \delta$ in the sense of distributions. Then, for any $u_1 \in \mathcal{D}'(\mathbb{R}^n)$, for $u = u_1 * u_2$ we have

$$Pu = P(u_1 * u_2) = u_1 * (Pu_2) = u_1 * \delta = u_1.$$

Thus, for any $u_1 \in \mathcal{D}'(\mathbb{R}^n)$, there exists a solution to $Pu = u_1$ in the sense of distributions, namely $u = u_1 * u_2$. This idea will be used more heavily next week.

There are many more situations in which the convolution of two distributions can be well-defined. One such situation is the following: suppose the map

$$\text{supp } u_1 \times \text{supp } u_2 \to \mathbb{R}^n, \quad (x, y) \mapsto x + y$$

is *proper*, meaning that the pre-image of compact sets is compact[11], then the convolution $u_1 * u_2$ can be defined as follows: for a fixed $\phi \in C_c^\infty(\mathbb{R}^n)$, let $K =$ supp ϕ, and let K_1 and K_2 be the projections of the preimage of K under the map supp $u_1 \times$ supp $u_2 \to \mathbb{R}^n$,$(x, y) \mapsto x + y$. (Thus, if $(x, y) \in$ supp $u_1 \times$ supp u_2 and $x + y \in K$, then $x \in K_1$ and $y \in K_2$). Note that K, K_1, and K_2 are all compact, by the properness assumption. We then define

$$(u_1 * u_2, \phi) := ((\psi_1 u_1) * (\psi_2 u_2), \phi)$$

where $\psi_1, \psi_2 \in C_c^\infty(\mathbb{R}^n)$ are identically 1 in neighborhoods of K_1 and K_2, respectively. Note then that $\psi_i u_i$ are compactly supported distributions, so their convolution is well-defined. The idea behind the definition is to cut off the distributions u_1 and u_2 "only where they matter" when trying to evaluate the pairing of $u_1 * u_2$ against ϕ.

To make this a well-defined definition, we need to check that this result is independent of the cutoffs ψ_i chosen. For example, to check the result is independent of the choice of ψ_2, suppose ψ_2 and $\tilde{\psi}_2$ are both identically 1 in a neighborhood of K_2. We need to check that $((\psi_1 u_1) * (\tilde{\psi}_2 u_2), \phi) = ((\psi_1 u_1) * (\psi_2 u_2), \phi)$. Unraveling the definition of compositions, this amounts to checking that

$$((\psi_1 u_1) * (((\tilde{\psi}_2 - \psi_2) u_2) * \check{\phi}))(0) = 0.$$

We note that $\tilde{\psi}_2 - \psi_2$ equals zero on a neighborhood of K_2, so $y \in \text{supp } (\tilde{\psi}_2 - \psi_2) \implies y \notin K_2$. In particular, if $x \in$ supp u_1 and $y \in \text{supp } (\tilde{\psi}_2 - \psi_2)$, then $x + y \notin$ supp ϕ.

[11]This can be phrased in other ways; two examples of which are the following: for any compact set $K \subset \mathbb{R}^n$:

- The set $(K -$ supp $u_1) \cap$ supp u_2 is compact.
- There exists $C > 0$ such that if $x \in$ supp u_1 and $y \in$ supp u_2, then $x + y \in K \implies |x| \leq C, |y| \leq C$.

But we also know that

$$\begin{aligned}\text{supp }((\psi_1 u_1) * (((\tilde{\psi}_2 - \psi_2)u_2) * \check{\phi})) &\subset \text{supp }(\psi_1 u_1) + \text{supp }((\tilde{\psi}_2 - \psi_2)u_2) + \text{supp }\check{\phi} \\ &\subset \text{supp }u_1 + \text{supp }(\tilde{\psi}_2 - \psi_2) - \text{supp }\phi,\end{aligned}$$

and the latter set does not contain 0 by the discussion above. This shows that $((\psi_1 u_1) * (\psi_2 u_2), \phi)$ does not depend on the choice of ψ_2, so long as ψ_2 is identically 1 in a neighborhood of K_2. Similar methods show this is independent of the choice of ψ_1 as well.

Example 3.9. For $\overline{\mathbb{R}_+} = [0, \infty)$, we have that $\overline{\mathbb{R}_+} \times \overline{\mathbb{R}_+} \to \mathbb{R}$, $(x, y) \to x + y$ is proper. Hence, if supp $u_1, u_2 \subset \overline{\mathbb{R}_+}$, then the convolution $u_1 * u_2$ is well-defined.

3.3. **Products and Schwartz Kernel.** One reason to study distributions, even if one is only interested in smooth solutions to a differential equation, is that they are intimately related to operators. First we consider products. In this subsection, let X and Y be open subsets of some Euclidean spaces.

Definition 3.10. For $\phi \in C_c^\infty(X)$ and $\psi \in C_c^\infty(Y)$, the *tensor product* is the function $\phi \otimes \psi \in C_c^\infty(X \times Y)$ defined by

$$(\phi \otimes \psi)(x, y) = \phi(x)\psi(y).$$

From this, we can define the tensor product of distributions:

Theorem 3.11. *Let $u_1 \in \mathcal{D}'(X)$ and $u_2 \in \mathcal{D}'(Y)$. Then there exists a unique distribution $u \in \mathcal{D}'(X \times Y)$ satisfying*

$$u(\phi \otimes \psi) = u_1(\phi)u_2(\psi).$$

This is called the tensor product *of the distributions u_1 and u_2 and is denoted $u_1 \otimes u_2$.*

We now consider the following situation: suppose we're *given* a distribution K on the product space $X \times Y$. Then the map

$$\psi \mapsto (\phi \mapsto (K, \phi \otimes \psi)_{X \times Y})$$

defines a linear operator $T : C_c^\infty(Y) \to \mathcal{D}'(X)$. That is, given $\psi \in C_c^\infty(Y)$, $T\psi$ is a distribution on X satisfying $(T\psi, \phi)_X = (K, \phi \otimes \psi)_{X \times Y}$. (In more informal terms, if we were to view K as a function $K(x, y)$, then the operator in question is

$$T\psi(x) = \int_{X \times Y} K(x, y)\psi(y)\, dy.)$$

Moreover, this operator T is continuous with respect to the respective topologies.

Definition 3.12. Let $T : C_c^\infty(Y) \to \mathcal{D}'(X)$ be linear, and suppose $K \in \mathcal{D}'(X \times Y)$ satisfies $(T\psi, \phi)_X = (K, \phi \otimes \psi)_{X \times Y}$ for all $\phi \in C_c^\infty(X)$ and $\psi \in C_c^\infty(Y)$. Then we say that K is a *Schwartz kernel* of T.

A remarkable fact is:

Theorem 3.13 (Schwartz Kernel Theorem)**.** *Every continuous linear operator $C_c^\infty(Y) \to \mathcal{D}'(X)$ has a* ***unique*** *Schwartz kernel associated to it.*

Example 3.14. Some examples of Schwartz kernels (here the coordinates on X and Y are denoted x and y):

- The identity operator $\mathrm{Id} : C_c^\infty(X) \to C_c^\infty(X) \subset \mathcal{D}'(X)$ has Schwartz kernel $\delta(x-y)$.
- The differential operator $P = \sum a_\alpha(x)\partial^\alpha$ has Schwartz kernel $\sum a_\alpha(x)\partial^\alpha \delta(x-y)$.
- If $u \in \mathcal{D}'(\mathbb{R}^n)$, the convolution operator $C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$, $\phi \mapsto u * \phi$ has Schwartz kernel $u(x-y)$.
- If $T : C_c^\infty(Y) \to \mathcal{D}'(X)$ has Schwartz kernel $K \in \mathcal{D}'(X \times Y)$, then ${}^tT : C_c^\infty(X) \to \mathcal{D}'(Y)$ has Schwartz kernel ${}^tK \in \mathcal{D}'(Y \times X)$, where "${}^tK(y,x) = K(x,y)$". Formally, if ${}^t\psi(x,y) = \psi(y,x)$ for $\phi \in C_c^\infty(Y \times X)$, then ${}^t\phi \in C_c^\infty(X \times Y)$, and
$$({}^tK, \psi)_Y = (K, {}^t\psi)_X.$$
- Suppose the Schwartz kernel, which *a priori* is a distribution on $X \times Y$, is actually in $L^2(X \times Y)$. Then the corresponding operator is a *Hilbert-Schmidt* operator. Such operators have some nice properties (e.g. they are compact operators).

***The following was not covered during lecture but may be of interest for some students*:**

One application of studying the Schwartz kernel is the following: we can often give an upper bound on the set of singularities of Tu, if we know the singularities of the Schwartz kernel of T and of u. We formalize the notions as follows:

Definition 3.15. Let $u \in \mathcal{D}'(X)$. The *singular support* of u, denoted sing supp u, is the set of $x \in X$ such that, for any neighborhood $V \ni x$, the restriction $u|_V$ does **not** agree with the restriction of any smooth function on V. (Equivalently, $x \in X$ is *not* in the singular support if there exists a neighborhood V of x such that $u|_V$ is smooth, i.e. agrees with the restriction of some smooth function).

Definition 3.16. Suppose $A \subset X \times Y$ and $B \subset Y$. The *composition* $A \circ B$ of sets is the set
$$A \circ B = \{x \in X : \text{ there exists } y \in B \text{ such that } (x,y) \in A\}.$$
Equivalently,
$$A \circ B = \pi_X(A \cap \pi_Y^{-1}(B))$$
where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are the projections onto X and Y.

Theorem 3.17. *Suppose $T : C_c^\infty(Y) \to \mathcal{D}'(X)$ satisfies that T maps into $C^\infty(X)$, and ${}^tT$ maps continuously from $C_c^\infty(X)$ to $C^\infty(Y)$, so that T can be extended to a map $T : \mathcal{E}'(Y) \to \mathcal{D}'(X)$. Let $K \in \mathcal{D}'(X \times Y)$ be the Schwartz kernel of T. Then,*
$$\text{sing supp } Tu \subset \text{sing supp } K \circ \text{sing supp } u \quad \text{for all } u \in \mathcal{E}'(Y).$$

Proof Sketch. The proof boils down to the following statement, which will not be proven in this sketch:

if $T : \mathcal{E}'(Y) \to \mathcal{D}'(X)$ has Schwartz kernel $K \in C^\infty(Y \times X)$, then $Tu \in C^\infty(X)$ for all $u \in \mathcal{E}'(Y)$.

Assuming the statement, we can prove as follows. Suppose $(x, y) \notin \text{sing supp}\, K \circ \text{sing supp}\, u$; we'd then like to show that $x \notin \text{sing supp}\, Tu$. Note that the assumptions then give that $\{x\} \times \text{sing supp}\, u$ is disjoint from sing supp K. Since the former set is compact and the latter set is closed, it follows that there exists open neighborhoods U and V of x and sing supp u, respectively, such that $U \times V$ is still disjoint from sing supp u. Let U' and V' be neighborhoods of x and sing supp u compactly contained in U and V, respectively, and let $\phi \in C_c^\infty(U)$ and $\psi \in C_c^\infty(V)$ be identically 1 on U' and V'. Then the operator $\phi T \psi$ has Schwartz kernel $(\phi \otimes \psi) \cdot K$ (or more colloquially $\phi(x)K(x, y)\psi(y)$), which is in $C^\infty(X \times Y)$ since the support of $\phi \otimes \psi$ is disjoint from the singular support of K, and hence $\phi T \psi u \in C^\infty(X)$. On the other hand, $u - \psi u \in C_c^\infty(Y)$ since $1 - \psi$ is supported away from sing supp u, and hence $\phi T(1 - \psi)u \in C^\infty(X)$ as well. Thus we have

$$\phi T u = \phi T \psi u + \phi T(1 - \psi)u \in C^\infty(X).$$

Since ϕ is identically 1 on U', it follows that the restriction of Tu to U' is equal to that of a C^∞ function on U', and hence $x \notin \text{sing supp}\, Tu$, as desired. $\square$

Next Lecture*: Tempered distributions and Fourier transform.*

4. Lecture 4 (04/07): Fourier Transform

4.1. **Fourier Transform on functions and Schwartz space.** Recall the Fourier transform on functions:

Definition 4.1. Let $f \in L^1(\mathbb{R}^n)$. The *Fourier transform* of f is the function $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ defined by[12]

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

Some properties:

Theorem 4.2. *Let $f \in L^1(\mathbb{R}^n)$. Then:*

(1) We have $\hat{f} \in C^0(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$.
(2) If in addition $f \in C^1(\mathbb{R}^n)$ and $\partial_j f \in L^1(\mathbb{R}^n)$, then $\widehat{\partial_j f}(\xi) = i\xi_j \hat{f}(\xi)$.
(3) If in addition $x_j f \in L^1(\mathbb{R}^n)$, then $\hat{f} \in C^1(\mathbb{R}^n)$ with $\partial_j \hat{f}(\xi) = \widehat{-ix_j f}(\xi)$.

It follows that the Fourier transform intertwines differentiation and multiplication by monomials. Hence, we are interested in a space of functions which behaves well under both operations:

Definition 4.3. The *Schwartz space*, denoted $\mathcal{S}$ or $\mathcal{S}(\mathbb{R}^n)$, is the set of all smooth functions $\phi \in C^\infty(\mathbb{R}^n)$ satisfying the property that

$$\sup_{x\in\mathbb{R}^n} |x^\beta \partial^\alpha \phi(x)| < \infty$$

for all multi-indices α, β. It is a topological vector space, when equipped with the topology induced by the seminorms appearing in the left-hand side of the above inequality.

That is, the Schwartz space consists of functions that are not only infinitely differentiable, but in addition decay faster than any inverse polynomial rate, with their derivatives decaying that fast as well. Note that $\phi \in \mathcal{S}(\mathbb{R}^n) \implies x^\beta\partial^\alpha\phi \in \mathcal{S}(\mathbb{R}^n)$ for any multi-indices α and β, i.e. $\mathcal{S}(\mathbb{R}^n)$ is closed under differentiation and multiplication by polynomials.

Example 4.4. We have the inclusion[13] $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$, i.e. any compactly supported smooth function will satisfy the above estimates.

---

[12] There are multiple commonly used conventions regarding the definition/normalization of the Fourier transform. This is the convention we'll use, since it works well with differentiation.

[13] More accurately, we should write that there is an inclusion $C_c^\infty(\mathbb{R}^n) \hookrightarrow \mathcal{S}(\mathbb{R}^n)$ which is continuous with respect to the respective topologies on $C_c^\infty(\mathbb{R}^n)$ and $\mathcal{S}(\mathbb{R}^n)$. Note that this inclusion, while continuous, is not an *embedding* of topological vector spaces, i.e. the topology on $C_c^\infty(\mathbb{R}^n)$ is *not* the subspace topology obtained by the inclusion into $\mathcal{S}(\mathbb{R}^n)$. In particular, a sequence $\{\phi_k\}$ in $C_c^\infty(\mathbb{R}^n)$ may converge in $\mathcal{S}(\mathbb{R}^n)$ without converging in $C_c^\infty(\mathbb{R}^n)$. To see this, fix a nonzero $\phi \in C_c^\infty(\mathbb{R}^n)$, let $\{a_k\}$ and $\{b_k\}$ be decreasing sequences of positive numbers, and let $\phi_k(x) = a_k\phi(b_k x)$. One can check that a sufficient condition for ϕ_k to converge to 0 in $\mathcal{S}(\mathbb{R}^n)$ is for $\lim_{k\to\infty} a_k b_k^m = 0$ for all $m \in \mathbb{Z}$; this can e.g. be arranged by taking $a_k = e^{-k}$ and $b_k = 1/k$. However, ϕ_k does not converge to 0 in $C_c^\infty(\mathbb{R}^n)$ since $\text{supp}\ \phi_k = b_k^{-1}\text{supp}\ \phi$, so that in particular the supports of ϕ_k are not all contained in some fixed compact set, thus violating a necessary condition for sequences to converge in $C_c^\infty(\mathbb{R}^n)$.

Example 4.5. If A is a symmetric positive definite $n \times n$ matrix, then $\phi(x) = e^{-\langle Ax, x\rangle/2}$ is in $\mathcal{S}(\mathbb{R}^n)$.

The decay requirement on Schwartz functions gives that any Schwartz function is integrable, and hence we can consider the Fourier transform of Schwartz functions. We then have:

Lemma 4.6. *For any $\phi \in \mathcal{S}(\mathbb{R}^n)$, we have $\hat{\phi} \in \mathcal{S}(\mathbb{R}^n)$ as well.*

Proof. It suffices to note that the intertwining of multiplication and differentiation allows us to conclude that

$$\xi^\beta \partial_\xi^\alpha \hat{\phi}(\xi) = \hat{\psi}(\xi), \quad \text{where } \psi = (-i\partial)_x^\beta \left((-ix)^\alpha \phi\right).$$

Then ψ is also a Schwartz function, and hence the Fourier transform is bounded. $\square$

4.2. **Tempered distributions and extending the Fourier transform.** From Lemma 4.6, we see that the Schwartz space is a nice space of "test functions" which behaves well with respect to the Fourier transform. This motivates considering a class of distributions dual to this nice test space:

Definition 4.7. The space of *tempered distributions*, denoted $\mathcal{S}'(\mathbb{R}^n)$, is the dual space (i.e. space of continuous linear functionals into $\mathbb{C}$) of $\mathcal{S}(\mathbb{R}^n)$, where $\mathcal{S}(\mathbb{R}^n)$ is equipped with the seminorm topology. The space of tempered distributions is also a topological vector space, equipped with the weak-* topology.

Remark 4. Since $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n)$, with all inclusions continuous with respect to the respective topologies, it follows that we have inclusions $\mathcal{E}'(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n)$, since $\mathcal{D}'(\mathbb{R}^n)$ and $\mathcal{E}'(\mathbb{R}^n)$ are the dual spaces of $C_c^\infty(\mathbb{R}^n)$ and $C^\infty(\mathbb{R}^n)$, respectively.

Remark 5. Note as well that $\mathcal{S}'(\mathbb{R}^n)$ is closed under differentiation, as well as multiplication by either functions in $\mathcal{S}(\mathbb{R}^n)$ or by polynomials, though not necessarily by arbitrary smooth functions.

Remark 6. It can be shown that $\mathcal{S}(\mathbb{R}^n)$ is in fact dense in $\mathcal{S}'(\mathbb{R}^n)$ (with respect to the weak-* topology on $\mathcal{S}'(\mathbb{R}^n)$). Thus, if we want to extend operators initially defined on $\mathcal{S}$ to continuous operators defined on $\mathcal{S}'$, such an extension would necessarily be unique due to the density of $\mathcal{S}$ in $\mathcal{S}'$.

We now ask how to define the Fourier transform for tempered distributions. We thus aim to find the adjoint of the Fourier transform, i.e. for $\phi, \psi \in \mathcal{S}(\mathbb{R}^n)$, see if we can rewrite the pairing $(\hat{\phi}, \psi)$ in terms of ϕ applied to an operator of ψ. Indeed, we see that

$$\begin{aligned}
(\hat{\phi}, \psi) = \int_{\mathbb{R}^n} \hat{\phi}(\xi)\psi(\xi)\, d\xi &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{-i\xi\cdot x}\phi(x)\, dx \right) \psi(\xi)\, d\xi \\
&= \int_{\mathbb{R}^n} \phi(x) \left(\int_{\mathbb{R}^n} e^{-i\xi\cdot x}\psi(\xi)\, d\xi \right) dx \\
&= \int_{\mathbb{R}^n} \phi(x)\hat{\psi}(x)\, dx = (\phi, \hat{\psi})
\end{aligned}$$

by Fubini's Theorem. Note as well that $\hat{\psi} \in \mathcal{S}(\mathbb{R}^n)$ if $\psi \in \mathcal{S}(\mathbb{R}^n)$. Thus, we define:

Definition 4.8. Given $u \in \mathcal{S}'(\mathbb{R}^n)$, the *Fourier transform* of u is the distribution $\hat{u} \in \mathcal{S}'(\mathbb{R}^n)$ defined by

$$(\hat{u}, \phi) := (u, \hat{\phi}).$$

We will sometimes denote the Fourier transform as an operator $\mathcal{F} : \mathcal{S}' \to \mathcal{S}'$ (i.e. $\mathcal{F}(u) = \hat{u}$). Note that $\mathcal{F}$ is continuous both as a map $\mathcal{S} \to \mathcal{S}$ and $\mathcal{S}' \to \mathcal{S}'$.

Some important properties (often proven by proving the analogous properties for Schwartz functions):

Theorem 4.9. *Let $u, v \in \mathcal{S}'(\mathbb{R}^n)$. Then:*

- *If $u \in L^1(\mathbb{R}^n)$, then the distribution $\hat{u}$ defined in Definition 4.8 agrees with the continuous bounded function $\hat{u}$ defined in Definition 4.1.*
- *If $u \in L^2(\mathbb{R}^n)$, then the distribution $\hat{u}$ is in fact in $L^2(\mathbb{R}^n)$. Moreover, for $v \in L^2(\mathbb{R}^n)$, we have the **Plancherel formula***

$$(\hat{u}, \overline{\hat{v}}) = (2\pi)^n (u, \overline{v}) \quad (\textit{in particular } \|\hat{u}\|_{L^2} = (2\pi)^{n/2}\|u\|_{L^2}).$$

- *If u is compactly supported, then $\hat{u}$ is in fact a C^∞ function, and moreover it satisfies*

$$\hat{u}(\xi) = (u, e^{-i\xi\cdot x})$$

(the RHS means $(u, \chi(x)e^{-i\xi\cdot x})$ for any $\chi \in C_c^\infty(\mathbb{R}^n)$ which is identically 1 on supp u.)

- *If $v \in \mathcal{E}'(\mathbb{R}^n)$, then*

$$\widehat{u * v} = \widehat{u}\widehat{v}.$$

(The formula continues to hold in many other situations as well.)

- *If u and v are sufficiently nice (e.g. in $\mathcal{S}$), then*

$$\widehat{uv} = (2\pi)^{-n}\widehat{u} * \widehat{v}.$$

- *In the sense of distributions, we have*

$$\widehat{\partial_{x_j} u} = i\xi_j \hat{u}, \quad \widehat{x_j u} = i\partial_{\xi_j}\hat{u}.$$

- *We have*

$$\hat{\hat{u}}(-x) = (2\pi)^n u,$$

and hence

$$\mathcal{F}^{-1}u(x) = (2\pi)^{-n}\mathcal{F}u(-x),$$

or more colloquially the inverse Fourier transform is given by

$$u(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{u}(\xi)\,d\xi.$$

4.3. **Examples.**

Example 4.10. Let $\phi(x) = e^{-ax^2/2}$ on $\mathbb{R}$, with $a > 0$. To calculate $\hat{\phi}$, we first calculate $\hat{\phi}(0) = \int_{\mathbb{R}} e^{-ax^2/2}\,dx$. We recall that

$$\int_{\mathbb{R}} e^{-x^2}\,dx = \sqrt{\pi} \implies \hat{\phi}(0) = \int_{\mathbb{R}} e^{-ax^2/2}\,dx = \sqrt{\frac{2\pi}{a}}.$$

Moreover, noting that $x\phi(x) = xe^{-ax^2/2} = -\frac{1}{a}\phi'(x)$, we have

$$(\hat{\phi})'(\xi) = -i\widehat{x\phi}(\xi) = \frac{i}{a}\widehat{\phi'}(\xi) = \frac{i}{a}(i\xi\hat{\phi})(\xi) = -\frac{\xi}{a}\hat{\phi}(\xi).$$

Recalling that $y'(t) = -cty(t) \implies y(t) = y(0)e^{-ct^2/2}$, it follows that

$$\hat{\phi}(\xi) = \hat{\phi}(0)e^{-\xi^2/(2a)} = \sqrt{\frac{2\pi}{a}}e^{-\xi^2/(2a)}.$$

In particular, for $a = 1$ we see that $e^{-x^2/2}$ is an eigenfunction of the Fourier transform[14]. This computes the Fourier transform of Gaussians in one dimension.

Suppose now that $\phi : \mathbb{R}^n \to \mathbb{R}$ is a multivariable Gaussian given by $\phi(x) = e^{-\langle Ax,x\rangle/2}$, where A is a symmetric positive definite $n \times n$ matrix. We now want to compute

$$\hat{\phi}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x}e^{-\langle Ax,x\rangle/2}\,dx, \quad \xi \in \mathbb{R}^n.$$

We diagonalize $A = Q^{-1}DQ$, where D is diagonal and Q is orthogonal; note then that $\langle Ax, x\rangle = \langle DQx, Qx\rangle$. If we now let $y = Qx$ (then $dy = dx$ since Q has determinant ±1), the above integral becomes

$$\int_{\mathbb{R}^n} e^{-i\xi\cdot Q^{-1}Qx}e^{-\langle DQx,Qx\rangle/2}\,dx = \int_{\mathbb{R}^n} e^{-i\xi\cdot Q^{-1}y}e^{-\langle Dy,y\rangle/2}\,dy.$$

Note that we can write $\xi \cdot Q^{-1}y = Q\xi \cdot y$. If we let $\eta = Q\xi$, with $\eta = (\eta_1, \ldots, \eta_n)$, and let the diagonal values of D be $a_1, \ldots, a_n$ (note these are all positive), then the above integral becomes

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{-i\eta\cdot y}e^{-\left(\sum_{j=1}^n a_j y_j^2\right)/2}\,dy &= \prod_{j=1}^n \left(\int_{\mathbb{R}} e^{-i\eta_j y_j}e^{-a_j y_j^2/2}\,dy_j\right) \\
&= \prod_{j=1}^n \left(\sqrt{\frac{2\pi}{a_j}}e^{-\eta_j^2/(2a_j)}\right) \\
&= \frac{(2\pi)^{n/2}}{\left(\prod_{j=1}^n a_j\right)^{1/2}}e^{-\left(\sum_{j=1}^n \frac{\eta_j^2}{a_j}\right)/2}.
\end{aligned}$$

[14]Under different conventions, the exact choice of Gaussian that ends up being an eigenfunction may differ, but it will always be the case that some Gaussian is an eigenfunction.

Note that $\prod_{j=1}^n a_j = \det A$, and $\sum_{j=1}^n \frac{\eta_j^2}{a_j} = \langle D^{-1}\eta, \eta\rangle = \langle D^{-1}Q\xi, Q\xi\rangle = \langle A^{-1}\xi, \xi\rangle$. It follows that we can write

$$\hat{\phi}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x} e^{-\langle Ax,x\rangle/2}\,dx = \frac{(2\pi)^{n/2}}{(\det A)^{1/2}} e^{-\langle A^{-1}\xi,\xi\rangle/2}.$$

Example 4.11. Consider the distribution given by the constant function $1 \in C^\infty(\mathbb{R}^n)$. This does not belong to L^1 or L^2, so we need to compute its Fourier transform in the sense of distributions. Thus, we consider the distribution

$$(\hat{1}, \phi) = (1, \hat{\phi}) = \int_{\mathbb{R}^n} \hat{\phi}(\xi)\,d\xi.$$

By the Fourier inversion formula, the right-hand side equals $(2\pi)^n\phi(0)$ (since $1 = e^{i(0\cdot\xi)}$). It follows that $(\hat{1}, \phi) = (2\pi)^n\phi(0) \implies \hat{1} = (2\pi)^n\delta_0$. Similar logic yields $\widehat{e^{i\xi_0\cdot x}} = (2\pi)^n\delta_{\xi_0}$ for any $\xi_0 \in \mathbb{R}^n$.

Another way to compute the Fourier transform is by approximating the distribution by Schwartz functions and then take the limit (in the sense of distributions): this works because the Fourier transform is continuous as a map $\mathcal{S}' \to \mathcal{S}'$. As such, note that for $\epsilon > 0$, the Gaussians $e^{-\epsilon|x^2|/2}$ converge to 1 *in the space of distributions*, meaning that $\lim_{\epsilon\to 0^+}(e^{-\epsilon|x|^2/2}, \phi) = (1, \phi)$ for all $\phi \in \mathcal{S}(\mathbb{R}^n)$. Hence, the Fourier transforms of $e^{-\epsilon|x|^2/2}$ should also converge to the Fourier transform of 1. From the previous example, we have

$$\widehat{e^{-\epsilon|x|^2/2}}(\xi) = \left(\frac{2\pi}{\epsilon}\right)^{n/2} e^{-|\xi|^2/(2\epsilon)}.$$

We note the following about the family of functions on the RHS:

- The integral of the RHS equals $(2\pi)^n$ for all ϵ.
- As $\epsilon \to 0$, the RHS converges, uniformly outside any neighborhood of the origin, to zero.

These are enough to guarantee that $\widehat{e^{-\epsilon|x|^2/2}} \to (2\pi)^n\delta_0$ in $\mathcal{S}'(\mathbb{R}^n)$ (cf. Problem 4 on HW 1).

***The following was not covered during lecture but may be of interest for some students*:**

One technique often used in computing Fourier transforms of distributions is to consider analytic families of distributions:

Definition 4.12. Let $U \subset \mathbb{C}$ be open, and let $\{u_z\}_{z\in U}$ be a collection of tempered distributions in $\mathcal{S}'(\mathbb{R}^n)$ indexed by U. We say that $\{u_z\}_{z\in U}$ is an *analytic* family of distributions on U if, for any $\phi \in \mathcal{S}(\mathbb{R}^n)$, the function

$$U \ni z \mapsto (u_z, \phi) \in \mathbb{C}$$

is a complex analytic function on U.

Most operations we've defined so far preserve the property of a family of distributions being analytic; in particular the Fourier transform of an analytic family of distributions is also an analytic family of distributions, since for any $\phi \in \mathcal{S}(\mathbb{R}^n)$ the function $z \mapsto (\widehat{u_z}, \phi)$ is, by definition, the function $z \mapsto (u_z, \hat{\phi})$, which is analytic since $\hat{\phi} \in \mathcal{S}(\mathbb{R}^n)$.

Example 4.13. Consider the function $u(x) = e^{iax^2/2}$ on $\mathbb{R}$, with $a \in \mathbb{R}\backslash\{0\}$. This defines a tempered distribution since $|u| = 1$ on $\mathbb{R}$. How do we compute its Fourier transform?

The trick is to use our computations for Gaussians $e^{-ax^2/2}$, $a > 0$ from before, and use analyticity to extend our results when "a is complex". Concretely, we note that for $U = \{\text{Re } z > 0\}$, the family $\{e^{-zx^2/2}\}_{z\in U}$ is analytic (note that requiring Re $z > 0$ guarantees that $e^{-zx^2/2}$ is bounded and thus defines a tempered distribution). Thus, its Fourier transform is also analytic in U. Moreover, the family of distributions

$$\left\{\left(\frac{2\pi}{z}\right)^{1/2} e^{-\xi^2/(2z)}\right\}_{z\in U}$$

is *also* an analytic family of distributions (here the square root is well-defined on U and sends $\mathbb{R}_+$ to $\mathbb{R}_+$; concretely $(re^{i\theta}) = r^{1/2}e^{i\theta/2}$ for $r > 0$, $-\pi/2 < \theta < \pi/2$), and $\left(\frac{2\pi}{z}\right)^{1/2} e^{-\xi^2/(2z)} = \mathcal{F}(e^{-zx^2/2})$ when $z \in \mathbb{R}_+$. Thus by analytic continuation the two families must agree for all $z \in U$, i.e.

$$\mathcal{F}(e^{-zx^2/2}) = \left(\frac{2\pi}{z}\right)^{1/2} e^{-\xi^2/(2z)} \text{ for all } z \text{ with Re } z > 0.$$

This does not quite give us our result, since we'd like to plug in $z = -ia$, which is not in this open set. Nonetheless, we note that $-ia+\epsilon \in U$ for $\epsilon > 0$, with $-ia+\epsilon \to -ia$ as $\epsilon \to 0^+$. Hence the Fourier transform of $e^{-(-ia+\epsilon)x^2/2}$ approaches that of $e^{iax^2/2}$, and hence

$$\mathcal{F}(e^{iax^2/2}) = \lim_{\epsilon\to 0^+}\left(\left(\frac{2\pi}{-ia+\epsilon}\right)^{1/2} e^{-\xi^2/(2(-ia+\epsilon))}\right).$$

The only subtlety in evaluating the limit on the RHS is the square root:

- If $a > 0$, then $\frac{2\pi}{-ia+\epsilon} \to \frac{2\pi}{|a|}i = \frac{2\pi}{|a|}e^{i\pi/2}$. Hence
$$\lim_{\epsilon\to 0^+}\left(\frac{2\pi}{-ia+\epsilon}\right)^{1/2} = \left(\frac{2\pi}{|a|}\right)^{1/2} e^{i\pi/4}.$$
- If $a < 0$, then $\frac{2\pi}{-ia+\epsilon} \to -\frac{2\pi}{|a|}i = \frac{2\pi}{|a|}e^{-i\pi/2}$. Hence
$$\lim_{\epsilon\to 0^+}\left(\frac{2\pi}{-ia+\epsilon}\right)^{1/2} = \left(\frac{2\pi}{|a|}\right)^{1/2} e^{-i\pi/4}.$$

Putting it altogether, we obtain

$$\mathcal{F}(e^{iax^2/2}) = \left(\frac{2\pi}{|a|}\right)^{1/2} e^{i\frac{\pi}{4}\text{sgn } a} e^{-i\xi^2/(2a)}.$$

Similarly, if A is a real symmetric non-singular $n \times n$ matrix, then similar arguments in Example 4.10 gives

$$\mathcal{F}(e^{i\langle Ax,x\rangle/2}) = \frac{(2\pi)^{n/2}}{|\det A|^{1/2}} e^{i\frac{\pi}{4}\text{sgn } A} e^{-i\langle A^{-1}\xi,\xi\rangle/2},$$

where sgn A is the sum of the sign of its eigenvalues, i.e the number of positive eigenvalues minus the number of negative eigenvalues.

Next Lecture*: Introduction to parabolic equations.*

5. Lecture 05 (04/12): Heat equation

The material in the next few weeks will be based on the textbook *Partial Differential Equations* by Evans [Eva10]. The material in the first half of this lecture is based on Section 2.3 of [Eva10].

5.1. Introduction and Motivation.

The heat equation is the differential equation

$$\partial_t u - \Delta u = 0.$$

In most problems regarding the heat equation, the differential equation is supplemented with additional conditions. The most common is the *Cauchy problem*, where $u|_{t=0}$ is specified:

$$\partial_t u - \Delta u = 0 \text{ in } (0,\infty)_t \times \mathbb{R}^n_x, \quad u(0,x) = f(x) \text{ with } f(x) \text{ specified.}$$

The heat equation models diffusion (e.g. of temperature). The idea is that the rate of change of a diffusive quantity should be given by $\partial_t u = \operatorname{div} F$, where F is the "flux" vector field. In some cases, we can model F as being proportional to the gradient of u, i.e. $F = A\nabla u$ (where A could be a constant or even a matrix). Then $\partial_t u = \operatorname{div}(A\nabla u)$; in the special case that $A = 1$, we get $\partial_t u = \Delta u$, i.e. the heat equation.

The heat equation is also related to Brownian motion. Let W^t_x denote a Brownian motion in $\mathbb{R}^n$ at t starting at x; this is a random process. Suppose, for $f : \mathbb{R}^n \to \mathbb{R}$, that we wanted to know the expected value of $f(W^t_x)$. It turns out that

$$u(t,x) := \mathbb{E}[f(W^t_x)] \implies \partial_t u - \frac{1}{2}\Delta u = 0 \text{ in } (0,\infty)_t \times \mathbb{R}^n_x, \quad u(0,x) = f(x),$$

i.e. u solves the Cauchy problem for the (rescaled) heat equation. See [Law10] for more details about Brownian motion.

5.2. Solving the Cauchy Problem.

Suppose for convenience that we seek a smooth solution $u(t,x)$ with initial data $f(x) \in \mathcal{S}(\mathbb{R}^n)$. In fact, we can obtain a solution u such that $u(t,\cdot) \in \mathcal{S}'(\mathbb{R}^n)$. We do so using the Fourier transform in $\mathbb{R}^n_x$. If we let $\hat{u}(t,\xi)$ denote the Fourier transform of $u(t,\cdot)$, then

$$\partial_t u(t,x) - \Delta(t,x) = 0 \implies \partial_t \hat{u}(t,\xi) + |\xi|^2 \hat{u}(t,\xi) = 0$$

since $\widehat{\partial_{x_j} u} = i\xi_j \hat{u}$. Thus, if we take the above equation, fix a value of $\xi \in \mathbb{R}^n$, and try to solve the resulting ODE in t, we get

$$\partial_t \hat{u}(t,\xi) = -|\xi|^2 \hat{u}(t,\xi) \implies \hat{u}(t,\xi) = e^{-|\xi|^2 t}\hat{u}(0,\xi).$$

Recalling that we have the Cauchy problem where we prescribed $u(0,x) = f(x)$, it follows that we'd need $\hat{u}(0,\xi) = \hat{f}(\xi)$. It follows that

$$\hat{u}(t,\xi) = e^{-|\xi|^2 t}\hat{f}(\xi) \implies u(t,x) = \mathcal{F}^{-1}_x \hat{u}(t,\cdot) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} e^{-t|\xi|^2} \hat{f}(\xi)\, d\xi.$$

There is another way to represent this: recalling that $\widehat{u * v} = \widehat{u}\widehat{v}$, it follows that

$$\hat{u} = e^{-|\xi|^2 t}\hat{f} = \widehat{\mathcal{F}^{-1}(e^{-|\xi|^2 t})}\hat{f} \implies u = \mathcal{F}^{-1}(e^{-|\xi|^2 t}) * f.$$

We now use the calculations from last lecture[15] to see that, for $t > 0$, we have

$$\mathcal{F}^{-1}(e^{-t|\xi|^2}) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/(4t)}.$$

Thus, for $H(t,x) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/(4t)}$, it follows that for $t > 0$ we have

(1)
$$u(t,x) = (H(t,\cdot) * f)(x) = \int_{\mathbb{R}^n} H(t, x-y) f(y)\, dy = \int_{\mathbb{R}^n} \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/(4t)} f(y)\, dy.$$

The function $H(t,x)$ is called the *heat* kernel. *A priori* this gives just *a solution* u to the heat equation with initial condition f, though we will show that, given some mild regularity conditions on u, that this is the unique such solution.

Note that for $t > 0$ we have the following about $H(t,x)$:

(1) $H(t,x) > 0$ for all x, and $\int_{\mathbb{R}^n} H(t,x)\, dx = 1$.
(2) $H(t,x)$ is smooth in both t and x (for $t > 0$).
(3) $|H(t,x)| \leq (4\pi t)^{-n/2}$ for all x.

Theorem 5.1. *For $u(t,x) = (H(t,\cdot) * f)(x)$ with $f \in \mathcal{S}(\mathbb{R}^n)$ (i.e. the solution of the heat equation obtained in (1)) and $t > 0$, we have:*

(1) $\int_{\mathbb{R}^n} u(t,x)\, dx = \int_{\mathbb{R}^n} f(x)\, dx$.
(2) $u(t,x)$ *is smooth in* t *and* x.
(3) $\sup_{x \in \mathbb{R}^n} |u(t,x)| \leq (4\pi)^{-n/2} \|f\|_{L^1(\mathbb{R}^n)} t^{-n/2}$.

Proof. (1) This follows from Observation 1 above by noting that for any $f, g \in L^1(\mathbb{R}^n)$ we have

$$\begin{aligned}
\int_{\mathbb{R}^n} (f * g)(x)\, dx &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y) g(y)\, dy \right) dx \\
&= \int_{\mathbb{R}^n} g(y) \left(\int_{\mathbb{R}^n} f(x-y)\, dx \right) dy \\
&= \left(\int_{\mathbb{R}^n} f(x)\, dx \right) \left(\int_{\mathbb{R}^n} g(y)\, dy \right).
\end{aligned}$$

(2) This follows from Observation 2 above, since in a convolution one can differentiate on either factor.
(3) This follows from Observation 3 above and writing out the convolution as an integral.

$\square$

[15]In more detail: we note that $e^{-t|\xi|^2} = e^{-\langle A^{-1}\xi, \xi\rangle/2}$ for $A^{-1} = 2t\mathrm{Id} \implies A = (2t)^{-1}\mathrm{Id}, \det A = (2t)^{-n}$. It follows that

$$\widehat{e^{-|x|^2/(4t)}} = \widehat{e^{-\langle Ax, x\rangle/2}} = \frac{(2\pi)^{n/2}}{(\det A)^{1/2}} e^{-\langle A^{-1}\xi,\xi\rangle/2} = (4\pi t)^{n/2} e^{-t|\xi|^2},$$

i.e. $\mathcal{F}^{-1}(e^{-t|\xi|^2}) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/(4t)}$.

Remark 7. The first observation can also be obtained using the equation for the Fourier transform of u: indeed, note that $\hat{u}(t,0)$ equals the integral of $u(t,\cdot)$, and from the formula for the Fourier transform, we see that $\hat{u}(t,0) = e^{-|\xi|^2\cdot 0}\hat{u}(0,0) = \hat{u}(0,0)$ is constant in t.

We can also study how $H(t,x)$ behaves as $t \to 0^+$, in hopes of studying how $u(t,x)$ behaves as $t \to 0^+$; ideally we want $u(t,x) \to f(x)$ in order to satisfy the initial condition in the Cauchy problem. We note that if we study the pointwise convergence behavior of $H(t,x)$ that the behavior depends on whether we fix x at the origin or away from the origin:

- If $x \neq 0$, then $\lim_{t\to 0^+} H(t,x) = 0$. Indeed, even though there is a factor of $t^{-n/2}$ which could *a priori* blow up as $t \to 0$, that factor is tempered by the exponential factor $e^{-|x|^2/(4t)}$ which decays *extremely fast*, notably faster than any inverse polynomial as $t \to 0^+$ (since $x \neq 0$).
- On the other hand, for $x = 0$ we have $H(0,x) = (4\pi t)^{-n/2} \to \infty$ as $t \to 0^+$.

This behavior can be described more precisely as follows:

Theorem 5.2. *Viewing $H(t,\cdot) \in \mathcal{S}'(\mathbb{R}^n)$ for $t > 0$, we have $H(t,\cdot) \to \delta$ as $t \to 0^+$. Moreover, for $u(t,\cdot) = (H(t,\cdot)) * f$, we have $u(t,x) \to f(x)$ as $t \to 0^+$, uniformly in x if $f \in \mathcal{S}(\mathbb{R}^n)$.*

Proof. This is easiest to see using the Fourier transform: we note that $\hat{H}(t,\xi) = e^{-|\xi|^2 t} \to 1$ in $\mathcal{D}'(\mathbb{R}^n)$ as $t \to 0^+$, so by the continuity of the (inverse) Fourier transform we have $H(t,x) = \mathcal{F}^{-1}\hat{H}(t,\cdot) \to \mathcal{F}^{-1}(1) = \delta$ in $\mathcal{S}'(\mathbb{R}^n)$. For $u(t,\cdot) = (H(t,\cdot)) * f$, we have $\hat{u}(t,\xi) = e^{-|\xi|^2 t}\hat{f}(\xi)$, from which we see that $\hat{u}(t,\cdot) \to \hat{f}$ in $L^1(\mathbb{R}^n)$ by the Dominated Convergence Theorem since $\hat{f} \in L^1(\mathbb{R}^n)$ if $f \in \mathcal{S}(\mathbb{R}^n)$. It follows that

$$u(t,x) - f(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi}(\hat{u}(t,\xi) - \hat{f}(\xi))\, d\xi \leq (2\pi)^{-n}\|\hat{u}(t,\cdot) - \hat{f}\|_{L^1(\mathbb{R}^n)}$$

for any $x \in \mathbb{R}^n$, and the latter quantity converges to 0 (independent of x) as $t \to 0^+$, as desired. ☐

5.3. **Fundamental Solutions.** We now take a small diversion to discuss fundamental solutions to differential operators.

Definition 5.3. Let P be a *constant-coefficient* differential operator on $\mathbb{R}^n$. We say that $E \in \mathcal{D}'(\mathbb{R}^n)$ is a *fundamental solution* of P if $PE = \delta$, where δ is the Dirac delta on $\mathbb{R}^n$.

Fundamental solutions play an important role for solving PDEs due to properties of convolution: recall that $u * \delta = u$ for any $u \in \mathcal{D}'(\mathbb{R}^n)$, and $P(u_1 * u_2) = Pu_1 * u_2 = u_1 * (Pu_2)$, assuming the convolution $u_1 * u_2$ is defined. As such, suppose we have a fundamental solution E, and we have $f \in \mathcal{D}'(\mathbb{R}^n)$ such that the convolution $u = f * E$ makes sense (for example if f is compactly supported). Then we have

$$Pu = P(f * E) = f * (PE) = f * \delta = f.$$

Thus, we have easily constructed (a) solution to the equation $Pu = f$ by taking $u = f * E$.

Example 5.4. Let $P = \frac{d}{dx}$ on $\mathbb{R}$. Then $E(x) = H(x)$, the Heaviside function, is a fundamental solution for P. Any other fundamental solution is of the form $H(x) + c$ for some $c \in \mathbb{R}$ (a notable example is $c = -1$, which gives $H(x) - 1 = \begin{cases} -1 & x \leq 0 \\ 0 & x > 0 \end{cases}$). Note then that if f is a compactly supported continuous function on $\mathbb{R}$, then the convolution

$$u(x) = (f * H)(x) = \int_{\mathbb{R}} H(x-y) f(y)\, dy = \int_{\mathbb{R}} \mathbb{1}_{x \geq y} f(y)\, dy = \int_{-\infty}^{x} f(y)\, dy$$

solves $\frac{d}{dx} u = f$ in the sense of distributions. This just follows from the Fundamental Theorem of Calculus.

Example 5.5. Let $P = \Delta = \sum_{j=1}^{n} \partial_{x_j}^2$ be the Laplacian on $\mathbb{R}^n$. Then

$$E(x) = \begin{cases} c_2 \log(|x|) & n = 2 \\ c_n |x|^{2-n} & n \neq 2 \end{cases}$$

gives a fundamental solution to Δ, for some constants c_n. (Note that these are all locally integrable functions, and hence define distributions.)

In the case $n = 3$, interpreting u as electric potential, we then have $\Delta u = -\rho/\epsilon_0$ where ρ is the charge density (which can be interpreted as a distribution, particularly if the charges are viewed as point charges, i.e. Dirac deltas). Then u can be recovered from ρ by

$$u(x) = (-\rho/\epsilon_0) * (c_3 |x|^{-1}).$$

If ρ is a continuous distribution, then $u(x) = \int_{\mathbb{R}^3} \frac{\rho(y)}{4\pi\epsilon_0 |x-y|}\, dy$, whereas if ρ is a point charge with charge Q say at the origin, then $u(x) = \frac{Q}{4\pi\epsilon_0 |x|}$. Both of these are variants of Coulomb's law.

Definition 5.6. Suppose one of the variables in $\mathbb{R}^n$ is denoted t. For a constant-coefficient differential operator P on $\mathbb{R}^n$, a fundamental solution E of P is called *forward* if supp $E \subset \{t \geq 0\}$. It is called *backward* if supp $E \subset \{t \leq 0\}$.

Example 5.7. For $P = \partial_t$ on $\mathbb{R}$, the Heaviside function $H(t)$ is a forward fundamental solution.

What can we do with this forward fundamental solution? Suppose we wanted to solve the *inhomogeneous Cauchy problem* for P, which in this case just means solving

$$u'(t) = f(t) \text{ for } t > 0, u(0) = u_0 \in \mathbb{C}.$$

One convoluted way to obtain a solution is to study the distribution $u(t)\mathbb{1}_{t \geq 0}$. Note that we are allowed to convolve this distribution with $H(t)$, since both distributions are supported in $\overline{\mathbb{R}_+}$. Then, using various properties of convolutions, we have

$$\begin{aligned} u(t)\mathbb{1}_{t\geq 0} = (u(t)\mathbb{1}_{t\geq 0}) * \delta &= (u(t)\mathbb{1}_{t\geq 0}) * \partial_t H(t) \\ &= \partial_t \left[(u(t)\mathbb{1}_{t\geq 0}) * H(t) \right] \\ &= (\partial_t (u(t)\mathbb{1}_{t\geq 0})) * H(t) \\ &= (\partial_t u(t)\mathbb{1}_{t\geq 0}) * H(t) + (u(t)\partial_t \mathbb{1}_{t\geq 0}) * H(t) \\ &= (f\mathbb{1}_{t\geq 0}) * H + (u(t)\delta_0) * H(t) \\ &= (f\mathbb{1}_{t\geq 0}) * H + u(0)\delta_0 * H = (f\mathbb{1}_{t\geq 0}) * H + u_0 H. \end{aligned}$$

It follows that for $t > 0$ we should have

$$u(t) = ((f\mathbb{1}_{t\geq 0}) * H)(t) + u_0 = \int_0^t f(s)\, ds + u_0.$$

This is of course just the Fundamental Theorem of Calculus. However, this gives some motivation for how to tackle inhomogeneous problems for other operators, as long as we can obtain a fundamental solution.

Next Lecture*: Computing and applying the fundamental solution for the heat operator.*

6. Lecture 06 (04/14): The inhomogeneous heat equation and uniqueness

6.1. **The inhomogeneous heat equation.** We now return to the heat equation and ask: what is a fundamental solution for the heat operator, i.e. what distribution $E \in \mathcal{D}'(\mathbb{R}^{n+1}_{t,x})$ satisfies

$$\partial_t E - \Delta E = \delta_{(0,0)}?$$

For emphasis, we write $\delta_{(0,0)}$ to point out that this is the Dirac delta **of the origin in $\mathbb{R}^{n+1}_{t,x}$, i.e. of the space-time origin** (not just in space).

Putting aside the question for a moment, let's see how we can use such a hypothetical fundamental solution. In fact, let's suppose we have a *forward* fundamental solution E. Let's see if we can use it to solve the inhomogeneous heat equation

$$\partial_t u - \Delta u = f \text{ in } (0,\infty) \times \mathbb{R}^n, \quad u(0,x) = g(x).$$

Following the same trick we used in the case of $P = \partial_t$ on $\mathbb{R}$, we write[16]

$$\begin{aligned} u\mathbb{1}_{t\geq 0} = (u\mathbb{1}_{t\geq 0}) * \delta_{(0,0)} &= (u\mathbb{1}_{t\geq 0}) * (\partial_t - \Delta)E \\ &= (\partial_t - \Delta)\left[(u\mathbb{1}_{t\geq 0}) * E\right] \\ &= ((\partial_t - \Delta)(u\mathbb{1}_{t\geq 0})) * E. \end{aligned}$$

Here, the convolution is taken in $\mathbb{R}^{n+1}_{t,x}$, i.e. with respect to both the t and x variables. We split up the last term as follows:

$$\partial_t(u\mathbb{1}_{t\geq 0}) = (\partial_t u)\mathbb{1}_{t\geq 0} + u\delta_{t=0}$$

while

$$\Delta(u\mathbb{1}_{t\geq 0}) = (\Delta u)\mathbb{1}_{t\geq 0}$$

since the cutoff in t is independent of the spatial variables. It follows that

$$u\mathbb{1}_{t\geq 0} = ((\partial_t u - \Delta u)\mathbb{1}_{t\geq 0}) * E + (u\delta_{t=0}) * E = (f\mathbb{1}_{t\geq 0}) * E + (g\delta_{t=0}) * E.$$

Note that if E were smooth[17], then the convolutions can be evaluated pointwise for $t > 0$ to give the integrals

$$((f\mathbb{1}_{t\geq 0}) * E)(t,x) = \int_0^t \int_{\mathbb{R}^n} E(t-s, x-y) f(s,y)\, dy\, ds$$

and

$$(g\delta_{t=0}) * E = \int_{\mathbb{R}^n} E(t, x-y) g(y)\, dy.$$

---

[16] One subtlety is justifying why the convolutions are well-defined. It turns out that the key property here is the *forward* part of the fundamental solution: this guarantees that, when taking the convolution, the integral in t is over a bounded interval. If the fundamental solution has sufficient decay for large x, which *a posteriori* we can certainly arrange, then the convolution in the spatial variables x makes sense against reasonable distributions in space-time; the forward part of the fundamental solution then guarantees that the convolution in the t variable is over a compact region.

[17] E is definitely not everywhere smooth, given that a combination of its derivatives gives the Dirac delta distribution.

Hence, our solution $u(t,x)$ should be given by

$$u(t,x) = \int_0^t \int_{\mathbb{R}^n} E(t-s, x-y) f(s,y)\, dy\, ds + \int_{\mathbb{R}^n} E(t, x-y) g(y)\, dy.$$

In particular, if $f \equiv 0$, then u should be given by $u(t,x) = \int_{\mathbb{R}^n} E(t, x-y) g(y)\, dy = (E(t,\cdot) * g)(x)$. But we already know how to solve that problem, namely $u = H(t,\cdot) * g$. This suggests that our fundamental solution $E(t,\cdot)$ should be given by $H(t,\cdot)$, at least for $t > 0$. In addition, if E is a *forward* fundamental solution, then it should equal 0 for $t < 0$. This suggests considering

$$E(t,x) = \begin{cases} H(t,x) & t > 0 \\ 0 & t < 0 \end{cases} = \begin{cases} \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/(4t)} & t > 0 \\ 0 & t < 0 \end{cases}.$$

Note then that E is locally integrable, and hence defines a distribution.

Theorem 6.1. *With E defined above, we have that E is a forward fundamental solution for $P = \partial_t - \Delta$ on $\mathbb{R}^{n+1}_{t,x}$.*

Proof. The support properties are clear, so it suffices to verify that $(\partial_t - \Delta)E = \delta_{(0,0)}$. There are at least two ways to do so:

Method 1: By definition, this involves verifying that $(E, -\partial_t \phi - \Delta\phi) = \phi(0,0)$ for all $\phi \in C_c^\infty(\mathbb{R}^{n+1})$. This can be done by integrating $E \cdot (-\partial_t \phi - \Delta \phi)$ over $(\epsilon, \infty) \times \mathbb{R}^n$ for $\epsilon > 0$, integrating by parts and noting that E solves the heat equation as a smooth function in $\{t > \epsilon\}$, and taking $\epsilon \to 0^+$. This will be an exercise on the homework.

Method 2: Note that E is uniformly bounded outside compact subsets, so it in fact defines a tempered distribution. We can then calculate (cf. Problem 6 on HW 1) that

$$\widehat{E}(\xi, \tau) = \frac{1}{|\xi|^2 + i\tau}.$$

This implies that $\widehat{(\partial_t - \Delta)E}(\xi,\tau) = (i\tau + |\xi|^2)\widehat{E}(\xi,\tau) = 1 \implies (\partial_t - \Delta)E = \mathcal{F}^{-1}(1) = \delta_{(0,0)}$. $\square$

As such, we have

Theorem 6.2. *Suppose $f \in C_c^\infty(\mathbb{R}^{1+n})$ and $g \in \mathcal{S}(\mathbb{R}^n)$. Then*

(2)
$$u(t,x) = \int_0^t \int_{\mathbb{R}^n} \frac{1}{(4\pi(t-s))^{n/2}} e^{-|x-y|^2/(4(t-s))} f(s,y)\, dy\, ds + \int_{\mathbb{R}^n} \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/(4t)} g(y)\, dy$$

solves the inhomogeneous heat equation $(\partial_t - \Delta)u = f$ in $(0,\infty) \times \mathbb{R}^n$, $u(0,x) = g(x)$.

Remark 8. This process of using the solution operator to the homogeneous equation and convolving it (in time, not just in space) to solve the inhomogeneous equation is part of a general method called *Duhamel's principle*.

6.2. **Sobolev spaces and solutions with less regular initial data.**

Definition 6.3. For $s \in \mathbb{R}$, the (L^2-based) Sobolev space of order s on $\mathbb{R}^n$, denoted $H^s(\mathbb{R}^n)$, consists of all tempered distributions $u \in \mathcal{S}'(\mathbb{R}^n)$ such that $(1+|\xi|^2)^{s/2}\hat{u}$, *a priori* well-defined as a tempered distribution, is in fact in $L^2(\mathbb{R}^n)$. More colloquially, $u \in H^s(\mathbb{R}^n)$ if

$$\int_{\mathbb{R}^n} (1+|\xi|^2)^s |\hat{u}(\xi)|^2 \, d\xi < \infty.$$

The square root of the above integral is called the H^s norm.

For example, $H^0(\mathbb{R}^n) = L^2(\mathbb{R}^n)$, and for $k \in \mathbb{N}$, $H^k(\mathbb{R}^n)$ coincides with the space of $u \in L^2(\mathbb{R}^n)$ whose distributional derivatives $\partial^\alpha u$ are also in L^2, i.e. can be identified with an L^2 function, for all $|\alpha| \leq k$. The Dirac delta δ belongs to $H^s(\mathbb{R}^n)$ for any $s < -n/2$, as its Fourier transform is identically 1, and $(1+|\xi|^2)^s$ is integrable precisely when $s < -n/2$.

Some important facts:

- $H^s(\mathbb{R}^n) \subset C^0(\mathbb{R}^n)$ for $s > n/2$.
- $\partial^\alpha : H^s(\mathbb{R}^n) \to H^{s-|\alpha|}(\mathbb{R}^n)$.
- We have $\mathcal{S}(\mathbb{R}^n) \subset \cap_{s\in\mathbb{R}^n} H^s(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n)$.

(The third fact in fact follows from the first two.)

We now note that, for $g \in H^s(\mathbb{R}^n)$ for any $s \in \mathbb{R}$, the convolution $u(t,x) = (H(t,\cdot) * g)(x)$ gives a *smooth* solution to the homogeneous heat equation for $t > 0$, with $\lim_{t\to 0} u(t,\cdot) = g$ (say in the $H^s(\mathbb{R}^n)$ norm or in $\mathcal{S}'(\mathbb{R}^n)$). To see that u is smooth in $t > 0$, we can show that $u(t,\cdot) \in H^{s'}(\mathbb{R}^n)$ for any $s' \in \mathbb{R}$ (so that $u(t,\cdot) \in \cap_{s'} H^{s'}(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n)$, by noting that for any $s' \in \mathbb{R}$ we have

$$(1+|\xi|^2)^{s'/2}\hat{u}(t,\xi) = (1+|\xi|^2)^{s'/2} e^{-|\xi|^2 t}\hat{g}(\xi) = \left((1+|\xi|^2)^{(s'-s)/2} e^{-|\xi|^2 t}\right)(1+|\xi|^2)^{s/2}\hat{g}(\xi),$$

with the prefactor $(1+|\xi|^2)^{(s'-s)/2}e^{-|\xi|^2 t}$ uniformly bounded for *any* s, s' if $t > 0$ due to the Gaussian decay of the $e^{-|\xi|^2 t}$ factor. Thus $(1+|\xi|^2)^{s'/2}\hat{u}(t,\xi)$ is a bounded factor times $(1+|\xi|^2)^{s/2}\hat{g}$ which is in L^2 by assumption, i.e. $u(t,\cdot) \in H^{s'}(\mathbb{R}^n)$ for any s'. We thus summarize as follows:

Theorem 6.4. *For $g \in H^s(\mathbb{R}^n)$, there exists a solution $u(t,x)$ to the heat equation which is* smooth *in (t,x) for $t > 0$, such that $\lim_{t\to 0} u(t,\cdot) = g$ with respect to the H^s norm, namely given by $u(t,x) = (H(t,\cdot) * g)(x)$.*

Remark 9. The fact that u is immediately smooth for $t > 0$, even if the initial condition g is not smooth, is an effect called the *instantaneous smoothing effect* for the heat equation.

6.3. **Uniqueness.** We finally address an issue sidestepped so far, which is whether the solution to the heat equation we obtained via convolution with the fundamental solution is the *unique* solution to the heat equation. Having uniqueness allows us to conclude that the properties we derive with the explicit formula are meaningful for the equation as a whole, in the sense that there are no "other" solutions (that we want to consider) where our approach using the explicit formula doesn't work.

In general, a mild *a priori* assumption on the functions being considered will be necessary (see the end of this section for a counterexample without such restrictions). We present two situations, as follows.

If F is a topological vector space space, we can consider functions $f : I \to F$ where I is an interval in $\mathbb{R}$. For t_0 in the interior of I, we can say that f is *differentiable* at t_0 if the limit $\frac{f(t_0+h)-f(t_0)}{h}$ exists in F as $h \to 0$ (the derivative can then be defined as this limit). We then let $C^k(I; F)$ denote the space of functions $f : I \to F$ which are k times differentiable which are all continuous (w.r.t. the topology on F).

Theorem 6.5. *Let $f \in C_c^\infty(\mathbb{R}^{1+n})$ and $g \in L^1(\mathbb{R}^n)$. Then the function defined in* (2) *is the unique solution of the inhomogeneous heat equation*

$$(\partial_t - \Delta)u = f \text{ in } (0, \infty) \times \mathbb{R}^n, \quad u(0, x) = g(x)$$

among functions u in $C^1([0, \infty); L^1(\mathbb{R}^n))$.

Proof. It suffices to assume that $f = 0$ and $g = 0$, since for general f and g we see that if u_1 and u_2 both solve the same equation, then $v = u_1 - u_2$ solves $(\partial_t - \Delta)v = 0$, $v(0, x) = 0$. Now, since $u(t, \cdot)$ belongs in $L^1(\mathbb{R}^n)$ for all $t \geq 0$ and varies continuously differentiably in t, it follows that $\hat{u}(t, \xi)$ is continuous in ξ for each t, and it is C^1 in t for each ξ (essentially since the assumptions allow us to differentiate under the integral). Thus, we see that for **each** $\xi \in \mathbb{R}^n$ the Fourier transform must satisfy

$$\partial_t \hat{u}(t, \xi) = -|\xi|^2 \hat{u}(t, \xi), \quad \hat{u}(0, \xi) = 0.$$

By uniqueness of 1st-order ODE, it follows that we must have $\hat{u}(t, \xi) = 0$ for all (t, ξ), i.e. $u(t, \cdot) \equiv 0$ for each $t \geq 0$, as desired. $\square$

7. Lecture 07 (04/19): Maximum Principle and Regularity

7.1. **Maximum Principle.** *Recall*: Last lecture we derived a *uniqueness* result for the heat equation, using the Fourier transform.

A different approach of obtaining uniqueness involves the *maximum principle*. We consider a bounded set U and a bounded time interval $[0, T]$, and ask: if u solves the heat equation, where does u attain its maximum value on $\overline{U} \times [0, T]$?

Theorem 7.1 (Maximum Principle for Bounded Domains)**.** *Let* $u \in C^2((0, T] \times U) \cap C^0([0, T] \times \overline{U})$*, and suppose* $\partial_t u - \Delta u = 0$ *in* $(0, T) \times U$*. Then*

$$\max_{(t,x)\in[0,T]\times\overline{U}} u(t, x) = \max\left(\max_{(t,x)\in[0,T]\times\partial U} u(t, x), \max_{x\in\overline{U}} u(0, x)\right),$$

i.e. $([0, T] \times \partial U) \cap (\{0\} \times \overline{U})$ *contains a point which maximizes* u *over all of* $[0, T] \times \overline{U}$*.*

As a consequence, we have

Corollary 7.2. *If* $\partial_t u - \Delta u = 0$ *in* $(0, T) \times U$*,* $u(t, x) = 0$ *for all* $(t, x) \in [0, T] \times \partial U$*, and* $u(0, x) = 0$ *for all* $x \in U$*, then* $u \equiv 0$ *on* $\overline{U} \times [0, T]$*.*

This follows by applying the maximum principle to both u and $-u$ to conclude that $\max_{[0,T]\times\overline{U}}(\pm u) = 0$.

Proof of Maximum Principle. For $\epsilon > 0$, let $u_\epsilon(t, x) = u(t, x) + \epsilon|x|^2$. Then

$$\partial_t u_\epsilon - \Delta u_\epsilon = \partial_t u - \Delta u - 2n\epsilon = -2n\epsilon < 0$$

in $(0, T) \times U$. I now claim that the maximum principle holds for u_ϵ, i.e. that

$$\max_{(t,x)\in[0,T]\times\overline{U}} u_\epsilon(t, x) = \max\left(\max_{(t,x)\in[0,T]\times\partial U} u_\epsilon(t, x), \max_{x\in\overline{U}} u_\epsilon(0, x)\right).$$

To see this, suppose (t_0, x_0) maximized u_ϵ in $[0, T] \times \overline{U}$. If $(t_0, x_0) \in (0, T) \times U$ (i.e. in the interior in both space and time), then necessarily we must have $\partial_t u_\epsilon(t_0, x_0) = 0$ (interior critical point), and $\Delta u_\epsilon(t_0, x_0) \le 0$ (since the Hessian $D^2 u_\epsilon$ cannot have any positive eigenvalues due to being an interior local minimum, and $\Delta u_\epsilon = \mathrm{tr}\ D^2 u_\epsilon$). In particular this would give $(\partial_t u_\epsilon - \Delta u_\epsilon)(t_0, x_0) \ge 0$, contradicting the calculation that $\partial_t u_\epsilon - \Delta u_\epsilon < 0$ in $(0, T) \times U$. Similarly, if $t_0 = T$ and $x_0 \in U$, then we necessarily must have $\partial_t u(t_0, x_0) \ge 0$ (otherwise there is a larger value for smaller t), and $\Delta u_\epsilon(t_0, x_0) \le 0$ (for the same reasons as above), so we'd still have $(\partial_t u_\epsilon - \Delta u_\epsilon)(t_0, x_0) \ge 0$, a contradiction. It follows that either $x_0 \in \partial U$ or $t_0 = 0$, thus showing the maximum principle holds for u_ϵ.

Since U is bounded, we have that $C = \max_{\overline{U}} |x|^2$ is finite. In particular, note that

$$u(t, x) \le u_\epsilon(t, x) \le u(t, x) + \epsilon C$$

for all $(t, x) \in [0, T] \times \overline{U}$. Thus, we have

$$\max_{(t,x)\in[0,T]\times\overline{U}} u(t, x) \le \max_{(t,x)\in[0,T]\times\overline{U}} u_\epsilon(t, x),$$

while

$$\max_{(t,x)\in\partial U\times[0,T]}(u_\epsilon(t,x)) \leq \max_{(t,x)\in\partial U\times[0,T]}(u(t,x)) + \epsilon C,$$

and similarly

$$\max_{x\in\overline{U}} u_\epsilon(0,x) \leq \max_{x\in\overline{U}} u(0,x) + \epsilon C.$$

It follows that

$$\max_{(t,x)\in[0,T]\times\overline{U}} u(t,x) \leq \max\left(\max_{(t,x)\in[0,T]\times\partial U]} u(t,x), \max_{x\in\overline{U}} u(0,x)\right) + \epsilon C.$$

Taking $\epsilon \to 0^+$ gives the desired statement. □

Remark 10. This formulation is sometimes called the *weak* maximum principle. A *strong* maximum principle also holds, which gives that the maximum *cannot* be attained in the interior unless the function is constant. For instance, the proof above shows that a function satisfying $\partial_t u - \Delta u < 0$ satisfies the strong maximum principle, but it is *a priori* not clear that the same holds passing to the limit (in this case it is true).

When the spatial domain U is unbounded, the maximum principle still holds with mild assumptions:

Theorem 7.3 (cf. [Eva10] Section 2.3, Theorem 6)**.** *Suppose $u \in C^2((0,T]\times\mathbb{R}^n)\times C^0([0,T]\times\mathbb{R}^n)$ solves $\partial_t u - \Delta u = 0$ in $(0,T)\times\mathbb{R}^n$, with $u(0,x) = g(x) \in L^\infty(\mathbb{R}^n)$. Suppose as well that u satisfies the growth estimate*

$$u(t,x) \leq Ae^{a|x|^2} \text{ for all } x \in \mathbb{R}^n, t \in [0,T]$$

for some constants $A, a > 0$. Then

$$\sup_{[0,T]\times\mathbb{R}^n} u = \sup_{\mathbb{R}^n} g.$$

Proof sketch. Suppose first that T is small enough to satisfy $4aT < 1$. Then, for any $\mu > 0$ and any sufficiently small $\epsilon > 0$, the function

$$v_\mu(t,x) = u(t,x) - \frac{\mu}{(T+\epsilon-t)^{n/2}} e^{|x-y|^2/(4(T+\epsilon-t))}$$

also solves the heat equation by direct computation and *is sufficiently negative* (notably less than $\sup_{\mathbb{R}^n} g$) for sufficiently large values of $|x|$ independent of t due to the growth estimate on u (the required lower bound depends on a, A, T, ϵ, and μ). We can then show that $\sup_{[0,T]\times\mathbb{R}^n} v_\mu \leq \sup_{\mathbb{R}^n} g$, by combining the maximum principle on a sufficiently large ball in $\mathbb{R}^n$ and the fact that v_μ was constructed to be less than $\sup_{\mathbb{R}^n} g$ outside large enough balls. Letting $\mu \to 0$ then gives the desired result. If $4aT \geq 1$, split up $[0,T]$ into subintervals whose lengths are less than $1/(4a)$, and iterate the argument on each subinterval. □

Corollary 7.4. *For $g \in C(\mathbb{R}^n)$ and $f \in C([0,T]\times\mathbb{R}^n)$, there is at most one solution $u \in C^2((0,T]\times\mathbb{R}^n)\cap C([0,T]\times\mathbb{R}^n)$ to the inhomogeneous heat equation*

$$u_t - \Delta u = f \text{ in } (0,T)\times\mathbb{R}^n, \quad u(0,x) = g(x) \text{ on } \mathbb{R}^n.$$

Remark 11. Without the growth condition $u(t,x) \le Ae^{a|x|^2}$, one can obtain examples of non-uniqueness. In particular, there exists a non-zero smooth function $u(t,x)$ solving the heat equation with *zero* initial data. Indeed, the power series

$$u(t,x) = \sum_{n=0}^{\infty} g^{(n)}(t)\frac{x^{2n}}{(2n)!}$$

for $g : \mathbb{R} \to \mathbb{R}$ smooth formally solves the heat equation; by choosing g appropriately (e.g. $g(t) = e^{-1/t}H(t)$) one can arrange for the above series to converge and vanish only for $t \le 0$.

7.2. **Regularity.** We now consider parabolic regularity estimates. These amount to asking: if we know the "right-hand sides" of the inhomogeneous heat equation (i.e. given a solution u we know $f = \partial_t u - \Delta u$ and $g(x) = u(0,x)$), can we estimate derivatives of u by corresponding derivatives of the right-hand sides?

Theorem 7.5. *Suppose u is a smooth solution to*

$$\partial_t u - \Delta u = f \text{ in } (0,T] \times \mathbb{R}^n, u|_{t=0} = g,$$

with $f(t,\cdot), g \in L^2(\mathbb{R}^n)$ for all $0 < t \le T$, and $u(t,\cdot)$ decays sufficiently quickly for all t. Then, there exist constants C_T, C_T', C_T'' such that:

(1) $\max_{0\le t\le T} \|u(t)\|_{L^2(\mathbb{R}^n)} \le C_T \left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{L^2(\mathbb{R}^n)}\right)$.
(2) $\|u\|_{L^2([0,T];H^1(\mathbb{R}^n))} \le C_T' \left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{L^2(\mathbb{R}^n)}\right)$.
(3) $\|\partial_t u\|_{L^2([0,T];H^{-1}(\mathbb{R}^n))} \le C_T'' \left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{L^2(\mathbb{R}^n)}\right)$.

Remark 12. In [Eva10] (e.g. Theorem 2 in Section 7.1.2), this is stated as a single estimate on the sum of the three items above, i.e.

$$\begin{aligned}&\left(\max_{0\le t\le T} \|u(t)\|_{L^2(\mathbb{R}^n)}\right) + \|u(t)\|_{L^2([0,T];H^1(\mathbb{R}^n))} + \|\partial_t u(t)\|_{L^2([0,T];H^{-1}(\mathbb{R}^n))}\\ &\le C\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{L^2(\mathbb{R}^n)}\right)\end{aligned}$$

for some C. This is equivalent to obtaining each of the three estimates above separately. The single estimate is certainly a more concise way to state the result, but it may be more useful to think about each estimate separately.

Proof. (1) Multiplying the differential equation by u and integrating in x gives

$$\int_{\mathbb{R}^n} u(t,x)u_t(t,x) - u(t,x)\Delta u(t,x)\,dx = \int_{\mathbb{R}^n} f(t,x)u(t,x)\,dx.$$

We now note that $uu_t = \frac{1}{2}\partial_t(|u|^2)$, so

$$\int_{\mathbb{R}^n} u(t,x)u_t(t,x)\,dx = \int_{\mathbb{R}^n} \frac{1}{2}\partial_t(|u(t,x)|^2)\,dx = \partial_t\left(\frac{1}{2}\|u(t,\cdot)\|^2_{L^2(\mathbb{R}^n)}\right).$$

On the other hand, we have *Green's first identity*: for any u, v smooth and any bounded open subset U we have

$$\begin{aligned}\int_U u\Delta v\,dx = \int_U u\mathrm{div}\ (\nabla v)\,dx &= \int_U \mathrm{div}\ (u\nabla v) - \nabla u\cdot\nabla v\,dx \\ &= \int_{\partial U} u(\nabla v\cdot\nu)\,dS - \int_U \nabla u\cdot\nabla v\,dx.\end{aligned}$$

In particular, if u and v decay sufficiently quickly at infinity, so that the boundary integral can be made arbitrarily small by taking U to be a sufficiently large ball, we have

$$\int_{\mathbb{R}^n} u\Delta v\,dx = -\int_{\mathbb{R}^n} \nabla u\cdot\nabla v\,dx.$$

Thus, we have

$$\frac{1}{2}\partial_t\left(\|u(t,\cdot)\|^2_{L^2(\mathbb{R}^n)}\right) + \int_{\mathbb{R}^n} |\nabla u(t,x)|^2\,dx = \int_{\mathbb{R}^n} f(t,x)u(t,x)\,dx. \tag{3}$$

In particular,

$$\begin{aligned}\frac{1}{2}\partial_t\left(\|u(t,\cdot)\|^2_{L^2(\mathbb{R}^n)}\right) &\le \int_{\mathbb{R}^n} f(t,x)u(t,x)\,dx \\ &\le \|u(t)\|_{L^2(\mathbb{R}^n)}\|f(t)\|_{L^2(\mathbb{R}^n)} \\ &\le \frac{1}{2}\|u(t)\|^2_{L^2(\mathbb{R}^n)} + \frac{1}{2}\|f(t)\|^2_{L^2(\mathbb{R}^n)}.\end{aligned}$$

Recall:

Lemma 7.6 (Gronwall's inequality)**.** *Suppose $y(t)$ satisfies the differential inequality*

$$y'(t) \le a(t)y(t) + b(t)$$

for some functions $a(t)$, $b(t)$. Then

$$y(t) \le e^{\int_0^t a(s)\,ds}\left(y(0) + \int_0^t e^{-\int_0^s a(r)\,dr}b(s)\,ds\right).$$

Applying Gronwall's inequality to $y(t) = \|u(t)\|^2_{L^2(\mathbb{R}^n)}$ thus yields

$$\begin{aligned}\|u(t)\|^2_{L^2(\mathbb{R}^n)} &\le e^t\left(\|u(0)\|^2_{L^2(\mathbb{R}^n)} + \int_0^t e^{-s}\|f(s)\|^2_{L^2(\mathbb{R}^n)}\,ds\right) \\ &\le e^T\left(\|g\|^2_{L^2(\mathbb{R}^n)} + \int_0^T \|f(s)\|^2_{L^2(\mathbb{R}^n)^2}\,ds\right) \\ &= e^T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{L^2(\mathbb{R}^n)}\right).\end{aligned}$$

From the inequality $(a^2+b^2)^{1/2} \le (a+b)/\sqrt{2}$, we see that

$$\max_{0\le t\le T}\|u(t)\|_{L^2(\mathbb{R}^n)} \le C_T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{L^2(\mathbb{R}^n)}\right)$$

holds with $C_T = e^{T/2}/\sqrt{2}$.

(2) From (3), integrating from 0 to T yields

$$\frac{1}{2}\left(\|u(T)\|^2_{L^2(\mathbb{R}^n)}-\|u(0)\|^2_{L^2(\mathbb{R}^n)}\right)+\int_0^T\int_{\mathbb{R}^n}|\nabla u(t,x)|^2\,dx\,dt=\int_0^T\int_{\mathbb{R}^n}f(t,x)u(t,x)\,dx\,dt.$$

Thus

$$\begin{aligned}\|u\|^2_{L^2([0,T];H^1(\mathbb{R}^n))}&=\int_0^T\|u(t)\|^2_{L^2(\mathbb{R}^n)}+\|\nabla u(t)\|^2_{L^2(\mathbb{R}^n)}\,dt\\&=\int_0^T\|u(t)\|^2_{L^2(\mathbb{R}^n)}\,dt+\frac{1}{2}(\|u(0)\|^2_{L^2(\mathbb{R}^n)}-\|u(T)\|^2_{L^2(\mathbb{R}^n)})+\int_0^T\int_{\mathbb{R}^n}f(t,x)u(t,x)\,dx\,dt\\&\le\|u\|^2_{L^2([0,T]\times\mathbb{R}^n)}+\frac{1}{2}\|g\|^2_{L^2(\mathbb{R}^n)}+\|f\|_{L^2([0,T]\times\mathbb{R}^n)}\|u\|_{L^2([0,T]\times\mathbb{R}^n)}\\&\le\frac{3}{2}\|u\|^2_{L^2([0,T]\times\mathbb{R}^n)}+\frac{1}{2}\|g\|^2_{L^2(\mathbb{R}^n)}+\frac{1}{2}\|f\|^2_{L^2([0,T]\times\mathbb{R}^n)}.\end{aligned}$$

We now note that

$$\begin{aligned}\|u\|^2_{L^2([0,T]\times\mathbb{R}^n)}&=\int_0^T\|u(t)\|^2_{L^2(\mathbb{R}^n)}\,dt\\&\le T\max_{0\le t\le T}\|u(t)\|^2_{L^2(\mathbb{R}^n)}\\&\le Te^T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)}+\|g\|_{L^2(\mathbb{R}^n)}\right).\end{aligned}$$

It follows that

$$\|u\|^2_{L^2([0,T];H^1(\mathbb{R}^n))}\le\left(\frac{1}{2}+Te^T\right)\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)}+\|g\|_{L^2(\mathbb{R}^n)}\right),$$

i.e. $\|u\|_{L^2([0,T];H^1(\mathbb{R}^n))}\le C'_T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)}+\|g\|_{L^2(\mathbb{R}^n)}\right)$ holds for $C'_T=\left(\frac{Te^T+1/2}{2}\right)^{1/2}$.

(3) This follows upon noting, for each t, that

$$\begin{aligned}\|\partial_t u(t)\|_{H^{-1}(\mathbb{R}^n)}&=\|\Delta u(t)+f(t)\|_{H^{-1}(\mathbb{R}^n)}\\&\le\|\Delta u(t)\|_{H^{-1}(\mathbb{R}^n)}+\|f(t)\|_{H^{-1}(\mathbb{R}^n)}\\&\le\|u(t)\|_{H^1(\mathbb{R}^n)}+\|f(t)\|_{L^2(\mathbb{R}^n)}.\end{aligned}$$

Squaring and integrating in t between 0 and T yields the desired result.

□

In short, we obtain estimates on u and some derivatives on u, by multiplying the PDE that u satisfies by itself and integrating in space.

We can obtain another kind of estimate by doing similar manipulations, but starting with the PDE and squaring it instead: again assuming that u decays sufficiently quickly, we have

$$\begin{aligned}\int_{\mathbb{R}^n}|f(t,x)|^2\,dx&=\int_{\mathbb{R}^n}\left(\partial_t u(t,x)-\Delta u(t,x)\right)^2\,dx\\&=\|\partial_t u(t)\|^2_{L^2(\mathbb{R}^n)}-2\int_{\mathbb{R}^n}\partial_t u(t,x)\Delta u(t,x)\,dx+\|\Delta u(t)\|^2_{L^2(\mathbb{R}^n)}.\end{aligned}$$

We now note that

$$\int_{\mathbb{R}^n} \partial_t u(t,x)\Delta u(t,x)\,dx = -\int_{\mathbb{R}^n} \nabla(\partial_t u)(t,x)\cdot\nabla u(t,x)\,dx$$
$$= -\frac{1}{2}\partial_t\left(\|\nabla u(t)\|^2_{L^2(\mathbb{R}^n)}\right).$$

We also have

Lemma 7.7. *Suppose $u \in H^2(\mathbb{R}^n)$. Then*

$$\|\Delta u\|^2_{L^2(\mathbb{R}^n)} = \sum_{i,j=1}^n \|\partial_i\partial_j u\|^2_{L^2(\mathbb{R}^n)}.$$

(Note that if we define $|D^2u|^2 := \sum_{i,j=1}^n |\partial_i\partial_j u|^2$, then the latter quantity can be written as $\|D^2u\|^2_{L^2(\mathbb{R}^n)}$.)

It follows that

$$\|f(t)\|^2_{L^2(\mathbb{R}^n)} = \|\partial_t u(t)\|^2_{L^2(\mathbb{R}^n)} + \partial_t\left(\|\nabla u(t)\|^2_{L^2(\mathbb{R}^n)}\right) + \|D^2u\|^2_{L^2(\mathbb{R}^n)}.$$

Following similar logic to the above regularity result, we have:

Theorem 7.8. *Suppose u is a smooth solution to*

$$\partial_t u - \Delta u = f \text{ in } (0,T]\times\mathbb{R}^n, u|_{t=0} = g,$$

with $f(t,\cdot) \in L^2(\mathbb{R}^n)$ for all $0 < t \le T$, $g \in H^1(\mathbb{R}^n)$, and $u(t,\cdot)$ decays sufficiently quickly for all t. Then, there exist constants C_T, C'_T, C''_T such that:

(1) $\max_{0\le t\le T}\|\nabla u(t)\|_{L^2(\mathbb{R}^n)} \le C_T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{H^1(\mathbb{R}^n)}\right)$.
(2) $\|\partial_t u\|_{L^2([0,T]\times\mathbb{R}^n)} \le C'_T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{H^1(\mathbb{R}^n)}\right)$.
(3) $\|D^2u\|_{L^2([0,T]\times\mathbb{R}^n)} \le C''_T\left(\|f\|_{L^2([0,T]\times\mathbb{R}^n)} + \|g\|_{H^1(\mathbb{R}^n)}\right)$.

8. Lecture 08 (04/21): Parabolic regularity on bounded domains

8.1. **Heat equation on bounded domains.** We now turn our attention to the heat equation, and more general parabolic equations, on bounded domains $U \subset \mathbb{R}^n$. We assume the following:

Assumption 8.1. There exists a collection of functions $\{w_k\}_{k=1}^\infty$ such that:

- Each $w_k \in C^\infty(\overline{U})$.
- The collection $\{w_k\}_{k=1}^\infty$ forms an orthonormal basis[18] for $L^2(U)$.
- Each w_k is an eigenfunction of $-\Delta$, say of eigenvalue λ_k. Furthermore, we assume there exist positive constants c, α, and N such that

$$\lambda_k \geq ck^\alpha \text{ for all } k \geq N.$$

As we'll see shortly, these assumptions are always satisfied, but the choice of functions $\{w_k\}$ is far from unique.

Given these assumptions, we note the following: *a* solution to the problem

$$\partial_t u - \Delta u = 0 \text{ in } (0,\infty) \times U, \quad u(0,x) = g(x) \text{ on } U$$

is given by

$$u(t,x) = \sum_{k=1}^\infty e^{-\lambda_k t}(g, w_k)_{L^2(u)} w_k(x).$$

Note that the assumptions of $w_k \in C^\infty(\overline{U})$ and the growth rate of λ_k guarantee that u is in fact *smooth* on $(0,\infty)\times\overline{U}$ (i.e. smooth in $(0,\infty)\times U$ with derivatives continuous up to $(0,\infty)\times\partial U$).

Example 8.2. Let $U = (0,1) \subset \mathbb{R}$. Examples of collections of $\{w_k\}$ which satisfy our assumptions include:

(1) $w_k(x) = \sqrt{2}\sin(k\pi x)$, in which case $\lambda_k = \pi^2 k^2$ (corresponding to "Dirichlet boundary conditions").

(2) $w_k(x) = \begin{cases} 1 & k = 0 \\ \sqrt{2}\cos(k\pi x) & k \geq 1 \end{cases}$, in which case $\lambda_k = \pi^2 k^2$ (corresponding to the "Neumann boundary conditions")[19].

(3) $w_k(x) = e^{2\pi i k x}$ for $k \in \mathbb{Z}$, in which case $\lambda_k = 4\pi^2 k^2$ (corresponding to "periodic boundary conditions")[20].

---

[18] Recall this means that the collection $\{w_k\}$ is orthonormal, and for every $f \in L^2(U)$ can be written as a convergent series $f = \sum_{k=1}^\infty f_k w_k$ for some sequence of numbers $\{f_k\}$, where the series converges in $L^2(U)$. The last condition can be rephrased as saying that the *algebraic span* of these functions (i.e. collections of finite linear combinations of w_k) is dense in $L^2(U)$.

[19] For purposes of ordering, we can consider $k \mapsto w_{k+1}$.

[20] For purposes of ordering, we can consider $k \mapsto \begin{cases} w_{k/2} & k \text{ even} \\ w_{-(k+1)/2} & k \text{ odd} \end{cases}$, or instead consider periodic sines and cosines.

We'll focus on the Dirichlet condition. This roughly means we want to only consider functions where "$u(t,x) = 0$ when $x \in \partial U$." To do so, we need to consider some Sobolev spaces associated to U:

Definition 8.3. The H^1 norm on U (definable e.g. on $C^\infty(\overline{U})$) is given by

$$\|u\|_{H^1(U)} := \left(\|u\|_{L^2(u)}^2 + \|\nabla u\|_{L^2(U)}^2\right)^{1/2}.$$

The space $H_0^1(U)$ is defined as the closure of $C_c^\infty(U)$ with respect to the $H^1(U)$ norm; this is a Banach space with respect to the H^1 norm (sometimes called the H_0^1 norm in this context).

Remark 13. We can also define $H^1(U)$, either via an extrinsic characterization ($u \in H^1(U) \iff u \in L^2(U), \partial_j u \in L^2(U)$ for all j), or define it as a completion analogously to the above definition, in which case $H^1(U)$ is the completion of $C^\infty(\overline{U})$ with respect to the H^1 norm. For bounded open sets U, an important fact is that $H_0^1(U) \subsetneq H^1(U)$; the subscript 0 heuristically corresponds to the property of "$u = 0$ on ∂U" (i.e. the Dirichlet boundary condition).

Note that $H_0^1(U)$ has the structure of a Hilbert space under the inner product

$$\langle u, v\rangle_{H_0^1(U)} = \langle u, v\rangle_{L^2(U)} + \langle \nabla u, \nabla v\rangle_{L^2(U)}.$$

Moreover, for $u, v \in C_c^\infty(U)$, integration by parts gives

$$\langle \nabla u, \nabla v\rangle_{L^2(U)} = \langle u, (-\Delta)v\rangle_{L^2(U)},$$

so for $u \in C_c^\infty(U)$ we have

$$\|u\|_{H_0^1(U)}^2 = \|u\|_{L^2(U)}^2 + \langle u, -\Delta u\rangle_{L^2(U)}.$$

By density, the above result still holds for $u \in C^2(\overline{U}) \cap H_0^1(U)$. In particular, functions in $C^2(\overline{U}) \cap H_0^1(U)$ have the property that their H^1 norm is controlled by the L^2 norms of u and Δu (this is *not* true for functions in $H^1(U)$ in general.)

An important fact is:

Theorem 8.4. *Let U be bounded open (with smooth boundary?). Then $-\Delta$ admits a sequence of eigenfunctions $\{w_k\}$ which all lie in $H_0^1(U)$, with the corresponding eigenvalues $\lambda_k \to +\infty$. Moreover, each $w_k \in C^\infty(\overline{U})$.*

Corollary 8.5. *For any $g \in L^2(U)$, there exists a solution $u(t,x)$ to the heat equation $\partial_t u - \Delta u = 0$ in U which is* ***smooth*** *on $(0,\infty)_t \times \overline{U}$ such that $u(t,x) = 0$ on $(0,\infty)_t \times \partial U$ and $u(t,\cdot) \to g$ in $L^2(U)$ as $t \to 0^+$.*

Definition 8.6. The space $H^{-1}(U)$ is defined as the dual space of $H_0^1(U)$. This is a Hilbert space, with respect to the norm

$$\|u\|_{H^{-1}(U)} := \sup_{v \in H_0^1(U), \|v\|_{H_0^1(U)} = 1} |(u, v)|.$$

Since $C_c^\infty(U)$ is contained in, and in fact is a dense subspace of, $H_0^1(U)$, it follows that there is a continuous inclusion $H^{-1}(U) \hookrightarrow D'(U)$ by identifying $u \in H^{-1}(U)$ with its restriction on $C_c^\infty(U)$. (Note that there abstractly exists an isometric isomorphism between $H^{-1}(U)$ and $H_0^1(U)$, but we will choose not to use that isomorphism to identify $H^{-1}(U)$ and $H_0^1(U)$; instead we identify $H^{-1}(U)$ with a subspace of $D'(U)$.)

8.2. **General parabolic operators: the Galerkin method.** We consider a general initial-value/Dirichlet boundary parabolic problem

$$\begin{aligned} \partial_t u + Lu = f \text{ in } (0,T] \times U, \\ u = 0 \text{ on } [0,T] \times \partial U, \\ u = g \text{ on } \{0\} \times U. \end{aligned} \tag{4}$$

Here, L is a elliptic differential operator, written in divergence form

$$Lu = -\sum_{i,j=1}^n \partial_{x_j} \left(a^{ij}(t,x)\partial_{x_i} u \right) + \sum_{i=1}^n b^i(t,x)\partial_{x_i} u + c(t,x)u$$

with[21] $a^{ij}, b^i, c \in L^\infty([0,T] \times \overline{U})$ and

$$\sum_{i,j=1}^n a^{ij}(t,x)\xi_i\xi_j \geq \theta|\xi|^2$$

for all $(t,x) \in (0,T) \times U$ for some $\theta > 0$ (i.e. the second order spatial derivative component of L is uniformly elliptic).

Let $L(t)$ denote the operator L, viewed as an operator on U, where the coefficients are frozen at some t. For $u, v \in H_0^1(U)$, let

$$B[u,v;t] := (L(t)u, v)_{L^2(U)}.$$

Then, an integration by parts arguments gives

$$B[u,v;t] = \int_U \sum_{i,j=1}^n a^{ij}(t,x)\partial_{x_i} u(x)\partial_{x_j} v(x) + \sum_{i=1}^n b^i(t,x)\partial_{x_i} u(x)v(x) + c(t,x)u(x)v(x)\,dx$$

(the condition of $u, v \in H_0^1(U)$ means we do not need to consider a boundary term in integration by parts). Note that if u solves our problem (4), say in the sense of distributions, and is sufficiently regular, say with $u(t) \in H^1(U)$ for each t (in which case it would be in $H_0^1(U)$ due to the Dirichlet boundary condition), and $v \in H_0^1(U)$, the integration by parts would give

$$(\partial_t u(t), v)_{L^2(U)} + B[u(t), v; t] = (f(t), v)_{L^2(U)}$$

for all t.

One can check:

[21] This convention has changed from the original convention introduced in class. See Remark 15 for more details.

Lemma 8.7 (cf. [Eva10] Section 6.2.2 Theorem 2)**.** *We have the upper bound*

$$|B[u,v;t]| \le \alpha \|u\|_{H_0^1(U)} \|v\|_{H_0^1(U)}$$

and the lower bound

$$B[u,u;t] \ge \beta \|u\|^2_{H_0^1(U)} - \gamma \|u\|^2_{L^2(U)}$$

for some $\alpha, \beta > 0$ and $\gamma \ge 0$, uniform in t.

(The first estimate is standard; the second estimate uses the uniform ellipticity of the coefficients a^{ij}.)

It thus follows that we would have

$$\begin{aligned} |(\partial_t u(t), v)_{L^2(U)}| &\le \alpha \|u(t)\|_{H_0^1(U)} \|v\|_{H_0^1(U)} + \|f(t)\|_{L^2(U)} \|v\|_{L^2(U)} \\ &\le C_{u,f} \|v\|_{H_0^1(U)}, \end{aligned}$$

so that in particular $\partial_t u(t)$ defines a continuous linear map on $H_0^1(U)$ via the L^2-pairing, i.e. $\partial_t u(t) \in H^{-1}(U)$. This motivates the notion of *weak solution* that we'll work with in this problem:

Definition 8.8. Let $u \in L^2([0,T]; H_0^1(U))$ with $\partial_t u \in L^2([0,T]; H^{-1}(U))$. We say that u is a *weak solution* to the problem (4), i.e. the problem

$$\begin{aligned} \partial_t u + Lu &= f \text{ in } (0,T] \times U, \\ u &= 0 \text{ on } [0,T] \times \partial U, \\ u &= g \text{ on } \{0\} \times U, \end{aligned}$$

if $(\partial_t u(t), v)_{L^2(U)} + B[u(t), v; t] = (f(t), v)_{L^2(U)}$ for almost every[22] $0 \le t \le T$ and every $v \in H_0^1(U)$, and $u(0) = g$ as well.

Remark 14. The assumptions on u in the definition actually give that $u \in C^0([0,T]; L^2(U))$ (possibly after redefining u on a measure zero set of times), so the condition $u(0) = g$ does make sense.

The goal now is to prove well-posedness results for weak solutions of the problem (4), i.e.:

- Existence of a weak solution for any $f \in L^2([0,T] \times U)$ and $g \in L^2(U)$.
- Uniqueness of weak solutions.
- Energy/regularity estimates on weak solutions.

These will be done via *Galerkin approximations*, which reduce the problem to studying problems on a finite-dimensional space of functions, and taking the limit.

Recall Assumption 8.1; in particular the existence of a smooth orthonormal basis $\{w_k\}$ for $L^2(U)$. Let $V_m = \text{span}\,\{w_1, \dots, w_m\}$.

Theorem 8.9 (cf. [Eva10] Section 7.1.2, Theorem 1)**.** *Let $f \in L^2([0,T] \times U)$, $g \in L^2(U)$, and $m \in \mathbb{N}$. Then there exists a unique $u_m \in C^1([0,T]; V_m)$ which satisfies*

$$(\partial_t u_m(t), v)_{L^2(U)} + B[u_m(t), v; t] = (f(t), v)_{L^2(U)}, \quad (u_m(0), v)_{L^2(U)} = (g, v)_{L^2(U)}$$

[22] This convention has changed from the original convention introduced in class. See Remark 15 for more details.

for all $v \in V_m$.

Proof. Writing $u_m(t) = \sum_{k=1}^m d^k(t) w_k$, we then get

$$\sum_{k=1}^m (d^k)'(t)(w_k, v) + d_k(t) B[w_k, v; t] = (f(t), v)_{L^2(U)}.$$

Writing $f^l(t) = (f(t), w_l)$, it follows by plugging in $v = w_l$ for $1 \leq l \leq m$ that

$$(d^l)'(t) + \sum_{k=1}^m B(w_k, w_l; t) d^k(t) = f^l(t).$$

This gives a system of m first-order ODEs, so by ODE theory there exists a unique C^1 solution $(d^1, \ldots, d^m)$, upon providing the initial data $d^l(0) = (g, w_l)$. $\square$

Next lecture: Regularity estimates on the Galerkin approximations, applied to obtain well-posedness results for weak solutions of problem (4).

9. Lecture 09 (04/26): Well-posedness of weak solutions via the Galerkin method

9.1. **Clarification on conventions.** First, a clarification regarding definitions:

Remark 15. In class, during Lecture 08, I originally defined a weak solution as satisfying the pairing equation

$$(\partial_t u(t), v)_{L^2(U)} + B[u(t), v; t] = (f(t), v)_{L^2(U)}$$

for *every* $0 \leq t \leq T$. This differed from the convention in [Eva10], Section 7.2, where he required the condition hold only for *almost every* $0 \leq t \leq T$. I had originally justified this by assuming that the coefficients a^{ij}, b^i, c involved the in the operator L were continuous up to the boundary.

It turns out that the "almost every" t condition shows up in many different places, and that it is much easier to incorporate that as part of the definition rather than trying to avoid it. As such, I will revert to the conventions of [Eva10]. Thus, the pairing condition should hold **for almost every** $0 \leq t \leq T$, and the coefficients a^{ij}, b^i, c only need to be uniformly bounded on $[0,T] \times \overline{U}$. (The change has been reflected in the lecture notes in the previous section.)

We continue to work with Assumption 8.1. Thus, let $\{w_k\}_{k=1}^{\infty}$ be an orthonormal basis for $L^2(U)$, such that each $w_k \in C^\infty(\overline{U})$, and each w_k is an eigenfunction of $-\Delta$. For the Dirichlet problem, we'll make a further assumption:

$$\text{each } w_k \in H_0^1(U) \text{ as well.}$$

This is possible due to Theorem 8.4. Note in that case that $\{w_k\}$ also forms an orthogonal basis of $H_0^1(U)$, since

$$(u, v)_{H_0^1(U)} = (u, v)_{L^2(U)} + (u, (-\Delta)v)_{L^2(U)}$$

for $u, v \in H_0^1(U)$.

9.2. **Galerkin approximation regularity and existence of weak solution.**

From last time: given $f \in L^2([0,T] \times U)$, $g \in L^2(U)$, and $m \in \mathbb{N}$. Then there exists a unique "Galerkin approximation" $u_m \in C^1([0,T]; V_m)$ which satisfies

$$(\partial_t u_m(t), v)_{L^2(U)} + B[u_m(t), v; t] = (f(t), v)_{L^2(U)}, \quad (u_m(0), v)_{L^2(U)} = (g, v)_{L^2(U)}$$

for all $v \in V_m$, where $V_m = \operatorname{span}\{w_1, \ldots, w_m\}$.

We also have a regularity estimate for these Galerkin approximations:

Theorem 9.1 (cf. [Eva10] Section 7.1.2, Theorem 2)**.** *Let $u_m(t) \in V_m$ be the unique solution obtained above. Then we have estimates*

$$\begin{aligned} &\max_{t\in[0,T]} \|u_m(t)\|_{L^2(U)} + \|u_m\|_{L^2([0,T];H_0^1(U))} + \|\partial_t u_m\|_{L^2([0,T];H^{-1}(U))} \\ &\leq C(\|f\|_{L^2([0,T]\times U)} + \|g\|_{L^2(U)}), \end{aligned}$$

where the constant C is independent *of m.*

Proof. The proof is similar to the proof of Theorem 7.5. Applying the weak formulation to $u_m(t)$, we see that

$$\frac{1}{2}\partial_t\left(\|u_m(t)\|_{L^2(U)}^2\right)+B[u_m(t),u_m(t);t]=(f(t),u_m(t))_{L^2(U)}. \tag{5}$$

Recall from Lemma 8.7 that we have a lower bound

$$B[u,u;t]\geq\beta\|u\|_{H_0^1(U)}^2-\gamma\|u\|_{L^2(U)}^2$$

for some $\beta>0$ and $\gamma\geq 0$. It follows that

$$\begin{aligned}\frac{1}{2}\partial_t\left(\|u_m(t)\|_{L^2(U)}^2\right)&\leq(f(t),u_m(t))_{L^2(U)}-B[u_m(t),u_m(t);t]\\&\leq(f(t),u_m(t))_{L^2(U)}-\beta\|u_m(t)\|_{H_0^1(U)}^2+\gamma\|u_m(t)\|_{L^2(U)}^2\\&\leq\frac{1}{2}\|f(t)\|_{L^2(U)}^2+\left(\frac{1}{2}+\gamma\right)\|u_m(t)\|_{L^2(U)}^2.\end{aligned}$$

By Gronwall's Lemma, we thus have

$$\|u_m(t)\|_{L^2(U)}^2\leq e^{(1+\gamma)t}\left(\|u_m(0)\|_{L^2(U)}^2+\int_0^t e^{-(1+\gamma)s}\|f(s)\|_{L^2(U)}^2\,ds\right).$$

It follows that

$$\max_{t\in[0,T]}\|u_m(t)\|_{L^2(U)}^2\leq e^{(1+\gamma)T}\left(\|f\|_{L^2([0,T]\times U)}^2+\|g\|_{L^2(U)}^2\right),$$

which gives the first part of the estimate. Moreover, from (5), we also have

$$\begin{aligned}&\frac{1}{2}\partial_t\left(\|u_m(t)\|_{L^2(U)}^2\right)+\beta\|u_m(t)\|_{H_0^1(U)}^2-\gamma\|u_m(t)\|_{L^2(U)}^2\\&\leq\frac{1}{2}\partial_t\left(\|u_m(t)\|_{L^2(U)}^2\right)+B[u_m(t),u_m(t);t]=(f(t),u_m(t))_{L^2(U)}\end{aligned}$$

and hence

$$\begin{aligned}\frac{1}{2}\partial_t\left(\|u_m(t)\|_{L^2(U)}^2\right)+\beta\|u_m(t)\|_{H_0^1(U)}^2&\leq\gamma\|u_m(t)\|_{L^2(U)}^2+(f(t),u_m(t))_{L^2(U)}\\&\leq\frac{1}{2}\|f(t)\|_{L^2(U)}+\left(\gamma+\frac{1}{2}\right)\|u_m(t)\|_{L^2(U)}^2.\end{aligned}$$

Integrating from $t=0$ to $t=T$ thus yields

$$\begin{aligned}&\frac{1}{2}\left(\|u_m(T)\|_{L^2(U)}^2-\|u_m(0)\|_{L^2(U)}^2\right)+\beta\|u_m\|_{L^2([0,T];H_0^1(U))}^2\\&\leq\frac{1}{2}\|f\|_{L^2([0,T]\times U)}^2+\int_0^T\left(\gamma+\frac{1}{2}\right)\|u_m(t)\|_{L^2(U)}^2\,dt,\end{aligned}$$

which yields an estimate on $\|u_m\|_{L^2([0,T];H_0^1(U))}$ since $\beta>0$ (cf. the proof in Theorem 7.5). Finally, from the estimate (cf. Lemma 8.7)

$$|B[u,v;t]|\leq\alpha\|u\|_{H_0^1(U)}\|v\|_{H_0^1(U)}$$

we have

$$|(\partial_t u_m(t), v)_{L^2(U)}| = |-B[u_m(t), v; t] + (f(t), v)_{L^2(U)}| \le \Big(\alpha \|u_m(t)\|_{H_0^1(U)} + \|f(t)\|_{L^2(U)}\Big) \|v\|_{H_0^1(U)}$$

for any $v \in V_m$. In general, for any $v \in H_0^1(U)$ we can write $v = v' + v''$ with $v' \in V_m$ and $v'' \in V_m^\perp$, in which case $\|v'\|_{H_0^1(U)} \le \|v\|_{H_0^1(U)}$, and

$$\begin{aligned} |(\partial_t u_m(t), v)_{L^2(U)}| &= |(\partial_t u_m(t), v')_{L^2(U)}| \\ &\le \Big(\alpha \|u_m(t)\|_{H_0^1(U)} + \|f(t)\|_{L^2(U)}\Big) \|v'\|_{H_0^1(U)} \\ &\le \Big(\alpha \|u_m(t)\|_{H_0^1(U)} + \|f(t)\|_{L^2(U)}\Big) \|v\|_{H_0^1(U)}. \end{aligned}$$

Thus we obtain

$$\|\partial_t u_m(t)\|_{H^{-1}(U)} \le \alpha \|u_m(t)\|_{H_0^1(U)} + \|f(t)\|_{L^2(U)}.$$

The estimate on $\|\partial_t u_m\|_{L^2([0,T];H^{-1}(U))}$ then follows by squaring and integrating. $\square$

Thus, in considering our Galerkin approximations u_m, we see that they are uniformly bounded, with respect to all three norms appearing in the LHS of Theorem 9.1. This helps prove existence via taking a subsequence. We recall:

Theorem 9.2 (Banach-Alaoglu). *Let X be a normed linear space, and X^* its continuous dual space. Then, the unit ball in X^**

$$\{\ell \in X^* \,:\, \|\ell\|_{X^*} \le 1\}$$

is compact in the weak- topology. In particular, if X is reflexive (e.g. X is a Hilbert space), so that the weak topology on X coincides with the weak-* topology on X viewing X as the dual of X^*, then the unit ball in X is compact with respect to the weak topology on X.*

We can use this to prove:

Theorem 9.3 (Existence of weak solution). *Let $f \in L^2([0,T] \times U)$ and $g \in L^2(U)$. Then there exists a weak solution to the problem* (4)*.*

Proof. From the regularity bounds on the Galerkin approximations u_m and the Banach-Alaoglu theorem, we see that $\{u_m\}$ is sequentially compact with respect to the weak topology on $L^2([0,T]; H_0^1(U))$, and $\{\partial_t u_m\}$ is sequentially compact with respect to the weak topology on $L^2([0,T]; H^{-1}(U))$. It follows that there exists a subsequence $\{u_{m_l}\}_{l=1}^\infty$ such that $\{u_{m_l}\}$ converges in $L^2([0,T]; H_0^1(U))$, say to u, with $\{\partial_t u_{m_l}\}$ also converges in $L^2([0,T]; H^{-1}(U))$, say to $\tilde{u}$. Then necessarily we must have $\tilde{u} = \partial_t u$ in the sense of distributions.

It now suffices to show that u is indeed a weak solution. We first note that if

$$v(t) = \sum_{k=1}^{N} d^k(t) w_k$$

for $d^1, \dots, d^N \in C^1([0,T])$, then for $m \ge m_0$ we have

$$(\partial_t u_m(t), v(t))_{L^2(U)} + B[u_m(t), v(t); t] = (f(t), v(t))_{L^2(U)}.$$

Integrating thus gives

$$\int_0^T (\partial_t u_m(t), v(t))_{L^2(U)} + B[u_m(t), v(t); t]\, dt = \int_0^T (f(t), v(t))_{L^2(U)}\, dt.$$

Letting $m = m_l$ and passing to the limit, we see, due to the weak convergences, that

$$\int_0^T (\partial_t u(t), v(t))_{L^2(U)} + B[u(t), v(t); t]\, dt = \int_0^T (f(t), v(t))_{L^2(U)}\, dt.$$

This holds for all functions of the form $v(t) = \sum_{k=1}^N d^k(t) w_k$, and hence by density for all $v \in L^2([0,T]; H_0^1(U))$. In particular, if we now fix $v_0 \in H_0^1(U)$ and consider $v(t) = \phi(t) v_0$ for $\phi \in L^2([0,T])$, we then see that

$$\int_0^T \big((\partial_t u(t), v_0)_{L^2(U)} + B[u(t), v_0; t]\big)\, \phi(t)\, dt = \int_0^T \big((f(t), v_0)_{L^2(U)}\big)\, \phi(t)\, dt,$$

for all $\phi \in L^2([0,T])$, from which we conclude that

$$(\partial_t u(t), v_0)_{L^2(U)} + B[u(t), v_0; t] = (f(t), v_0)_{L^2(U)}$$

for almost every[23] $0 \leq t \leq T$. This shows that u satisfies the distributional formulation of a weak solution.

It remains to show that $u(0) = g$. To do so, note that from integration by parts in t we have

$$\int_0^T -(u(t), v'(t))_{L^2(U)} + B[u(t), v(t); t]\, dt = \int_0^T (f(t), v(t))_{L^2(U)}\, dt + (u(0), v(0))_{L^2(U)}$$

if $v \in C^1([0,T]; H_0^1(U))$ with $v(T) = 0$. Similarly, for the Galerkin approximations we have correspondingly

$$\int_0^T -(u_m(t), v'(t))_{L^2(U)} + B[u_m(t), v(t); t]\, dt = \int_0^T (f(t), v(t))_{L^2(U)}\, dt + (u_m(0), v(0))_{L^2(U)},$$

again for $v \in C^1([0,T]; H_0^1(U))$ with $v(T) = 0$. Taking $m = m_l$ and passing the limit, noting that $u_m(0) \to g$ in $L^2(U)$ since $u_m(0)$ is the projection of g onto V_m, we thus obtain by weak convergence

$$\int_0^T -(u(t), v'(t))_{L^2(U)} + B[u(t), v(t); t]\, dt = \int_0^T (f(t), v(t))_{L^2(U)}\, dt + (g, v(0))_{L^2(U)}.$$

Thus $(u(0), v(0))_{L^2(U)} = (g, v(0))_{L^2(U)}$ for any $v \in C^1([0,T]; H_0^1(U))$ with $v(T) = 0$, and since $v(0)$ can take any value in $H_0^1(U)$ given those restrictions, it follows that $u(0) = g$, as desired. □

[23]This upgrades to an equality for *all* times if we assume in addition that $f \in C([0,T]; L^2(U))$.

9.3. **Uniqueness and improved regularity of weak solutions.** Thus, we see that the Galerkin method produces the existence of a weak solution. Moreover, any weak solution can be obtained as a limit of Galerkin approximations, if we are able to show uniqueness:

Theorem 9.4 (Uniqueness)**.** *A weak solution of problem* (4) *is unique.*

Proof. It suffices to show that if $f = 0$ and $g = 0$, then the only weak solution to (4) is identically zero. Plugging in $v = u(t)$ in the weak formulation gives

$$\frac{1}{2}\partial_t(\|u(t)\|^2_{L^2(U)}) + B[u(t), u(t); t] = (\partial_t u(t), u(t))_{L^2(U)} + B[u(t), u(t); t] = (f(t), u(t))_{L^2(U)} = 0.$$

From Lemma 8.7, we have

$$\frac{1}{2}\partial_t(\|u(t)\|^2_{L^2(U)}) = -B[u(t), u(t); t] \le -\beta\|u(t)\|^2_{H^1_0(U)} + \gamma\|u(t)\|^2_{L^2(U)} \le \gamma\|u(t)\|^2_{L^2(U)},$$

so by Gronwall's inequality we have

$$\|u(t)\|^2_{L^2(U)} \le e^{2\gamma t}\|u(0)\|^2_{L^2(U)}.$$

If $u(0) = g = 0$, then this gives $u(t) = 0$ for all t. □

If the initial data g has a bit more regularity, then we can obtain more regularity on u. For convenience, we now assume:

$$a^{ij}, b^i, c \in C^\infty(\overline{U}) \quad \text{(i.e. independent of } t).$$

Theorem 9.5. *Suppose u is a weak solution to Problem* (4)*, and $g \in H^1_0(U)$. Then $u \in L^2([0,T]; H^2(U)) \cap L^\infty([0,T]; H^1_0(U))$, $\partial_t u \in L^2([0,T]; L^2(U))$, and*

$$\operatorname*{ess\,sup}_{0\le t\le T} \|u(t)\|_{H^1_0(U)} + \|u\|_{L^2([0,T];H^2(U))} + \|\partial_t u\|_{L^2([0,t];L^2(U))} \le C(\|f\|_{L^2([0,T]\times U)} + \|g\|_{H^1_0(U)}).$$

Proof sketch. We prove the corresponding estimates on the Galerkin approximations and pass to the limit. To do so, we plug in $v = \partial_t u_m$ in the weak formulation of the Galerkin solutions, and rewrite $B[u_m, \partial_t u_m; t]$ as the t derivative of a quadratic form of ∇u_m plus some remainders.

To be continued next lecture... □

10. Lecture 10 (04/28): Improved Regularity and Heat Kernels on Compact Manifolds

10.1. **Improved Regularity of Weak Solutions.** We start by finishing the proof from the end of last class:

Proof of Theorem 9.5. The Galerkin approximations satisfy

$$(\partial_t u_m(t), \partial_t u_m(t))_{L^2(U)} + B[u_m(t), \partial_t u_m(t)] = (f(t), \partial_t u_m(t))_{L^2(U)}.$$

(Recall that we assume the coefficients are now independent of t, so the bilinear form B is independent of t as well.)

Note that if $L = -\Delta$, then $B[v(t), \partial_t v(t)] = \int_U \nabla v(t) \cdot \partial_t \nabla v(t)\, dt = \frac{1}{2}\partial_t(\|\nabla v(t)\|^2_{L^2(U)})$, so morally we have an estimate on $\|\partial_t u_m(t)\|^2_{L^2(U)} + \frac{1}{2}\partial_t(\|\nabla u_m(t)\|^2_{L^2(U)})$. In fact, if we let

$$A[u,v] = \int_U \sum_{i,j=1}^n a^{ij}(x)\partial_{x_i} u(x) \partial_{x_i} v(x)\, dx$$

then

$$\frac{1}{2}\frac{d}{dt}(A[v(t), v(t)]) = B[v(t), \partial_t v(t)] - \int_U \sum_{i=1}^n b^i(x)\partial_{x_i} v(t,x)\partial_t v(t,x) + c(x) v(t,x) \partial_t v(t,x)\, dx.$$

Thus

(6)

$$\begin{aligned}
&\|\partial_t u_m(t)\|^2_{L^2(U)} + \frac{d}{dt}\left(\frac{1}{2}A[u_m(t), u_m(t)]\right) \\
&= \left(f(t) + \sum_{i=1}^n b^i \partial_{x_i} u_m(t) + c u_m, \partial_t u_m\right)_{L^2(U)} \\
&\le \left\| f(t) + \sum_{i=1}^n b^i \partial_{x_i} u_m(t) + c u_m \right\|_{L^2(U)} \|\partial_t u_m(t)\|_{L^2(U)} \\
&\le \frac{1}{2}\left(\|f(t)\|_{L^2(U)} + \left\| \sum_{i=1}^n b^i \partial_{x_i} u_m(t)\right\|_{L^2(U)} + \|c u_m(t)\|_{L^2(U)} \right)^2 + \frac{1}{2}\|\partial_t u_m(t)\|^2_{L^2(U)}.
\end{aligned}$$

We now note that

$$\left\| \sum_{i=1}^n b^i \partial_{x_i} u_m(t) \right\|_{L^2(U)} \le \left(\sum_{i=1}^n \|b^i\|^2_{L^\infty}\right)^{1/2} \|u_m(t)\|_{H^1_0(U)}$$

and

$$\|c u_m(t)\|_{L^2(U)} \le \|c\|_{L^\infty} \|u_m(t)\|_{L^2(U)}.$$

Thus, subtracting $\frac{1}{2}\|\partial_t u_m(t)\|_{L^2(U)}$ from both sides of (6), multiplying by 2, and plugging in the above estimates yields

$$\|\partial_t u_m(t)\|^2_{L^2(U)} + \frac{d}{dt}\left(A[u_m(t), u_m(t)]\right) \le C\left(\|f(t)\|^2_{L^2(U)} + \|u_m(t)\|^2_{H^1_0(U)}\right),$$

the constant depending on $\|b^i\|_{L^\infty}$ and $\|c\|_{L^\infty}$. Applying the same arguments in the other regularity theorems (i.e. applying Gronwall's inequality and then integrating) yields

$$\begin{aligned}\|\partial_t u_m\|_{L^2([0,T]\times U)} + \sup_{0\le t\le T} A[u_m(t),u_m(t)] &\le C\left(A[u_m(0),u_m(0)] + \|f\|^2_{L^2([0,T]\times U)} + \|u_m\|^2_{L^2([0,T];H^1_0(U))}\right)\\ &\le C'\left(A[u_m(0),u_m(0)] + \|f\|^2_{L^2([0,T]\times U)} + \|g\|^2_{L^2(U)}\right).\end{aligned}$$

We now note that the ellipticity assumptions on a^{ij} guarantee that

$$c_1\|\nabla v\|^2_{L^2(U)} \le A[v,v] \le c_2\|\nabla v\|^2_{L^2(U)} \le c_2\|v\|^2_{H^1_0(U)}$$

so

$$\|\partial_t u_m\|_{L^2([0,T]\times U)} + \sup_{0\le t\le T}\|\nabla u_m(t)\|_{L^2(U)} \le C''\left(\|f\|^2_{L^2([0,T]\times U)} + \|g\|^2_{H^1_0(U)}\right).$$

Since $\sup_{0\le t\le T}\|u_m(t)\|_{L^2(U)}$ also satisfies the above estimate, we can replace the term $\sup_{0\le t\le T}\|\nabla u_m(t)\|_{L^2(U)}$ in the above estimate by $\sup_{0\le t\le T}\|u_m(t)\|_{H^1_0(U)}$ (up to changing the constant on the RHS), thus yielding two of the three desired estimates at least for the Galerkin approximations u_m. Applying the estimates to the subsequence u_{m_l} converging weakly to u, and letting $l\to\infty$ (taking advantage of the weak convergences of u_{m_l} and $\partial_t u_{m_l}$) yields $u\in L^\infty([0,T];H^1_0(U))$ and $\partial_t u\in L^2([0,T]\times U)$, with the desired bounds.

Finally, note that

$$Lu(t) = f(t) - \partial_t u(t)$$

in the sense of distributions, with L an elliptic operator. By elliptic regularity (cf. [Eva10] Chapter 6.3), we have

$$\|u(t)\|^2_{H^2(U)} \le C(\|f(t)\|^2_{L^2(U)} + \|\partial_t u(t)\|^2_{L^2(U)}).$$

It follows that

$$\begin{aligned}\|u\|^2_{L^2([0,T];H^2(U))} &\le C\left(\|f\|^2_{L^2([0,T]\times U)} + \|\partial_t u\|^2_{L^2([0,T]\times U)}\right)\\ &\le C'\left(\|f\|^2_{L^2([0,T]\times U)} + \|g\|^2_{H^1_0(U)}\right),\end{aligned}$$

as desired. □

Some other regularity results to mention from [Eva10] Section 7.2:

- If in addition $\partial_t f\in L^2([0,T]\times U)$ and $g\in H^2(U)$, then

$$\begin{aligned}u &\in L^\infty([0,T];H^2(U))\\ \partial_t u &\in L^\infty([0,T]\times U)\cap L^2([0,T];H^1_0(U))\\ \partial_t^2 u &\in L^2([0,T];H^{-1}(U)).\end{aligned}$$

  The proof essentially amounts to finding the PDE that $\partial_t u$ satisfies, and applying the regularity results obtained above to $\partial_t u$.

- If $g \in H^{2m+1}(U)$ and $\frac{\partial^k}{\partial t^k} f \in L^2([0,T]; H^{2m-2k}(U))$ for $0 \le k \le m$, with the right-hand sides satisfying a "compatibility condition[24]" at the boundary of U, then
$$\frac{\partial^k}{\partial t^k} u \in L^2([0,T]; H^{2m+2-2k}(U)),$$
with corresponding estimates.
The argument is an induction argument, essentially again taking a look at what equation the derivatives of u must satisfy.
- In particular, if the right-hand sides are both C^∞ (on the closures, i.e. on $[0,T] \times \overline{U}$ and on $\overline{U}$, respectively), and the compatibility conditions are satisfied, then the solution u is also smooth up to all boundaries, i.e. $u \in C^\infty([0,T] \times \overline{U})$.

10.2. **Heat Kernel on Compact Manifolds.** We conclude our treatment of parabolic equations by considering the "heat kernel" (an analogue of the fundamental solution studied in Lecture 05) for compact manifolds. It turns out that the structure of the heat kernel (reflecting the behavior of solving the Cauchy problem for the heat equation) is related to the geometry and topology of the manifold.

The source for this section is [Cha84], specifically Chapter 6.

Thus, let (M, g) be a Riemannian manifold (i.e. M is a smooth manifold, g a Riemannian metric on M). Associated to the metric is the so-called Laplace-Beltrami operator Δ_g (which can be defined as the composition of the divergence and the gradient, once those notions are defined). In local coordinates, if $g = \sum_{i,j=1}^n g_{ij}(x)\, dx_i\, dx_j$ (i.e. $(g_{ij}(x))_{i,j=1}^n$ is the matrix associated to this metric, with $g_{ij}(x) = g(\partial_{x_i}|_x, \partial_{x_j}|_x)$), then
$$\Delta_g u = \frac{1}{\sqrt{\det g}} \sum_{i,j=1}^n \partial_{x_i}(g^{ij}\sqrt{\det g}\partial_{x_j} u)$$
where $(g^{ij}(x)) = (g_{ij}(x))^{-1}$ is the dual metric obtained in coordinates by inverting the matrix (g_{ij}), and $\det g$ is the determinant of the metric viewing it as a matrix (g_{ij}).

We can then consider the Cauchy problem for the heat equation on (M, g):
$$(\partial_t - \Delta_g)u = 0 \text{ in } (0,\infty) \times M, \quad u(0,x) = f(x) \text{ on } M.$$
Recall that for Euclidean space $\mathbb{R}^n$ we were able to derive an explicit formula for a *fundamental solution* of the heat operator, which can subsequently be used to obtain a formula for the solution of the Cauchy problem via convolution. For a general manifold the notion of a convolution does not quite make sense (some additive structure or other group action is needed); furthermore in local coordinates the Laplace-Beltrami operator Δ_g need not have "constant coefficients" (and one key feature of convolutions is the compatibility with *constant-coefficient* differential operators). Nonetheless, we can still generalize this notion:

[24]This holds e.g. if f and g are compactly supported in $[0,T] \times U$ and in U, respectively, i.e. vanish in a neighborhood of ∂U

Definition 10.1. A *fundamental solution of the heat equation*, also known as a *heat kernel*, is a continuous function $p : (0, \infty) \times M \times M \to \mathbb{R}$, which is C^2 in x and C^1 in t, such that

$$L_x p = 0 \quad \text{in } (0, \infty) \times M \times M \quad \text{and} \quad \lim_{t \to 0^+} p(t, \cdot, y) = \delta_y.$$

The last equation means if $\phi \in C^\infty(M)$, then

$$\lim_{t \to 0^+} \int_M p(t, x, y)\phi(x)\, d\mathrm{Vol}_g(x) = \phi(y)$$

where $d\mathrm{Vol}_g$ is the volume form induced by the Riemannian metric g.

Note that if such p exists, then for $f : M \to \mathbb{R}$ sufficiently nice (say continuous) we have that if

$$u(t, x) := \int_M p(t, x, y) f(y)\, d\mathrm{Vol}_g(y),$$

then u solves the heat equation for $t > 0$, while for any $\phi \in C^\infty(M)$ we have[25]

$$\int_M u(t, x)\phi(x)\, dx = \int_M \left(\int_M p(t, x, y)\phi(x)\, d\mathrm{Vol}_g(x) \right) f(y)\, d\mathrm{Vol}_g(y) \xrightarrow{t \to 0^+} \int_M \phi(y) f(y)\, d\mathrm{Vol}_g(y),$$

i.e. $u(t, x) \to f(x)$ in the sense of distributions as $t \to 0^+$.

An "example" is in the case $M = \mathbb{R}^n$ with the Euclidean metric, where $p(t, x, y) = E(t, x - y) = \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/(4t)}$ for $t > 0$. (Note that this "example" does not quite fit our framework since M is non-compact.)

We now want to ask two questions about the heat kernel:

(1) Does such a heat kernel p exist?
(2) What structure does p have, if it exists?

For the first question, we take advantage of the fact that Δ_g has an orthonormal basis of eigenfunctions $\{\phi_k\}_{k=1}^\infty$ in $L^2(M, \mathrm{dVol}_g)$ (hereafter denoted $L^2(M)$) which in fact belong to $C^\infty(M)$. In particular, we suppose

$$-\Delta_g \phi_k = \lambda_k \phi_k, \quad \|\phi_k\|_{L^2(M)} = 1.$$

Then[26] $\lambda_k \to +\infty$. Moreover, for any $f \in L^2(M)$, we can write

$$f(x) = \sum_{k=1}^\infty \langle f, \phi_k \rangle_{L^2(M)} \phi_k(x),$$

[25]To be precise, some uniform control on p as $t \to 0^+$ would be needed in order for the statement to go through; moreover more precise control allows for a more precise statement of convergence. *A posteriori* we see that this will be satisfied.

[26]In fact, it seems plausible, from first principles, that one can show that $\lambda_k \geq ck^\alpha$ for some $c, \alpha > 0$ for sufficiently large k, though I do not know how to do it. It turns out *a posteriori* to be true.

where the sum is interpreted as a convergent sum in $L^2(M)$. It follows that if we explicitly defined

$$u(t,x) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \langle f, \phi_k \rangle_{L^2(M)} \phi_k(x),$$

(i.e. take the "Fourier series solution"), then by construction u is smooth on $(0,\infty) \times M$ (note that for $t > 0$ the factor $e^{-\lambda_k t}$ is exponentially decaying and thus guarantees all series involved converge uniformly on M) and solves the heat equation with initial condition f. Noting that

$$\langle f, \phi_k \rangle_{L^2(M)} = \int_M f(y) \overline{\phi_k(y)} \, \mathrm{dVol}_g(y),$$

plugging this into the above sum, and interchanging the sum and integral (which is permissible given the exponential decay of $e^{-\lambda_k t}$), we obtain

$$u(t,x) = \int_M \left(\sum_{k=1}^{\infty} e^{-\lambda_k t} \phi_k(x) \overline{\phi_k(y)} \right) f(y) \, \mathrm{dVol}_g(y).$$

This suggests we take

$$p(t,x,y) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \phi_k(x) \overline{\phi_k(y)},$$

and indeed one can verify afterward that this does satisfy all of the requirements of the heat kernel. Moreover, from this formula, we see that there is a strong relationship between the structure of the heat kernel and the eigenvalues of the Laplacian: indeed, if we plug in $y = x$ and then integrate in x, i.e. take the *trace* of the heat kernel, we obtain

$$\int_M p(t,x,x) \, \mathrm{dVol}_g(x) = \int_M \sum_{k=1}^{\infty} e^{-\lambda_k t} |\phi_k(x)|^2 \, d\mathrm{Vol}_g(x) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \|\phi_k\|_{L^2(M)}^2 = \sum_{k=1}^{\infty} e^{-\lambda_k t}$$

since $\{\phi_k\}$ is orthonormal in $L^2(M)$.

There is another way to approach the question of existence which also reveals some structure on the heat kernel–namely, we attempt to guess a form of the heat kernel. For example, we can consider the Euclidean heat kernel

$$p_{\mathbb{R}^n}(t,x,y) = \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/(4t)}$$

and attempt to generalize this to manifolds, via

$$p_0(t,x,y) = \frac{1}{(4\pi t)^{n/2}} e^{-d_g(x,y)^2/(4t)}$$

where d_g is the geodesic distance[27] induced by the metric g.

[27] That is, $d_g(x,y)$ is the infimum of lengths of paths connecting x and y, where the length of a path is measured with respect to the metric g. On a compact manifold, this infimum is attained by some geodesic.

Perhaps unsurprisingly this does not quite work. However, we can attempt to modify p_0 to get something closer. More specifically, we multiply p_0 by an ansatz of the form $\sum_{j=0}^k t^j u_j(x,y)$ and consider

$$p_k(t,x,y) = p_0(t,x,y)\sum_{j=0}^k t^j u_j(x,y).$$

Lemma 10.2. *There exist a sequence of $u_j \in C^\infty(M \times M)$ such that*

$$(\partial_t - (\Delta_g)_x)\left(p_0(t,x,y)\sum_{j=0}^k t^j u_j(x,y)\right) \in p_0 t^k C^\infty([0,\infty) \times M \times M).$$

Moreover, the u_j can be chosen so that $p_0(t,x,y)\sum_{j=0}^k t^j u_j(x,y) \to \delta_y(x)$ as $t \to 0^+$ for each $y \in M$.

Furthermore, we can say some things about this sequence u_j. For example:

- The condition $p_0 \sum_{j=0}^k t^j u_j(x,y) \to \delta_y(x)$ as $t \to 0^+$ for any fixed y turns out to force $u_0(x,x) = 1$ for all x.
- Under the above normalization, we then have $u_1(x,x) = \frac{1}{6}S(x)$, where $S(x)$ is the scalar curvature at x.

Remark 16. The functions u_j are derived using the so-called *Minakshisundaram-Pleijel recursion formulas*. See HW 2, Problem 10 for more details.

We now note that if k is large enough, then functions in $p_0 t^k C^\infty([0,\infty) \times M \times M)$ will in fact vanish near $t = 0$, and more generally can be made to have small L^∞ norm when restricted to sufficiently small times t.

This suggests that, at least formally, for $L = \partial_t - (\Delta_g)_x$ we should have

$$p = p_k - (Lp_k *_t p_k) + (Lp_k)^{*_t(2)} *_t p_k - \dots$$

satisfies $Lp = 0$ and $p(t,x,y) \to \delta_y(x)$ as $t \to 0^+$. In fact, this works by making quantitative estimates on $\|(Lp_k)^{*_t(l)}\|_{L^\infty(M\times M\times[0,T])}$ to show that the above series converges (which we can do for T sufficiently small whenever k is sufficiently large). Thus we have that

$$p = p_k - p_k * F_k, \quad F_k = \sum_{l=1}^\infty (Lp_k)^{*_t(l)}$$

and hence:

Theorem 10.3. *The heat kernel indeed satisfies*

$$p(t,x,y) = p_0(t,x,y)\left(\sum_{j=0}^k t^j u_j(x,y) + O(t^{k+1})\right)$$

for any k.

In particular, noting that $p_0(t,x,x) = (4\pi t)^{-n/2}$ for any x, taking the trace of both sides yields

$$\sum_{k=1}^{\infty} e^{-t\lambda_j} = (4\pi t)^{-n/2}\left(\sum_{j=0}^{N} a_j t^j + O(t^{N+1})\right)$$

for any N, where

$$a_j = \int_M u_j(x,x)\,dx.$$

Note that the

$$a_0 = \int_M u_0(x,x)\,dx = \int_M 1\,dx = \text{vol }(M),$$

so

$$\sum_{k=1}^{\infty} e^{-t\lambda_j} = t^{-n/2}\left(\frac{\text{vol }(M)}{(4\pi)^{-n/2}} + O(t)\right). \tag{7}$$

Moreover, if M is two-dimensional, i.e. a surface, then the scalar curvature is twice the Gaussian curvature, so by the Gauss-Bonnet theorem we have

$$a_1 = \int_M u_1(x,x)\,dx = \frac{1}{6}\int_M S(x)\,dx = \frac{1}{3}\int_M K(x)\,dx = \frac{1}{3}(2\pi\chi(M)),$$

where $\chi(M)$ is the Euler characteristic of M. It follows that

$$\sum_{j=0}^{\infty} e^{-\lambda_j t} = \frac{\text{vol}(M)}{4\pi}t^{-1} + \frac{1}{6}\chi(M) + O(t),$$

so in theory knowledge of the eigenvalues of the Laplacian can help determine the volume and topology of the surface by analyzing the behavior of $\sum_{j=0}^{\infty} e^{-\lambda_j t}$ near $t = 0$.

Remark 17. One can use the asymptotics in (7), combined with a theorem attributed variously to Hardy and Littlewood or to Karamata:

$$\int_0^\infty e^{-tx}\,d\mu(x) = \alpha t^{-\beta}(1+o(1)) \quad \text{as } t \to 0 \implies \mu([0,b]) = \frac{\alpha}{\Gamma(\beta+1)}b^{\beta}(1+o(1)) \quad \text{as } b \to \infty$$

(where μ is a positive measure) to prove the *Weyl law* regarding the distribution of eigenvalues of the Laplacian:

Theorem 10.4 (Weyl law)**.** *If $N(\lambda)$ denotes the number of eigenvalues of $-\Delta$ which are at most λ, then*

$$N(\lambda) = \frac{\omega_n}{(2\pi)^n}vol(M)\lambda^{n/2}(1+o(1))$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$.

11. Lecture 11 (05/03): Introduction to Wave Equation

We now consider the wave equation

$$(\partial_t^2 - \Delta)u = 0 \text{ in } (0,\infty)_t \times \mathbb{R}^n_x.$$

Since this involves two derivatives in t, the Cauchy problem requires two pieces of initial data:

$$u(0,x) = f_0(x) \quad \text{and} \quad \partial_t u(0,x) = f_1(x) \text{ on } \mathbb{R}^n.$$

Linear wave equations appear in physics in several contexts, e.g. vibrating strings, Maxwell's equations in electromagnetism, etc.. Nonlinear wave equations also appear in several physical contexts, such as in general relativity (cf. "gravitational waves"), and methods of studying nonlinear wave equations often involve getting an understanding of the linear wave equation first.

Some properties to show:

- Existence/Uniqueness of solutions: Standard in PDE theory.
- Finite speed of propagation: initial data should "propagate" outwards with finite speed. Contrast this with the heat equation, where initial data will generically spread *everywhere* for any $t > 0$ (note that the fundamental solution is positive for *all* (t,x) with $t > 0$).
- Pointwise decay
- Energy conservation/bounds

References: Much of the treatment in the next few lectures will come from the lecture notes of Jonathan Luk and Sung-Jin Oh: [Luk, Oh]

11.1. Solving the Cauchy Equation.

Just like with the heat equation, we will use the Fourier transform (in the spatial variables, i.e. in x) to solve the Cauchy problem for the wave equation. Thus, for $\hat{u}(t,\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x} u(t,x)\,dx$, the equation $(\partial_t^2 - \Delta)u = 0$ turns into

$$\partial_t^2 \hat{u}(t,\xi) + |\xi|^2 \hat{u}(t,\xi) = 0.$$

Recalling that solutions of the ODE $y''(t) + k^2 y(t) = 0$ are given by $y(t) = A\cos(kt) + B\sin(kt)$, it follows that, viewing the above equation as an ODE in t with parameter ξ, we have

$$\hat{u}(t,\xi) = A(\xi)\cos(|\xi|t) + B(\xi)\sin(|\xi|t)$$

for each ξ, where $A(\xi)$ and $B(\xi)$ are some numbers. We can determine $A(\xi)$ and $B(\xi)$ through the initial conditions: we have

$$u(0,x) = f_0(x) \implies \hat{u}(0,\xi) = \widehat{f}_0(\xi)$$

and hence

$$A(\xi)(1) + B(\xi)(0) = \hat{u}(0,\xi) = \widehat{f}_0(\xi) \implies A(\xi) = \widehat{f}_0(\xi).$$

Similarly, $\partial_t \hat{u}(0,\xi) = \widehat{f}_1(\xi)$, and since

$$\partial_t \hat{u}(t,\xi) = -A(\xi)|\xi|\sin(|\xi|t) + B(\xi)|\xi|\cos(|\xi|t)$$

it follows that

$$B(\xi)|\xi| = \partial_t \hat{u}(0,\xi) = \widehat{f}_1(\xi) \implies B(\xi) = \frac{\widehat{f}_1(\xi)}{|\xi|}.$$

Thus we have[28]

$$\hat{u}(t,\xi) = \cos(|\xi|t)\widehat{f}_0(\xi) + \frac{\sin(|\xi|t)}{|\xi|}\widehat{f}_1(\xi).$$

It follows that

$$u(t,x) = (f_0 * \mathcal{F}^{-1}(\cos(|\xi|t)))(x) + \left(f_1 * \mathcal{F}^{-1}\left(\frac{\sin(|\xi|t)}{|\xi|}\right)\right)(x).$$

In particular, we see that if[29] $f_0, f_1 \in \mathcal{S}(\mathbb{R}^n)$, then there exists a solution to the Cauchy problem for the wave equation

Moreover, writing cos and sin in terms of complex exponentials, we have

$$\hat{u}(t,\xi) = e^{i|\xi|t}\left(\frac{\widehat{f}_0(\xi)}{2} + \frac{\widehat{f}_1(\xi)}{2i|\xi|}\right) + e^{-i|\xi|t}\left(\frac{\widehat{f}_0(\xi)}{2} - \frac{\widehat{f}_1(\xi)}{2i|\xi|}\right).$$

From this, we can use the inverse Fourier transform to show:

Proposition 11.1 (Local decay)**.** *For $f_0, f_1 \in C_c^\infty(\mathbb{R}^n)$, we have, for any $R > 0$,*

$$|u(t,x)| \le \frac{C}{t^{n-1}} \text{ for all } t \ge 1, |x| \le R.$$

The constant depends on f_0, f_1, and R, but otherwise not on t or x.

Proof. For convenience, take $n \ge 3$ (the cases $n = 1$ and $n = 2$ can be handled similarly). From the Fourier inversion formula, we have

$$u(t,x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{u}(t,\xi)\,d\xi = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\left(e^{i|\xi|t}\frac{\widehat{f}_1(\xi)}{2i|\xi|} + \dots\right)d\xi$$

(there are three other terms in the ... above). We thus analyze the term

$$\int_{\mathbb{R}^n} e^{ix\cdot\xi}e^{i|\xi|t}\frac{\widehat{f}_1(\xi)}{|\xi|}\,d\xi$$

and aim to show it decays as a power of $t^{-(n-1)}$; the analysis of the remaining three terms is similar. We write the above integral in polar coordinates, with $\xi = r\omega$, $r \in [0,\infty)$, $\omega \in \mathbb{S}^{n-1}$; the integral then becomes

$$\int_{\mathbb{S}^{n-1}}\int_0^\infty e^{itr}e^{irx\cdot\omega}\widehat{f}_1(r\omega)r^{n-2}\,dr\,d\omega$$

[28]Note that the function $\frac{\sin(|\xi|t)}{|\xi|}$ is in fact smooth despite the presence of $|\xi|$ in the denominator.

[29]In fact we can make less restrictive assumptions as well.

(the factor r^{n-2} comes from the factor r^{n-1} from the polar coordinates change of variables divided by r originally in the integrand). Noting that $e^{itr} = \frac{d}{dr}\left(\frac{e^{itr}}{it}\right)$, we can integrate by parts to obtain a factor of t in the denominator; doing so gives

$$\int_{\mathbb{S}^{n-1}} \left[\frac{e^{itr}}{it} g_1(x,r,\omega) r^{n-2}\right]\Big|_0^\infty - \int_0^\infty \frac{e^{itr}}{it}\frac{\partial}{\partial r}\left(g_1(x,r,\omega)r^{n-2}\right)\,dr\,d\omega$$

where $g_1(x,r,\omega) = e^{irx\cdot\omega}\widehat{f_1}(r\omega)$; note g_1 is smooth in all variables considered. We now note:

- If $n-2>0$ (i.e. $n \geq 3$), then the term r^{n-2} vanishes at $r=0$. Moreover, $g_1(x,r,\omega) = e^{irx\cdot\omega}\widehat{f_1}(r\omega)$ decays rapidly as $r \to +\infty$ since $f_1 \in C_c^\infty(\mathbb{R}^n) \implies \widehat{f_1} \in \mathcal{S}(\mathbb{R}^n)$. It follows that the boundary terms in the integration by parts vanish.
- Furthermore, we have
$$\frac{\partial}{\partial r}\left(g_1(x,r,\omega)r^{n-2}\right) = g_2(x,r,\omega)r^{n-3}$$
for some smooth function g_2 (explicitly $g_2 = (n-2)g_1 + r\partial_r g_1$), such that g_2 also decays rapidly as $r \to +\infty$.

Thus, the integral becomes

$$-\frac{1}{it}\int_{\mathbb{S}^{n-1}}\int_0^\infty e^{itr}g_2(x,r,\omega)r^{n-3}\,dr\,d\omega.$$

If $n-3$ is still positive, we can apply the exact same integration by parts argument (in particular that the boundary terms vanish), to get that the integral equals

$$-\frac{1}{t^2}\int_{\mathbb{S}^{n-1}}\int_0^\infty e^{itr}g_3(x,r,\omega)r^{n-4}\,dr\,d\omega$$

for $g_3 = (n-3)g_2 + r\partial_r g_2$. Thus, we keep iterating the integration by parts argument until the exponent in front of r is no longer positive, in which case we get that the integral equals

$$\frac{1}{t^{n-2}}\int_{\mathbb{S}^{n-1}}\int_0^\infty e^{itr}g_{n-1}(x,r,\omega)\,dr\,d\omega$$

for some smooth g_{n-1}, after which a final integration by parts argument yields

$$\frac{1}{it^{n-1}}\int_{\mathbb{S}^{n-1}}\left(\left[e^{itr}g_{n-1}(x,r,\omega)\right]\Big|_0^\infty - \int_0^\infty e^{itr}\partial_r g_{n-1}(x,r,\omega)\,dr\right)d\omega.$$

The boundary term need not vanish, so we cannot iterate the integration by parts procedure again. In this case, we note that the post-factor after $\frac{1}{t^{n-1}}$ is uniformly bounded in t and in x, as long as x itself is restricted to vary in a compact set. This gives the desired estimate. $\square$

Remark 18. Without the assumption of looking only at x in a compact region, we can only get an estimate of the form

$$|u(t,x)| \leq \frac{C}{t^{(n-1)/2}} \text{ for } t \geq 1, x \in \mathbb{R}^n.$$

11.2. **Fundamental solution of the wave operator.** We are interested in a fundamental solution of the wave operator, i.e. a distribution $E \in \mathcal{D}'(\mathbb{R}^{n+1})$ satisfying

$$(\partial_t^2 - \Delta)E = \delta_{(0,0)}.$$

In particular, we are interested in a *forward* fundamental solution, i.e. a fundamental solution E_+ further satisfying

$$\text{supp } E_+ \subset \{(t,x) \in \mathbb{R}^{n+1} : t \geq 0\}.$$

It turns out we can in fact do better:

Proposition 11.2. *There exists a fundamental solution E_+ which satisfies*

$$\text{supp } E_+ \subset \{(t,x) \in \mathbb{R}^{n+1} : |x| \leq t\}.$$

(Note that such a fundamental solution is forward, since $|x| \leq t \implies t \geq 0$.)

Assuming the existence of such a fundamental solution, we can use it as follows:

Proposition 11.3. *Let $F \in \mathcal{D}'(\mathbb{R}^{n+1})$, with*

$$\text{supp } F \subset \{(t,x) \in \mathbb{R}^{n+1} : t \geq -T\} \text{ for some } T > 0$$

(more colloquially "$F \equiv 0$ for $t < -T$"). Then

$$u = E_+ * F$$

is the unique solution in $\mathcal{D}'(\mathbb{R}^{n+1})$ to the equation

$$(\partial_t^2 - \Delta)u = F$$

which also satisfies supp $u \subset \{(t,x) \in \mathbb{R}^{n+1} : t \geq -T\}$.

The main content of the proposition is that the convolution $E_+ * F$ is well-defined. Recall from Lecture 03 that the convolution of two distributions u_1 and u_2 is well-defined if the map supp $u_1 \times$ supp $u_2 \to \mathbb{R}^n$, $(x,y) \mapsto x+y$ is *proper*, or equivalently that $(K - \text{supp } u_1) \cap \text{supp } u_2$ is compact whenever K is compact. In this case, if $K \subset \mathbb{R}^{n+1}$ is compact, and we let $T' = \max_{(t,x)\in K} t$, then

$$(t,x) \in K - \text{supp } F \implies t \leq T' + T$$

since supp $F \subset \{t \geq -T\}$, and hence

$$(K - \text{supp } F) \cap \text{supp } E_+ \subset \{(t,x) \in \mathbb{R}^{n+1} : |x| \leq t \leq T' + T\}.$$

The latter space can be verified to be compact.

Proof. Since $E_+ * F$ is well-defined, we can apply the rules of convolution (in particular relating to constant-coefficient differential operators) to see that

$$(\partial_t^2 - \Delta)(E_+ * F) = ((\partial_t^2 - \Delta)E_+) * F = \delta_{(0,0)} * F = F.$$

Moreover, supp $(E_+ * F) \subset$ supp $E_+ +$ supp $F \subset \{(t,x) : t \geq -T\}$ from the support properties of E_+ and F.

To check uniqueness, it suffices to check that if $u \in \mathcal{D}'(\mathbb{R}^{n+1})$ solves

$$(\partial_t^2 - \Delta)u = 0, \quad \text{supp } u \subset \{(t,x) \in \mathbb{R}^{n+1} : t \geq -T\},$$

then $u = 0$. To do so, we note that

$$u = u * \delta_{(0,0)} = u * ((\partial_t^2 - \Delta)E_+) = (\partial_t^2 - \Delta)(u * E_+) = ((\partial_t^2 - \Delta)u) * E_+ = 0$$

using the rules of convolution; again the main thing to check here is that all convolutions are well-defined (which is indeed the case here due to support properties). □

Corollary 11.4. *The fundamental solution E_+ from Proposition 11.2 is the* ***unique*** *forward fundamental solution of the wave operator.*

That is, even if we only required supp $E \subset \{t \geq 0\}$, we get for free that it must be actually supported in $\{|x| \leq t\}$.

Finally, the forward fundamental solution E_+ can be used to solve the inhomogeneous wave equation:

Theorem 11.5. *Suppose $u \in C^\infty(\mathbb{R}^{n+1})$, and let*

$$F = (\partial_t^2 - \Delta)u, \quad f_0(x) = u(0, x), \quad f_1(x) = \partial_t u(0, x).$$

Then, as distributions, we have

$$u\mathbb{1}_{t\geq 0} = (f_0\delta_{t=0}) * \partial_t E_+ + (f_1\delta_{t=0}) * E_+ + (F\mathbb{1}_{t\geq 0}) * E_+. \tag{8}$$

The proof is similar to the calculation for the fundamental solution of the heat equation.

In particular, since $u(t, \cdot) = f_0 * \mathcal{F}^{-1}(\cos(|\xi|t)) + f_1 * \mathcal{F}^{-1}\left(\frac{\sin(|\xi|t)}{|\xi|}\right)$ solves the wave equation with $F = 0$, (8) suggests that we should have

$$E_+(t, x) = \mathcal{F}^{-1}\left(\frac{\sin(t|\xi|)}{|\xi|}\right)(x)$$

(note then that we'd have $\partial_t E_+(t, x) = \mathcal{F}^{-1}(\cos(t|\xi|))(t, x)$, which is consistent with the solution above.

We can compute this explicitly in some cases:

- For $n = 1$, note that

$$\frac{\sin(t|\xi|)}{|\xi|} = \int_0^t \cos(|\xi|x)\,dx = \frac{1}{2}\int_{-t}^t \cos(\xi x)\,dx = \frac{1}{2}\int_{-t}^t e^{-i\xi x}\,dx = \mathcal{F}\left(\frac{1}{2}\mathbb{1}_{|x|\leq t}\right)(\xi).$$

  (The middle equations follow from the even/odd property of cosine and sine.) This suggests

$$E(t, x) = \frac{1}{2}\mathbb{1}_{t\geq 0}\mathbb{1}_{|x|\leq t}.$$

  Note then that, for $t > 0$, we have

$$\partial_t E(t, x) = \frac{1}{2}\left(\delta(x + t) + \delta(x - t)\right),$$

  in which case (8) becomes

$$u(t, x) = \frac{1}{2}\left(f_0(x - t) + f_0(x + t)\right) + \frac{1}{2}\int_{x-t}^{x+t} f_1(y)\,dy.$$

This recovers *d'Alembert's formula* for the solution of the wave equation in 1 dimension.

- For $n = 3$, we can explicitly compute E to obtain *Kirchoff's formula*:

  Lemma 11.6. *A solution to $(\partial_t^2 - \Delta)u = 0$ in $(0,\infty)\times\mathbb{R}^3$, $u(0,x) = f_0(x)$, $\partial_t u(0,x) = f_1(x)$ is given by*

$$u(t,x) = \frac{1}{4\pi t}\int_{\mathbb{S}_t^2(x)} f_1(y)\, d\mathbb{S}_t^2(x) + \frac{\partial}{\partial t}\left(\frac{1}{4\pi t}\int_{\mathbb{S}_t^2(x)} f_0(y)\, d\mathbb{S}_t^2(x)\right) \tag{9}$$

  where $\mathbb{S}_t^2(x)$ is the ball of radius t centered at x.

  This will be discussed in more detail next lecture.

In general, it turns out we have

$$E(t,x) = c_n \mathbb{1}_{t\geq 0}\chi_+^{-(n-1)/2}(t^2 - |x|^2),$$

where χ_+^a is defined, for Re $a > -1$, as in Example 1.9, and extended for all $a \in \mathbb{C}$ via the property

$$\chi_+^a = \frac{d}{dx}\chi_+^{a+1}$$

(cf. Homework 1). Note that for $k \in \mathbb{N}_{>0}$ we have

$$\chi_+^{-k} = \delta^{(k-1)},$$

so that in particular

$$\text{supp } \chi_+^a = \begin{cases}\{0\} & \text{if } a \in -\mathbb{N}_{>0} \\ [0,\infty) \text{ otherwise} & \end{cases}.$$

Hence, we always have supp $E \subset \{(t,x) : t^2 - |x|^2 \geq 0\} = \{|x| \leq t\}$, while if $n > 1$ is odd (i.e. $(n-1)/2 \in \mathbb{N}_{>0}$), then we in fact have supp $E \subset \{|x| = t\}$. This is known as the *strong Huygens principle.*

Next time: Finite speed of propagation and energy conservation

12. Lecture 12 (05/05): More on the wave equation, finite speed of propagation, energy methods

12.1. **More details on topics from last lecture.** Recall from last lecture that we solved the homogeneous wave equation by using the Fourier transform to convert the PDE into an ODE for each Fourier mode ξ. In particular, we obtained

$$\hat{u}(t,\xi) = \cos(t|\xi|)\widehat{f}_0(\xi) + \frac{\sin(t|\xi|)}{|\xi|}\widehat{f}_1(\xi).$$

Note that $\cos(t|\xi|)$ is uniformly bounded in ξ for any t, while $\frac{\sin(t|\xi|)}{|\xi|}$ is uniformly bounded in ξ *and* decays as $1/|\xi|$ as $\xi \to +\infty$ for any t. It follows that if $f_0 \in H^1(\mathbb{R}^n)$ and $f_1 \in L^2(\mathbb{R}^n)$, then

$$\|u(t,\cdot)\|_{H^1(\mathbb{R}^n)} \leq \|f_0\|_{H^1(\mathbb{R}^n)} + C_t\|f_1\|_{L^2(\mathbb{R}^n)}$$

for some constant C_t depending[30] on t. In particular, $u(t,\cdot) \in H^1(\mathbb{R}^n)$ for each $t > 0$, and $t \mapsto u(t,\cdot)$ is continuous as a map $[0,\infty) \to H^1(\mathbb{R}^n)$. Moreover,

$$\partial_t \hat{u}(t,\xi) = -\sin(t|\xi|)|\xi|\widehat{f}_0(\xi) + \cos(t|\xi|)\widehat{f}_1(\xi),$$

from which we see similarly that $t \mapsto \partial_t u(t,\cdot)$ is continuous from $[0,\infty)$ to $L^2(\mathbb{R}^n)$. Finally, we obtained the above solution by solving a second-order ODE, for which we specified an initial data of initial value and initial first derivative, so we obtain uniqueness. Thus, to summarize:

Theorem 12.1. *For $f_0 \in H^1(\mathbb{R}^n)$ and $f_1 \in L^2(\mathbb{R}^n)$, the unique solution*

$$u \in C([0,\infty); H^1(\mathbb{R}^n)) \cap C^1([0,\infty); L^2(\mathbb{R}^n))$$

to the equation $(\partial_t^2 - \Delta)u = 0$, $u(0,x) = f_0$, $\partial_t u(0,x) = f_1$ is given by

$$u(t,x) = (\mathcal{F}^{-1}(\cos(t|\xi|)) * f_0)(x) + \left(\mathcal{F}^{-1}\left(\frac{\sin(t|\xi|)}{|\xi|}\right) * f_1\right)(x),$$

where the inverse Fourier transform and convolutions are taken in x.

We also mentioned that if E_+ is a forward fundamental solution for the wave operator satisfying

$$\operatorname{supp} E_+ \subset \{(t,x) : |x| \leq t\},$$

then for any $u \in C^\infty(\mathbb{R}^{n+1})$ we have

$$u\mathbb{1}_{t\geq 0} = (f_0\delta_{t=0}) * \partial_t E_+ + (f_1\delta_{t=0}) * E_+ + (F\mathbb{1}_{t\geq 0}) * E_+$$

[30] Explicitly, we have

$$C_t^2 = \sup_{x\in\mathbb{R}} (1+x^2)\frac{\sin^2(tx)}{x^2}.$$

By estimating $\sin^2(tx) \leq 1$ for $|x| \geq 1$ and $\sin^2(tx) \leq t^2x^2$ for $|x| \leq 1$, we can obtain the crude estimate

$$C_t^2 \leq \max(2, 2t^2).$$

As far as I can tell, the estimate cannot be made uniform in t as $t \to +\infty$.

where $F = (\partial_t^2 - \Delta)u$, $f_0(x) = u(0,x)$, $f_1(x) = \partial_t u(0,x)$. This follows by noting that

$$\begin{aligned}(\partial_t^2 - \Delta)(u\mathbb{1}_{t\geq 0}) &= \partial_t(\partial_t(u\mathbb{1}_{t\geq 0})) - (\Delta u)\mathbb{1}_{t\geq 0} \\ &= \partial_t((\partial_t u)\mathbb{1}_{t\geq 0}) + \partial_t(u\delta_{t=0}) - (\Delta u)\mathbb{1}_{t\geq 0} \\ &= ((\partial_t^2 u - \Delta u)\mathbb{1}_{t\geq 0}) + (\partial_t u)\delta_{t=0} + \partial_t(u\delta_{t=0}) \\ &= F\mathbb{1}_{t\geq 0} + f_1\delta_{t=0} + \partial_t(f_0\delta_{t=0}).\end{aligned}$$

Thus, applying Proposition 11.3 gives

$$u\mathbb{1}_{t\geq 0} = (F\mathbb{1}_{t\geq 0}) * E_+ + (f_1\delta_{t=0}) * E_+ + \partial_t(f_0\delta_{t=0}) * E_+,$$

which gives the desired result upon rewriting

$$\partial_t(f_0\delta_{t=0}) * E_+ = \partial_t\left((f_0\delta_{t=0}) * E_+\right) = (f_0\delta_{t=0}) * \partial_t E_+.$$

For $n = 3$, we can obtain an explicit formula for E_+. We note by inspection that the Fourier transform of the surface measure[31] on a sphere of radius r at the origin is

$$\mathcal{F}(d\mathbb{S}_r^2)(\xi) = 4\pi r\frac{\sin(r|\xi|)}{|\xi|}.$$

This is a straightforward computation using polar coordinates in $\mathbb{R}^3$. From this, we see that

$$\mathcal{F}^{-1}\left(\frac{\sin(t|\xi|)}{|\xi|}\right)(x) = \frac{1}{4\pi t}d\mathbb{S}_t^2.$$

From this, using (8) we recover Kirchoff's formula (cf. (9); note that the first term in the RHS of (8) can also be written as $\partial_t\left((f_0\mathbb{1}_{t\geq 0}) * E_+\right)$):

$$u(t,x) = \frac{1}{4\pi t}\int_{\mathbb{S}_t^2(x)} f_1(y)\,d\mathbb{S}_t^2(x) + \frac{\partial}{\partial t}\left(\frac{1}{4\pi t}\int_{\mathbb{S}_t^2(x)} f_0(y)\,d\mathbb{S}_t^2(x)\right)$$

12.2. **Finite Speed of Propagation.** Recall we have:

$$\text{supp } E_+ \subset \{(t,x) \in \mathbb{R}^{n+1} \,:\, |x| \leq t\}.$$

As such, we have:

Theorem 12.2 (Finite Speed of Propagation)**.** *Let $(t_0, x_0) \in (0,\infty) \times \mathbb{R}^n$. Suppose f_0 and f_1 are identically 0 on the set*

$$\{y \in \mathbb{R}^n \,:\, |y - x_0| \leq t_0\}.$$

Then, for u solving the homogeneous wave equation with initial data (f_0, f_1), we have $u(t_0, x_0) = 0$.

[Proof by picture]

[31]That is, the distribution which takes $\phi \in C_c^\infty(\mathbb{R}^3)$ and integrates it on the sphere of radius r centered at the origin.

Corollary 12.3. *Suppose u_1 and u_2 both solve the homogeneous wave equation, and for some $t_1 < t_2$ and $x_0 \in \mathbb{R}^n$, we have*

$$u_1(t_1, y) = u_2(t_1, y) \quad \text{and} \quad \partial_t u_1(t_1, y) = \partial_t u_2(t_1, y) \quad \text{for all } y \text{ s.t. } |y - x_0| \le t_2 - t_1.$$

Then $u_1(t_2, x_0) = u_2(t_2, x_0)$.

Remark 19. The wave equation presented is in some sense a rescaled wave equation (essentially "setting $c = 1$" where c is the speed of the wave). For a more general wave operator $\partial_t^2 - c^2\Delta$, the fundamental solution turns out to be supported in

$$\{(t, x) \in \mathbb{R}^{n+1} : |x| \le ct\}.$$

Then all remarks hold above after replacing t by ct. This then says that "information cannot travel faster than c": i.e. the behavior of the solution at some (t_0, x_0) depends only on the behavior inside the "light cone"; the solution could have drastically different behavior outside the light cone without affecting its behavior inside.

Since for $n \ge 3$ odd we have

$$\text{supp } E_+ \subset \{(t, x) \in \mathbb{R}^{n+1} : |x| = t\}$$

it follows that in such situations we have:

Theorem 12.4 (Strong Huygens Principle)**.** *Let $n \ge 3$ be odd, and let $(t_0, x_0) \in (0, \infty) \times \mathbb{R}^n$. Suppose f_0 and f_1 are identically 0 on the set*

$$\{y \in \mathbb{R}^n : |y - x_0| = t_0\}.$$

Then, for u solving the homogeneous wave equation with initial data (f_0, f_1), we have $u(t_0, x_0) = 0$.

In particular, if f_0 and f_1 are compactly supported, then for any fixed $x_0 \in \mathbb{R}^n$ we have that $u(t, x_0) = 0$ for all *sufficiently large t.*

12.3. **Energy Methods.** Suppose u is a solution to the homogeneous wave equation $\partial_t^2 u - \Delta u = 0$ such that u decays sufficiently quickly as $|x| \to \infty$, and u has some level of differential regularity, say C^1 in both t and x. (For example, in light of the finite speed of propagation proven above, it would suffice to consider solutions u where the initial data f_0 and f_1 are both in $C_c^\infty(\mathbb{R}^n)$.) We define the *energy* associated to the solution as

$$E(t) := \frac{1}{2} \int_{\mathbb{R}^n} |\partial_t u(t, x)|^2 + |\nabla u(t, x)|^2 \, dx, \quad t \ge 0.$$

(In physical terms, the first term in the integral above corresponds to "kinetic energy", while the second term corresponds to "potential energy".)

It turns out we have "conservation of energy" (which is a theme that pops up in physics):

Theorem 12.5. *E is constant in t.*

Proof. Two methods:

(1) Note that by Parseval's theorem we have
$$E(t) = \frac{1}{2}\left(\|\partial_t u(t)\|_{L^2}^2 + \|\nabla u\|_{L^2}^2\right) = C\left(\|\partial_t \hat{u}(t)\|_{L^2}^2 + \|\widehat{\nabla u(t)}\|_{L^2}^2\right),$$
so it suffices to show the last quantity is constant in t. Noting that $|\widehat{\nabla u}(t,\xi)| = |\xi||\hat{u}(t,\xi)|$, the last term can be written as
$$\int_{\mathbb{R}^n} |\partial_t \hat{u}(t,\xi)|^2 + |\xi \hat{u}(t,\xi)|^2 \, d\xi.$$
From
$$\hat{u}(t,\xi) = \cos(|\xi|t)\widehat{f_0}(\xi) + \frac{\sin(|\xi|t)}{|\xi|}\widehat{f_1}(\xi)$$
we have
$$\begin{aligned} |\xi|\hat{u}(t,\xi) &= \cos(|\xi|t)|\xi|\widehat{f_0}(\xi) + \sin(|\xi|t)\widehat{f_1}(\xi) \\ \partial_t \hat{u}(t,\xi) &= -\sin(|\xi|t)|\xi|\widehat{f_0}(\xi) + \cos(|\xi|t)\widehat{f_1}(\xi). \end{aligned}$$
Thus
$$\begin{pmatrix} |\xi|\hat{u}(t,\xi) \\ \partial_t \hat{u}(t,\xi) \end{pmatrix} = \begin{pmatrix} \cos(|\xi|t) & \sin(|\xi|t) \\ -\sin(|\xi|t) & \cos(|\xi|t) \end{pmatrix} \begin{pmatrix} |\xi|\widehat{f_0}(\xi) \\ \widehat{f_1}(\xi) \end{pmatrix}.$$
Note that the matrix is unitary. Thus the norm $(z_1, z_2) \mapsto \sqrt{|z_1|^2 + |z_2|^2}$ on $\mathbb{C}^2$ is preserved, so
$$|\xi \hat{u}(t,\xi)|^2 + |\partial_t \hat{u}(t,\xi)|^2 = |\xi \widehat{f_0}(\xi)|^2 + |\widehat{f_1}(\xi)|^2.$$
Thus, the LHS is constant in t, so E (which can be written in terms of an integral of the LHS above) is also constant in t.

(2) Alternatively, we multiply the equation $\partial_t^2 u - \Delta u = 0$ by $\partial_t u$ and integrate in space. We obtain
$$0 = \int_{\mathbb{R}^n} \partial_t u(t,x)\partial_t^2 u(t,x)\, dx - \int_{\mathbb{R}^n} \partial_t u(t,x)\Delta u(t,x)\, dx.$$
The first term can be written as a time derivative of an integral in x:
$$\int_{\mathbb{R}^n} \partial_t u(t,x)\partial_t^2 u(t,x)\, dx = \frac{d}{dt}\left(\frac{1}{2}\int_{\mathbb{R}^n} |\partial_t u(t,x)|^2\, dx\right).$$
The second term can be rewritten via integration by parts using Green's first identity:
$$-\int_{\mathbb{R}^n} \partial_t u(t,x)\Delta u(t,x)\, dx = \int_{\mathbb{R}^n} \nabla\partial_t u(t,x)\cdot\nabla u(t,x)\, dx = \frac{d}{dt}\left(\frac{1}{2}\int_{\mathbb{R}^n} |\nabla u(t,x)|^2\, dx\right)$$
(note that there is no "boundary term" in the integration by parts by assuming u decays at infinity). It follows that we have
$$0 = \frac{d}{dt}\left(\frac{1}{2}\int_{\mathbb{R}^n} |\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2\, dx\right) = \frac{d}{dt}(E(t)),$$

i.e. E is constant in t.

□

Note that this method does not require us to know the form of the solution ahead of time, only that it satisfies some decay at infinity.

We also have a local energy estimate: for $x_0 \in \mathbb{R}^n$, $R > 0$, and $0 \le t \le R$, define

$$E(t; x_0, R) = \frac{1}{2} \int_{|x-x_0| \le R-t} |\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2 \, dx.$$

In other words, if we draw the "backwards light cone" from (R, x_0), then for $0 \le t \le R$ we have that $E(t; x_0, R)$ integrates the energy of the solution at time t only over the region which slices this cone.

We then have:

Theorem 12.6 (Local energy decay)**.** *For fixed $x_0 \in \mathbb{R}^n$ and $R > 0$, we have that $E(t; x_0, R)$ is non-increasing in t.*

One may expect E to be decreasing "most of the time" given that the region of integration $|x - x_0| \le R - t$ is decreasing in t and in fact should approach 0 as $t \to R^-$; the striking conclusion here is that this decreasing (or more accurately non-increasing) holds *at all times.*

Proof. We proceed similarly as above, noting that in calculating $\frac{dE}{dt}$, we note that we get both a derivative under the integral, and a boundary term from the fact that the domain is changing:

$$\frac{d}{dt}(E(t; x_0, R)) = \int_{|x-x_0| \le R-t} \partial_t u \partial_t^2 u + \nabla u \cdot \nabla \partial_t u \, dx - \frac{1}{2} \int_{|x-x_0|=R-t} |\partial_t u|^2 + |\nabla u|^2 \, d\mathbb{S}^{n-1}_{R-t}(x).$$

We multiply $(\partial_t^2 - \Delta)u$ by $\partial_t u$ and integrate on $\{x \in \mathbb{R}^n : |x - x_0| \le R - t\}$. We note that in integrating the Laplacian by parts, we get

$$-\int_{|x-x_0| \le R-t} \partial_t u \Delta u \, dx = -\int_{|x-x_0|=R-t} \partial_t u \nabla u \cdot \nu \, dS + \int_{|x-x_0| \le R-t} \nabla \partial_t u \cdot \nabla u \, dx.$$

It follows that

$$\begin{aligned}
0 &= \int_{|x-x_0| \le R-t} \partial_t u \partial_t^2 u - \partial_t u \Delta u \, dx \\
&= \int_{|x-x_0| \le R-t} \partial_t u \partial_t^2 u + \nabla \partial_t u \cdot \nabla u \, dx - \int_{|x-x_0|=R-t} \partial_t u \nabla u \cdot \nu \, dS \\
&= \frac{d}{dt}(E(t; x_0, R)) + \int_{|x-x_0|=R-t} \frac{1}{2}|\partial_t u|^2 - \partial_t u \nabla u \cdot \nu + \frac{1}{2}|\nabla u|^2 \, dS.
\end{aligned}$$

We now note that

$$|\partial_t u \nabla u \cdot \nu| \le |\partial_t u||\nabla u| \le \frac{1}{2}(|\partial_t u|^2 + |\nabla u|^2)$$

(the first inequality following from Cauchy-Schwarz, noting that ν is a unit vector, and the second inequality following from the AM-GM inequalitiy). It follows that

$$0 = \frac{d}{dt}(E(t; x_0, R)) + \int_{|x-x_0|=R-t} \frac{1}{2}|\partial_t u|^2 - \partial_t u \nabla u \cdot \nu + \frac{1}{2}|\nabla u|^2 \, dS \geq \frac{d}{dt}(E(t; x_0, R)).$$

Thus $\frac{d}{dt}(E(t; x_0, R)) \leq 0$, as desired. □

This gives another way to derive the finite speed of propagation for the wave equation:

Corollary 12.7 (Finite Speed of Propagation)**.** *If (f_0, f_1) are identically zero on $\{x : |x - x_0| \leq t_0\}$, then $u(t_0, x_0) = 0$.*

Proof. The assumptions give $E(0; x_0, t_0) = 0$, so by the local energy decay, we have $E(t; x_0, t_0) \leq 0$ for $0 \leq t \leq t_0$. Since E is defined as an integral of non-negative terms, it follows that we must have $E(t; x_0, t_0) = 0$ for all t. This in turn implies that $\partial_t u(t, x) = 0$ and $\nabla u(t, x) = 0$ for all (t, x) such that $0 \leq t \leq t_0$ and $|x - x_0| \leq t_0 - t$, so in particular u is constant on the region $\{(t, x) : 0 \leq t \leq t_0, |x - x_0| \leq t_0 - t\}$. Moreover, this constant is zero, since $u(0, x) = f_0(x)$ equals 0 for $|x - x_0| \leq t_0$. It follows that $u(t_0, x_0) = 0$ as well. □

13. Lecture 13: Geometric Optics and nonlinear first-order PDE: method of characteristics (05/10)

The references for this section are Stefanov's *Lecture Notes on Geometric Optics* [Ste], as well as [Eva10] Sections 3.2 and 4.5.

13.1. **Motivation for Geometric Optics Ansatz.** This week we study the behavior of "highly oscillating" solutions for hyperbolic equations. Recall that the solution to

$$(\partial_t^2 - \Delta)u = 0 \text{ in } (0,\infty)\times\mathbb{R}^n, \quad u(0,x) = 0, \quad \partial_t u(0,x) = f(x)$$

is given (via the formula for the Fourier transform $\hat{u}(t,\xi)$) by

$$u(t,x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\frac{\sin(t|\xi|)}{|\xi|}\hat{f}(\xi)\,d\xi = \sum_{\pm}(2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x\cdot\xi\pm t|\xi|)}\frac{\pm 1}{2i|\xi|}\hat{f}(\xi)\,d\xi$$

by writing $\sin(t|\xi|) = \frac{e^{it|\xi|}-e^{-it|\xi|}}{2i}$. We can view the above integrals as saying that the solution u is a superposition of complex exponentials of the form

$$e^{i\phi_\pm(t,x;\xi)}, \quad \phi_\pm(t,x;\xi) = x\cdot\xi \pm t|\xi|.$$

The functions $\phi_\pm$ are the so-called *phase functions* associated to the complex exponential solutions. Moreover, note that these phase functions are (positively) homogeneous of degree 1 in ξ, meaning that

$$\lambda > 0 \implies \phi_\pm(t,x;\lambda\xi) = \lambda\phi_\pm(t,x;\xi).$$

Thus, if for $\xi\in\mathbb{R}^n$ we view it as a scaled version of a unit vector, i.e. $\xi = \lambda\xi_0$ for $\xi_0\in\mathbb{S}^{n-1}$, then the complex exponential is $e^{i\lambda\phi_\pm(t,x;\xi_0)}$, or equivalently if $h = \lambda^{-1}$ this is $e^{i\phi_\pm(t,x;\xi_0)/h}$; note then that large $\xi\in\mathbb{R}^n$ correspond to large $\lambda > 0$ or small $h > 0$. This means that in the regime of "high frequency" (i.e. large "wavenumber" ξ, corresponding to small wavelength), these complex exponentials become increasingly oscillatory.

Thus, when studying hyperbolic equations $(\partial_t^2 - L)u = 0$ in general (where L is a second-order elliptic operator), we are motivated to consider highly oscillatory "solutions" (really ansatzs) of the form

$$u(t,x) = e^{i\phi(t,x)/h}a(t,x)$$

where ϕ is a "phase function", a is an "amplitude profile", and $h > 0$ is viewed as a small number. The idea is that as $h\to 0$, the ansatz above describes a solution which oscillates rapidly, with the oscillation profile determined by ϕ, times a slower-varying amplitude profile a. In general there is no reason to hope that this ansatz actually gives an actual solution, but we may ask if we can obtain an approximate solution, i.e. such that $(\partial_t^2 - L)u$ is "small" in h as $h\to 0$ (for example, perhaps it is "$O(h^N)$" for some N, for a suitably defined notion of big-O).

This is called the *geometric optics* ansatz. The name follows from studying light, which satisfies a wave-particle duality property where, on one hand, it can be modeled by a solution to the wave equation, and on the other hand it has features which are consistent with treating it as a particle and considering certain classical geometric dynamics regarding that particle. The two features end up being related in the high-frequency regime, where certain features of high-frequency solutions to the wave equation are well-described by classical dynamics.

13.2. **Plugging in the ansatz.** To set the setting, we consider a general hyperbolic operator[32]

$$\partial_t^2 - L, \quad L = \sum_{j,k=1}^{n} g^{jk}(x)\partial_{x_j}\partial_{x_k} + \sum_{k=1}^{n} b^k(x)\partial_{x_k} + q(x).$$

For convenience, we assume all coefficients are smooth and bounded, and furthermore the coefficients g^{jk} satisfy that $\{g^{jk}(x)\}_{j,k=1}^n$ forms a positive-definite symmetric matrix for each x. Then, given functions $\phi(t,x)$ and $a(t,x)$ on $\mathbb{R}^{n+1}$ (say smooth for convenience), we want to calculate

$$(\partial_t^2 - L)(e^{i\phi(t,x)/h}a(t,x)).$$

Without doing the full calculation, we can easily see that the result will be of the form

$$e^{i\phi/h}\left(\frac{1}{h^2}(\dots) + \frac{1}{h}(\dots) + (\dots)\right),$$

where each $\dots$ consist of expressions involving derivatives of ϕ and a (but otherwise not depending on h). (See equation (10) below for the precise result.) This is because when we take derivatives on the product $e^{i\phi/h}a$, they either land on the complex exponential $e^{i\phi/h}$, which produces a factor of h^{-1}, or on the amplitude (which produces something involving a and later possibly ϕ, but no factors of h). Since we have a second-order differential operator, this means that we could produce terms as large as h^{-2}.

It follows that if we just arbitrarily choose ϕ and a, then $e^{i\phi/h}a$ not only is not a solution, but $(\partial_t^2 - L)(e^{i\phi/h}a)$ might get quite large as $h \to 0$. However, if we take a closer look at what the $\dots$ coefficients are, in terms of ϕ and a, we may be able to arrange for those coefficients to vanish, meaning that $e^{i\phi/h}$ does not get large as $h \to 0$. Thus, we are motivated to find out what those coefficients are.

Note that we have

$$\partial_k(e^{i\phi/h}a) = e^{i\phi/h}\left(i\frac{\partial_k\phi}{h}a + \partial_k a\right).$$

It follows that

$$\begin{aligned}\partial_j\partial_k(e^{i\phi/h}a) &= e^{i\phi/h}\left(i\frac{\partial_j\phi}{h}\left(i\frac{\partial_k\phi}{h}a + \partial_k a\right) + \partial_j\left(i\frac{\partial_k\phi}{h}a + \partial_k a\right)\right)\\ &= e^{i\phi/h}\left(\frac{1}{h^2}(-\partial_j\phi\partial_k\phi)a + \frac{i}{h}\left(\partial_j\phi\partial_k a + \partial_j(\partial_k\phi a)\right) + \partial_j\partial_k a\right).\end{aligned}$$

[32]It's probably more natural to call these coefficients a^{jk}, b^k, and c. However, I want to use the letter a for the amplitude of our ansatz, and the letter c often denotes the wave speed; hence I use g^{jk} (which later on turns out to be associated to the dual metric of some Riemannian metric) and q (which is another letter often used for potentials).

Hence,

$$\partial_t^2(e^{i\phi/h}a) = e^{i\phi/h}\left(\frac{1}{h^2}(-(\partial_t\phi)^2 a) + \frac{i}{h}\left(\partial_t\phi\partial_t a + \partial_t(\partial_t\phi a)\right) + \partial_t^2 a\right)$$

and

$$\begin{aligned}\sum_{j,k=1}^n g^{jk}\partial_j\partial_k(e^{i\phi/h}a) = e^{i\phi/h}\Bigg[&\frac{1}{h^2}\left(-\sum_{j,k=1}^n g^{jk}\partial_j\phi\partial_k\phi\right)a \\ &+\frac{i}{h}\sum_{j,k=1}^n g^{jk}\left(\partial_j\phi\partial_k a + \partial_j(\partial_k\phi a)\right) + \sum_{j,k=1}^n g^{jk}\partial_j\partial_k a\Bigg]\end{aligned}$$

and

$$\sum_{k=1}^n b^k\partial_k(e^{i\phi/h}a) = e^{i\phi/h}\left(\frac{i}{h}\left(\sum_{k=1}^n b^k\partial_k\phi\right)a + \sum_{k=1}^n b^k\partial_k a\right).$$

It follows that

(10)

$$(\partial_t^2 - L)(e^{i\phi/h}a) = e^{i\phi/h}\left[\frac{1}{h^2}\left(-(\partial_t\phi)^2 + \sum_{j,k=1}^n g^{jk}\partial_j\phi\partial_k\phi\right)a + \frac{i}{h}\mathcal{L}_\phi a + \left(\partial_t^2 - L\right)a\right]$$

where

$$\mathcal{L}_\phi a = \partial_t\phi\partial_t a + \partial_t(\partial_t\phi a) - \sum_{j,k=1}^n g^{jk}\left(\partial_j\phi\partial_k a + \partial_j(\partial_k\phi a)\right) - \left(\sum_{k=1}^n b^k\partial_k\phi\right)a.$$

It follows that for the h^{-2} coefficient above to vanish, we need

$$-(\partial_t\phi)^2 + \sum_{j,k=1}^n g^{jk}\partial_j\phi\partial_k\phi = 0.$$

This is called the *eikonal equation* for the phase function ϕ. Note that this is a *nonlinear* first-order PDE on ϕ. We can rephrase this equation as

$$H(t,x,\partial_t\phi,\nabla\phi) = 0, \quad H(t,x,\tau,\xi) = -\frac{1}{2}\tau^2 + \frac{1}{2}\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k$$

(the factor $1/2$ is introduced for later convenience[33]). Note that the formula for H does not depend on t (though it otherwise depends on τ, ξ, and possibly x depending on how the coefficients g^{jk} depend on x), though it turns out to be convenient to include t as one of the independent variables.

We note that the eikonal equation only depends on the leading-order (i.e. 2nd order) terms of L; note that the first-order terms appear in the h^{-1} and h^0 coefficients only, while the zeroth order term only appears in the h^0 coefficient.

[33]This was not the convention introduced in lecture, but it is one that I would like to adopt in retrospect.

Example 13.1. Suppose $L = \Delta$, so that $(g^{jk}) = \mathrm{Id}$. Then the eikonal equation becomes

$$(\partial_t \phi(t,x))^2 - |\nabla \phi(t,x)|^2 = 0.$$

Then $\phi_\pm(t,x;\xi) = x \cdot \xi \pm t|\xi|$ solves the eikonal equation for any ξ, since $\partial_t \phi = |\xi|$ while $\nabla \phi = \xi$ (in particular both derivatives are constant).

More generally, if $L = c^2\Delta$ for $c > 0$, then $(g^{jk}) = c^2\mathrm{Id}$, and the eikonal equation becomes

$$(\partial_t \phi(t,x))^2 - c^2|\nabla \phi(t,x)|^2 = 0.$$

Then $\phi_\pm(t,x;\xi) = x \cdot \xi \pm ct|\xi|$ solves the eikonal equation.

13.3. **Nonlinear first-order PDE/Hamilton-Jacobi equations and the method of characteristics.** We now turn our attention to a general nonlinear first-order PDE (sometimes called[34] the Hamilton-Jacobi equation)

$$H(x, \nabla \phi(x)) = 0$$

(of which the eikonal equation above is a special case, essentially by renaming t as an x variable, e.g. as x_0). The function $H(x,\xi) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called the *Hamiltonian* in this equation.

To solve this nonlinear PDE on all of $\mathbb{R}^n$, we instead ask if we can solve for the solution, perhaps not on all of $\mathbb{R}^n$ at once, but at least along certain (specially chosen) curves $x(s)$. The advantage of restricting to curves is that this will likely reduce our problem to an ODE (i.e. a differential equation of one variable), whose theory we understand more completely.

As such, suppose that ϕ solves the above equation, and let $x(s)$ be some curve in $\mathbb{R}^n$ (which we can choose later). We can then keep track of how the derivatives of our solution ϕ look along this curve, i.e. we let

$$\xi(s) = \nabla \phi(x(s)).$$

Then ξ must satisfy

$$\dot{\xi}_i(s) = \sum_{j=1}^n \partial_{ij}^2 \phi(x(s)) \dot{x}_j(s).$$

On the other hand, differentiating $H(x, \nabla\phi(x)) = 0$ with respect to the x_i variable gives

$$\partial_{x_i} H(x, \nabla \phi(x)) + \sum_{j=1}^n \partial_{\xi_j} H(x, \nabla\phi(x)) \partial_{ij}^2 \phi(x) = 0.$$

It follows that if we *choose* our curve $x(s)$ to satisfy

$$\dot{x}_j(s) = \partial_{\xi_j} H(x(s), \nabla \phi(x(s)))$$

[34]In some sources, a Hamilton-Jacobi equation is specifically of the form $\partial_t \phi(t,x) + H(x, \nabla\phi(t,x)) = 0$, which is of the form considered here after renaming t as another x variable.

(note that if ϕ is known, then this determines $x(s)$ up to a choice of initial condition $x(0)$), then $\xi(s) = \nabla\phi(x(s))$ *must* satisfy

$$\begin{aligned}\dot{\xi}_i(s) &= \sum_{j=1}^n \partial_{ij}^2\phi(x(s))\dot{x_j}(s)\\ &= \sum_{j=1}^n \partial_{\xi_j}H(x(s),\nabla\phi(x(s)))\partial_{ij}^2\phi(x)\\ &= -\partial_{x_i}H(x(s),\nabla\phi(x(s))) = -\partial_{x_i}H(x(s),\xi(s)).\end{aligned}$$

This means that x (determined if ϕ is known), together with $\xi = \nabla\phi(x)$, satisfy *Hamilton's equations*

$$\dot{x} = \partial_\xi H, \quad \dot{\xi} = -\partial_x H.$$

Thus, if both $(x(0),\xi(0))$ are specified, then a unique trajectory $(x(s),\xi(s))$ is determined from Hamilton's equations.

Moreover, if we then let $z(s) = \phi(x(s))$, we see that

$$\dot{z}(s) = \nabla\phi(x(s))\cdot\dot{x}(s) = \xi(s)\cdot\partial_\xi H(x(s),\xi(s)).$$

If we know $(x(s),\xi(s))$, then we can integrate the above equation to recover $\phi(x(s))$ for any $x(s)$ along our Hamilton trajectory.

This in theory gives us a way to solve for ϕ along certain trajectories. To be pedantic, we'd need to check that this actually does give a solution to the claimed equation. In particular, we need to specify "initial data" for ϕ, say on some hypersurface S, and we want to check that any point away from S can be connected to a point in S via a Hamiltonian trajectory. That is, for every $x \in \mathbb{R}^n$, we'd like for there to exist a Hamiltonian trajectory $(x(s),\xi(s))_{0\le s\le T}$ for some $T \ge 0$ such that $x(T) = x$ and $x(0) \in S$. Moreover, the choice of ξ in the Hamiltonian trajectory cannot be arbitrary either: if we want ξ to represent the derivatives of our solution ϕ, then the starting momentum $\xi(0)$ in the above trajectory must be consistent with the prescribed data ϕ on S. More specifically, we need

$$\xi(0)\cdot\vec{v} = D_{\vec{v}}(\phi|_S) \text{ for all } \vec{v} \in T_{x(0)}S,$$

where $T_{x(0)}S$ is the tangent space of S at the starting point $x(0)$, and $D_{\vec{v}}(\phi|_S)$ is the directional derivative of ϕ in the direction $\vec{v}$.

It turns out that if those dynamical assumptions are satisfied, then our construction does work. The details are checked in Evans, Section 3.2.4.

To conclude this lecture, we see what the Hamiltonian equations are for our situation with the eikonal equation, i.e. with

$$H(t,x,\tau,\xi) = -\frac{1}{2}\tau^2 + \frac{1}{2}\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k.$$

Then Hamilton's equations become

$$\dot{t}(s) = -\tau, \qquad \dot{x}_i(s) = \partial_{\xi_i}\left(\frac{1}{2}G(x,\xi)\right)$$

$$\dot{\tau}(s) = 0, \qquad \dot{\xi}_i(s) = -\partial_{x_i}\left(\frac{1}{2}G(x,\xi)\right)$$

where

$$G(x,\xi) = \sum_{j,k=1}^{n} g^{jk}(x)\xi_j\xi_k.$$

We will study these equations more next lecture, but it suffices to observe that τ is constant (so s is just a linear reparametrization of t), and that (x,ξ) themselves satisfy Hamilton's equations for the Hamiltonian $G(x,\xi)$. It turns out that, if we consider a Riemannian metric (g_{jk}) defined in coordinates by $(g_{jk}(x)) = (g^{jk}(x))^{-1}$, then the geodesic flow with respect to the Riemannian metric, lifted to the cotangent bundle, corresponds exactly to Hamiltonian flow with respect to the Hamiltonian G. More on this next time.

14. Lecture 14 (05/12): Examples of Method of Characteristics and the relationship of Geodesic Flow with Hamiltonian Dynamics

14.1. **Method of characteristics: examples.** Recall the general strategy to solve $H(x, \nabla\phi(x)) = 0$, say with specified "initial data" $\phi|_S$ on some hypersurface S:

- Solve Hamilton's equations
$$\dot{x}(s) = \partial_\xi H(x(s), \xi(s)), \quad \dot{\xi} = -\partial_x H(x(s), \xi(s))$$
to understand the dynamics of Hamiltonian trajectory.
- For $x \in \mathbb{R}^n$, find a Hamiltonian trajectory $(x(s), \xi(s))_{0\le s\le T}$ such that:
  - $x(T) = x$,
  - $x(0) \in S$,
  - $\xi(0)$ is "compatible" with the initial data $\phi|_S$, i.e. that
  $$\xi(0) \cdot \vec{v} = D_{\vec{v}}(\phi|_S)$$
  at $x(0)$ for all $\vec{v} \in T_{x(0)}S$.
- Along the trajectory $x(s)$, $z(s) := \phi(x(s))$ should satisfy
$$\dot{z}(s) = \xi(s) \cdot \dot{x}(s) = \xi(s) \cdot \partial_\xi H(x(s), \xi(s)),$$
with $z(0) = \phi(x(0))$ known. Solve the above ODE for z to recover $\phi(x) = z(T)$.

Remark 20. A general fact is that along a Hamiltonian trajectory, we have that $H(x(s), \xi(s))$ is constant, i.e. "the Hamiltonian is preserved". Moreover, if H is homogeneous of degree m in ξ, then Euler's formula gives
$$\xi \cdot \partial_\xi H = mH,$$
which can simplify the above calculations for $z(s) = \phi(x(s))$ especially in light of the fact that H is constant along the trajectory.

Example 14.1. Transport equations: suppose $H(x, \xi) = b(x)\cdot\xi - f(x)$ for some vector field $b(x)$ (which we assume for convenience is non-vanishing) and some function $f(x)$. Then the corresponding PDE is
$$b(x) \cdot \nabla\phi(x) = f(x).$$
Then the Hamiltonian equations become
$$\dot{x}_i(s) = b_i(x(s)), \quad \dot{\xi}_i(s) = -\sum_{j=1}^n \partial_i b_j(x(s))\xi_j(s) + \partial_i f(x(s)).$$
The first equation tells us that the characteristic curves (in the x variable) are given by *integral curves* of the vector field b (for example, if b is a constant nonzero vector, then the curves are *straight lines* with tangent vector b). In this case, we can actually ignore the equation for ξ, and just directly note that
$$\frac{d}{ds}(\phi(x(s))) = b(x(s)) \cdot \nabla\phi(x(s)) = f(x(s)) \implies \phi(x(T)) = \phi(x(0)) + \int_0^T f(x(s))\, ds.$$

Example 14.2. Suppose $H(t,x,\tau,\xi) = \tau - \frac{1}{2}|\xi|^2$, i.e. the PDE is

$$\partial_t \phi(t,x) = \frac{1}{2}|\nabla \phi(t,x)|^2.$$

Then the Hamiltonian equations become

$$\dot{t} = 1, \quad \dot{x} = -\xi, \quad \dot{\tau} = 0, \quad \dot{\xi} = 0.$$

Note that $\dot{\tau} = 0$ and $\dot{\xi} = 0$ imply that, along the characteristics where we'll be solving for ϕ, we have that $\partial_t \phi$ and $\nabla \phi$ will be *constant*. Furthermore, $\dot{t} = 1 \implies t = s + C$, i.e. t and s are reparametrizations of each other up to a constant.

For example, if $S = \{t = 0\}$, then for a trajectory to satisfy $(t(0), x(0)) \in S$ we clearly have $t = s$. Furthermore, suppose $\phi|_S$ is given by $\phi(0,x) = \frac{1}{2}|x|^2$. Then, along *any* trajectory $(s, x(s), \tau(s), \xi(s))$, we must have $\xi(s) = \xi(0) = \nabla\phi(0, x(0)) = x(0)$, while $\tau(s) = \tau(0) = \frac{1}{2}|\xi(0)|^2 = \frac{1}{2}|x(0)|^2$. Then

$$\dot{x}(s) = -\xi(s) = -x(0) \implies x(s) = x(0) - x(0)s = (1-s)x(0).$$

Note that for any $s < 1$ and any $x \in \mathbb{R}^n$ there exists a unique choice of $x(0)$ such that $x(s) = x$, namely $x(0) = \frac{x}{1-s}$, though all characteristics end up colliding at $s = 1$. Furthermore,

$$\frac{d}{ds}(\phi(s, x(s))) = \tau(s)\dot{t}(s) + \xi(s) \cdot \dot{x}(s) = \frac{1}{2}|x(0)|^2 - |x(0)|^2 = -\frac{1}{2}|x(0)|^2.$$

It follows that

$$\phi(t, x(t)) = \phi(0, x(0)) + \int_0^t \frac{d}{ds}\phi(s, x(s))\, ds = \frac{1}{2}|x(0)|^2 - \frac{t}{2}|x(0)|^2 = \frac{1-t}{2}|x(0)|^2.$$

Thus,

$$\begin{aligned} \phi(t,x) &= \phi(t, x(t)) \quad \text{if } x(0) = \frac{x}{1-t} \\ &= \frac{1-t}{2}|x(0)|^2 \\ &= \frac{|x|^2}{2(1-t)}. \end{aligned}$$

We can go back and check:

$$\partial_t \phi = \frac{|x|^2}{2(1-t)^2}, \quad \nabla\phi = \frac{x}{1-t} \implies \frac{1}{2}|\nabla\phi|^2 = \frac{|x|^2}{2(1-t)^2},$$

i.e. $\partial_t \phi = \frac{1}{2}|\nabla\phi|^2$, as desired. Note that this solution blows up as $t \to 1^-$.

We now return to our nonlinear PDE of interest, the eikonal equation

$$H(t, x, \partial_t\phi, \nabla\phi) = 0 \text{ where } H(t,x,\tau,\xi) = -\tau^2 + \sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k.$$

Our initial hypersurface will be $S = \{t = 0\}$.

We see that Hamilton's equations become

$$\dot{t}(s) = -\tau, \qquad \dot{x}_i(s) = \partial_{\xi_i}\left(\frac{1}{2}G(x,\xi)\right)$$

$$\dot{\tau}(s) = 0, \qquad \dot{\xi}_i(s) = -\partial_{x_i}\left(\frac{1}{2}G(x,\xi)\right)$$

where

$$G(x,\xi) = \sum_{j,k=1}^{n} g^{jk}(x)\xi_j\xi_k.$$

In particular, τ is constant, $t(s) = t(0) - \tau s$, and $(x(s), \xi(s))$ is a Hamiltonian trajectory for the Hamiltonian $\frac{1}{2}G(x,\xi)$. Note then that $G(x(s),\xi(s))$ is constant, and it equals $G(x(0),\xi(0))$ (which also equals $\tau(0)$).

14.2. **Some Riemannian geometry: geodesic flow.** How do we make sense of the above Hamiltonian flow? It turns out we can do the following: let g be the Riemannian metric whose dual Riemannian metric is $\sum_{j,k=1}^{n} g^{jk}(x)\partial_{x_j} \otimes \partial_{x_k}$, i.e. $g = \sum_{j,k=1}^{n} g_{jk}(x)dx_j \otimes dx_k$ where, as matrices, we have

$$(g_{jk}(x)) = (g^{jk}(x))^{-1}.$$

Note for any tangent vector $v = \sum v^k \partial_{x^k} \in T_x\mathbb{R}^n$, if we let[35]

$$\xi = \sum \xi_j \, dx^j, \quad \xi_j = \sum g_{jk}(x)v^k,$$

then for any other tangent vector $w \in T_x\mathbb{R}^n$ we have

$$\xi(w) = \sum_j \xi_j w^j = \sum_{jk} g_{jk}(x)v^k w^j = g_x(v,w).$$

Moreover, $v^j = \sum_k g^{jk}(x)\xi_k$, and

$$g_x(v,v) = \xi(v) = \sum_j \xi_j v^j = \sum_{jk} g^{jk}(x)\xi_j\xi_k = G(x,\xi).$$

Thus, $G(x,\xi)$ is the dual metric function on $T^*\mathbb{R}^n$ with respect to the metric given by (g_{jk}).

Given this Riemannian metric, we can consider when a curve is a *geodesic* with respect to this metric. Recall a curve $x(s)$ is a geodesic if $\nabla_{x'}x' = 0$, where ∇ is the Levi-Civita connection associated with g; in coordinates this equation becomes the system of n equations

$$x_k''(s) + \sum_{i,j=1}^{n} \Gamma_{ij}^k(x(s))x_i'(s)x_j'(s) = 0,$$

[35]This association of a covector ξ to a vector v via the metric is sometimes called the *musical isomorphism* or *raising/lowering indices.*

where the Γ_{ij}^k are the Christoffel symbols satisfying $\nabla_{\partial_i}\partial_j = \sum_{k=1}^n \Gamma_{ij}^k \partial_k$; in coordinates we have

$$\Gamma_{ij}^k(x) = \frac{1}{2}\sum_{m=1}^n g^{km}(x)(\partial_i g_{jm}(x) + \partial_j g_{im}(x) - \partial_m g_{ij}(x)).$$

Then, it turns out we have:

Theorem 14.3. *Suppose $(x(s), \xi(s))$ is a Hamiltonian trajectory for $\frac{1}{2}G$, where $G(x,\xi) = \sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k$, and let τ^2 be the constant value of G along this trajectory. Then $x(s)$ is a geodesic with respect to the metric g_{ij}. Moreover, the tangent vector $\dot{x}(s)$ satisfies*

$$g_{x(s)}(\dot{x}(s), v) = \xi(s)\cdot v \tag{11}$$

for all $v \in T_{x(s)}\mathbb{R}^n$, and $g_{x(s)}(\dot{x}(s), \dot{x}(s)) = \tau^2$ for all s.

The proof is given in Appendix A. Note that (11) is easy to verify, since we have

$$\dot{x_i}(s) = \partial_{\xi_i}\left(\frac{1}{2}\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k\right) = \sum_{j=1}^n g^{ij}(x)\xi_j(s),$$

from which we see that $\xi_j(s) = \sum_{k=1}^n g_{jk}(x)\dot{x_k}(s)$, i.e. that

$$\xi\cdot\left(\sum_{j=1}^n v^j\partial_{x^j}\right) = \sum_{j=1}^n \xi_j(s)v^j = \sum_{j,k=1}^n g_{jk}(x)\dot{x_k}(s)v^j = g_{x(s)}\left(\dot{x}(s), \left(\sum_{j=1}^n v^j\partial_{x^j}\right)\right),$$

as desired.

Since t is also progressing in s at a rate of $|\tau|$, it follows that:

Corollary 14.4. *Suppose $(t(s), x(s), \tau(s), \xi(s))$ is a Hamiltonian trajectory of*

$$H(t,x,\xi,\tau) = \frac{1}{2}\left(\tau^2 - \sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k\right).$$

Then $(t(s), x(s))$ traces out a curve in $\mathbb{R}^{n+1}$ which can be reparametrized in terms of t, i.e. as $(t, x(t))$, such that $x(t)$ is a unit-speed *geodesic with respect to the Riemannian metric (g_{ij}).*

Example 14.5. Suppose $G = \xi_1^2 + \frac{1}{x_1^2}\xi_2^2$. Then

$$\dot{x_1} = \xi_1, \quad \dot{x_2} = \frac{\xi_2}{x_1^2}, \quad \dot{\xi_1} = \frac{1}{x_1^3}\xi_2^2, \quad \dot{\xi_2} = 0.$$

Then ξ_2 is constant. If $\xi_2 = 0$, then ξ_1 is constant as well. Then x_2 is constant, and x_1 travels at constant speed. If $\xi_2 \neq 0$, then we note that G is constant, and $\xi_1 = \dot{x_1}$ implies

$$\dot{x_1}^2 = G - \frac{\xi_2^2}{x_1^2}.$$

One can check that, up to constant-speed reparametrization, we have

$$x_1(s) = \sqrt{\frac{\xi_2^2}{G} + Gs^2}$$

in which case

$$\dot{x_2}(s) = \frac{\xi_2}{\frac{\xi_2^2}{G} + Gs^2} = \frac{\xi_2 G}{\xi_2^2 + G^2 s^2} \implies x_2(s) = \arctan\left(\frac{Gs}{\xi_2}\right) + x_2(0).$$

Note that if we let $r_0 = |\xi_2|/\sqrt{G}$ and $v = \sqrt{G}$, then

$$x_1(s) = \sqrt{r_0^2 + (vs)^2}, \quad x_2(s) = x_2(0) \pm \arctan\left(\frac{vs}{r_0}\right),$$

which is in fact the radius and angle of a straight line in $\mathbb{R}^2$ located a distance of r_0 from the origin parametrized at speed v. (In that case, $|\xi_2| = r_0 v$, and ξ_2 being constant is the "conservation of angular momentum."). This is no surprise, given that the corresponding metric $g = dx_1^2 + x_1^2 dx_2^2$ is the Euclidean metric in polar coordinates (x_1 the radial variable, x_2 the angular variable; rewriting this gives $g = dr^2 + r^2 d\theta^2$ which may be more familiar).

Moreover, if we now call $r = x_1$ and $\theta = x_2$, the corresponding operator

$$L = \partial_r^2 + r^{-2}\partial_\theta^2$$

is, up to lower-order terms, the Laplacian in polar coordinates in $\mathbb{R}^2$, since in fact

$$\Delta = \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\theta^2.$$

15. Lecture 15 (05/17): Back to the geometric optics ansatz and applications

Recall that the reason we investigated solving first-order PDE was in considering geometric optics ansatz:

$$(\partial_t^2 - L)(e^{i\phi/h}a) = e^{i\phi/h}\left(\frac{1}{h^2}H(t,x,\partial_t\phi,\nabla\phi)a + \frac{i}{h}\mathcal{L}_\phi(a) + (\partial_t^2 - L)a\right), \tag{12}$$

from which we'd like to impose $H(t,x,\partial_t\phi,\nabla\phi) = 0$, i.e. that ϕ solves the *eikonal* equation. To do so, we saw that we wanted to study the Hamilton trajectories with respect to the Hamiltonian

$$H(t,x,\tau,\xi) = -\tau^2 + \sum_{j,k=1}^{n} g^{jk}(x)\xi_j\xi_k,$$

whose trajectories (projected to the (t,x) variables) traced out a curve which in turn can be traced by $(t,x(t))$, where $x(t)$ is a unit-speed geodesic with respect to the metric $(g_{jk}(x)) = (g^{jk}(x))^{-1}$.

15.1. **Finishing up the eikonal equation.** For concreteness, we also need to impose initial conditions on ϕ on some initial hypersurface, which we take to be $S = \{t = 0\}$. We'll make the particular choice

$$\phi(0,x) = x\cdot\xi_0$$

for some $\xi_0 \in \mathbb{R}^n$. This corresponds to solving the following problem:

$$\begin{gathered}\text{find an “approximate solution” } u \text{ to } (\partial_t^2 - L)(e^{i\phi/h}a) = 0\\ \text{with initial condition } u(0,x) = e^{ix\cdot\xi_0/h}a(x).\end{gathered}$$

If a is compactly supported in a small neighborhood of some point, say x_0, then one can view the initial value $e^{ix\cdot\xi_0/h}a(x)$ as a "wavepacket" spatially supported near x_0 with momentum ξ_0/h; as $h \to 0$ the momentum gets larger in magnitude (though with a fixed direction).

Thus, for (t_0,x_0), we want to find a Hamiltonian trajectory $(t(s),x(s),\tau(s),\xi(s))$ and some "final[36] time" T such that

- $(t(T),x(T)) = (t_0,x_0)$,
- $t(0) = 0$,
- $\xi(0) = \xi_0$.

Recall that $\dot\tau = \partial_t H = 0$, so τ is constant, and in fact we must have

$$\tau = \pm|\xi_0|_{g,x(0)},$$

where $|\xi|^2_{g,x} = G(x,\xi)$. In particular, we have two possible choices for τ (this corresponds roughly to having an "incoming" or "outgoing" wave.) Moreover, we have

$$\dot t = -\tau \implies t = -\tau s.$$

[36]Actually, we can take $T < 0$: as we see below, this is needed if we subsequently take $\tau > 0$.

(Thus, $T = -t_0/\tau$.) Finally, $x(s)$ traces a geodesic of speed $|\tau|$. Reparametrizing by t, we see then that $x(t)$ is a geodesic of speed 1, since $\left|\frac{dt}{ds}\right| = |\tau|$. Moreover, the initial velocity $\dot{x}(0)$ (when parametrized by t) has a corresponding momentum parallel to ξ_0.

This describes the trajectories, though as a reminder we also want to solve the actual eikonal equation, for which understanding the trajectories gets us almost there. We now note that

$$\frac{d}{ds}(\phi(t(s), x(s))) = (\tau, \xi) \cdot \partial_{\tau,\xi} H = 2H = 0,$$

since H is homogeneous of degree 2 in (τ, ξ), and our trajectories stay in $\{H = 0\}$. Thus, the initial value $\phi(0, x) = x \cdot \xi_0$ is simply transported along the trajectories, with no modification.

Example 15.1. Suppose $(g^{ij}) = c^2 \mathrm{Id}$. In this case, τ is independent of (t, x), since $\tau = \pm c|\xi_0|$. Moreover, ξ is constant, so $\xi = \xi_0$. Finally,

$$\dot{x} = c^2 \xi = c^2 \xi_0 \implies x(s) = x(0) + c^2 \xi_0 s.$$

Thus, if $(t(s), x(s)) = (t, x)$, then $t = -\tau s = \mp c|\xi_0| s \implies s = \mp \frac{t}{c|\xi_0|}$, so

$$x = x(s) = x(0) + c^2 \xi_0 \left(\mp \frac{t}{c|\xi_0|} \right) = x(0) \mp ct \frac{\xi_0}{|\xi_0|} \implies x(0) = x \pm ct \frac{\xi_0}{|\xi_0|}.$$

Finally, from above we have that ϕ maintains its value along the trajectory, so

$$\phi(t, x) = x(0) \cdot \xi_0 = \left(x \pm ct \frac{\xi_0}{|\xi_0|} \right) \cdot \xi_0 = x \cdot \xi_0 \pm ct|\xi_0|.$$

Remark 21. Note that the speed of the trajectory $x(t)$ measured with respect to t does not depend on the magnitude of ξ_0, only its direction (the speed measured with respect to the original trajectory parameter s *does* depend on $|\xi_0|$, but then so does $t(s)$; the two end up canceling here). This is a sign that the linear wave equation is *not dispersive*, i.e. high frequencies are not "spreading" faster than low frequencies.

15.2. **Solving the transport equation in the h^{-1} coefficient.** Now that we've solved the eikonal equation, thus removing the h^{-2} component in (12), let's consider the h^{-1} coefficient. This is $i\mathcal{L}_\phi a$, where

$$\mathcal{L}_\phi a = \partial_t \phi \partial_t a + \partial_t (\partial_t \phi a) - \sum_{j,k=1}^{n} g^{jk} \left(\partial_j \phi \partial_k a + \partial_j (\partial_k \phi a) \right) - \left(\sum_{k=1}^{n} b^k \partial_k \phi \right) a.$$

Note that we can separate $\mathcal{L}_\phi$ into a first order and zeroth order part, i.e. $\mathcal{L}_\phi a = -\mathcal{L}_\phi^{(1)} a + \mathcal{L}_\phi^{(0)} a$ where

$$\mathcal{L}_\phi^{(1)} a = 2 \left(-\partial_t \phi \partial_t a + \sum_{j,k=1}^{n} g^{jk} \partial_j \phi \partial_k a \right)$$

and

$$\mathcal{L}_\phi^{(0)}a = \psi(t,x)a(t,x), \quad \psi = \left(\partial_t^2\phi - \sum_{j,k=1}^n g^{jk}\partial_j\phi\partial_k\phi - \sum_{k=1}^n b^k\partial_k\phi\right).$$

Thus, we can rewrite the condition as $\mathcal{L}_\phi^{(1)}a = \mathcal{L}_\phi^{(0)}a$, with the LHS a vector field applied to a, and the RHS a function multiplied by a. This is essentially a transport equation, with a zeroth order term.

If we wanted to solve $\mathcal{L}_\phi^{(1)}a = 0$, we would use the method of characteristics (cf. Example 14.1), in which case we want to consider curves in $\mathbb{R}^{n+1}$ whose velocities agreed with the coefficients in $\mathcal{L}_\phi^{(1)}a$, i.e. curves $(t(s), x(s))$ satisfying

$$\dot{t}(s) = -\partial_t\phi(t(s), x(s)), \quad \dot{x_k}(s) = \sum_{j,k=1}^n g^{jk}(x(s))\partial_j\phi(t(s), x(s)). \tag{13}$$

Such curves are *precisely* the Hamiltonian trajectories with respect to H. So, in considering the dynamics needed to solve the eikonal equation, to solve away the h^{-2} term, we also come up with the relevant dynamics to solve the transport equation in the h^{-1} term.

In this case, we do have a zeroth order term. It turns out, to solve the transport equation with zeroth order terms, that we can still consider the same characteristics; the resulting ODE to solve for the value ends up differing. Explicitly, if $(t(s), x(s))$ satisfies (13), so that $\frac{d}{ds}(a(t(s), x(s))) = \mathcal{L}_\phi a(t(s), x(s))$, then our ODE becomes

$$\frac{d}{ds}(a(t(s), x(s))) = \psi(t(s), x(s))a(t(s), x(s)),$$

which, given a known ψ (which can be determined having solved for ϕ), is just an ODE on the unknown $a(t(s), x(s))$. So we can integrate that ODE to solve for a along these characteristics. In particular, the geodesic dynamics (i.e. Hamiltonian dynamics for our Hamiltonian H) once again show up in solving for a.

15.3. **Solving the remaining terms: asymptotic series ansatz.** Now that we've solved away the h^{-1} component, we see that we are left with an order h^0 component $(\partial_t^2 - L)a$. Given that we've already solved for a, it is in general unreasonable to expect that $(\partial_t^2 - L)a = 0$. So it seems initially that we can say:

$$\text{there exists } \phi, a \text{ such that } (\partial_t^2 - L)(e^{i\phi/h}a) = O(1) \text{ as } h \to 0.$$

However, we can do better. Note that a itself is independent of h–what if we added a correction term to a, depending on h, to possibly solve away the $O(1)$ error? That is, we make the ansatz

$$a(t, x; h) = a^0(t, x) + ha^1(t, x).$$

Let us suppose for convenience that ϕ still solves the eikonal equation. Then

$$(\partial_t^2 - L)(e^{i\phi/h}a) = e^{i\phi/h}\left(\frac{i}{h}\mathcal{L}_\phi a^0 + i\mathcal{L}_\phi a^1 + (\partial_t^2 - L)a^0 + h(\partial_t^2 - L)a^0\right).$$

We once again require $\mathcal{L}_\phi a^0 = 0$, and now we see if we can arrange

$$i\mathcal{L}_\phi a^1 + (\partial_t^2 - L)a^0 = 0.$$

This is indeed possible, once a^0 has been solved for: it's essentially the same process as described above, namely decomposing $\mathcal{L}_\phi$ into a first-order term and a zeroth-order term, in which case we are solving a transport equation (where the vector field is still the same one as above, i.e. the ones obtained by Hamiltonian dynamics!), now just with an inhomogeneity (namely the $(\partial_t^2 - L)a^0$ term).

It follows, by adding the order 1 correction, that

$$\text{there exists } \phi(t,x), a(t,x;h) \text{ such that } (\partial_t^2 - L)(e^{i\phi/h}a) = O(h) \text{ as } h \to 0.$$

We don't have to stop there of course: if in general we make the ansatz

$$a(t,x;h) = \sum_{j=0}^{N} a^j(t,x)h^j,$$

then (continuing to assume that ϕ solves the eikonal equation) we have

$$(\partial_t^2 - L)(e^{i\phi/h}a) = e^{i\phi/h}\left(\frac{i}{h}\mathcal{L}_\phi a^0 + \sum_{j=1}^{N} h^{j-1}(i\mathcal{L}_\phi a^j - (\partial_t^2 - L)a^{j-1}) + h^N(\partial_t^2 - L)a^N\right).$$

It follows that if we inductively solve

$$i\mathcal{L}_\phi a^j - (\partial_t^2 - L)a^{j-1} = 0,$$

then

$$(\partial_t^2 - L)(e^{i\phi/h}a) = O(h^N).$$

Having solved for these a^j, we may be tempted to consider

$$a(t,x;h) := \sum_{j=0}^{\infty} a^j(t,x)h^j$$

which should heuristically satisfy

$$(\partial_t^2 - L)(e^{i\phi/h}a) = \text{“}O(h^\infty)\text{''} \quad (\text{i.e. } O(h^N) \text{ for each N}).$$

However, there's no guarantee that the sum $\sum_{j=0}^\infty a^j(t,x)h^j$ converges in any meaningful way. Nonetheless, we can make sense of it in the sense of *asymptotic series* (cf. Taylor expansions), as follows:

Lemma 15.2 (Borel Summation Lemma)**.** *Let $a^0(y), a^1(y), \dots$ be any sequence of functions in $C^\infty(Y)$ (where Y is any manifold, say $\mathbb{R}^n$). Then, there exists a function $a(y,h) \in C^\infty(Y \times [0,1))$ such that, for any N and any compact subset $K \subset Y$, there exists $C_{N,K} > 0$ such that*

$$\left|a(y,h) - \sum_{j=0}^{N-1} a^j(y)h^j\right| \le C_{N,K}h^N$$

for all $y \in K$.

The proof is somewhat explicit: we can take

$$a(y,h) = \sum_{j=0}^{\infty} \chi\left(\frac{h}{\epsilon_j}\right) a^j(y) h^j,$$

where $\chi \in C_c^\infty(\mathbb{R})$ is identically one for $|h| < 1/2$ and supported in $|h| < 1$, and ϵ_j is a rapidly decreasing sequence of positive numbers whose rate of decrease we can control. Note then that, for a *fixed* $h > 0$, the above sum is a finite sum (we sum only over j where $h < \epsilon_j$, which happens for only finitely many j), and we can control ϵ_j to decrease sufficiently fast so that the sum is uniformly bounded as $h \to 0$.

Remark 22. The "zero-dimensional" version of this (i.e. when "Y is a point", so functions in $C^\infty(Y)$ are just numbers) says that the Taylor coefficients of a smooth function can be *completely arbitrary*: any sequence (even those growing as fast as you like) is the set of Taylor coefficients for some *smooth* function. Such a function will necessarily non-analytic; in fact the Taylor coefficients of a (say 1-dimensional) analytic function cannot grow faster than exponential (equivalently, the jth derivative at a point cannot grow faster than $C^j j!$ for some $C > 0$), essentially due to Cauchy's integral formula.

For $t \geq 0$ and $\xi \in \mathbb{R}^n$ nonzero, let $(x(t), \xi(t))$ be a Hamiltonian trajectory where $x(0) = x$ and $\xi(0) = \frac{\xi}{|\xi|_g}$; here $|\xi|_g^2 = \sum g^{jk}\xi_j\xi_k$. Note then that $x(t)$ is a geodesic of speed 1. Set

$$\exp_{t,\xi}(x) = x(t).$$

Let's make the following dynamical assumption:

Assumption 15.3. The set $U \subset \mathbb{R}^n$ is bounded, and $T > 0$ satisfies the property that $\exp_{t,\xi} : U \to \mathbb{R}^n$ is a diffeomorphism onto its image for all $0 \leq t \leq T$ and all $\xi \in \mathbb{R}^n \backslash \{0\}$.

Under this assumption, we have that for any $\xi \in \mathbb{R}^n \backslash \{0\}$ there exists a unique solution to the eikonal equation

$$H(t, x, \partial_t \phi, \nabla \phi) = 0, \quad \phi(0, x) = x \cdot \xi \text{ for } x \in U$$

which is well-defined on

$$\Omega_\xi := \{(t, x) \,:\, 0 \leq t \leq T, x = \exp_{t,\xi}(x_0) \text{ for some } x_0 \in U\}.$$

Then, the geometric optics ansatz developed in the previous few lectures give the following result:

Theorem 15.4. *Suppose U and T satisfy Assumption 15.3, and let $a \in C_c^\infty(U)$, $\xi \in \mathbb{R}^n \backslash \{0\}$. Then there exists a family of smooth functions $u(t, x; h)$ for $h > 0$, supported on Ω_ξ, with the property that*

$$u(0, x; h) = e^{ix\cdot\xi/h} a(x)$$

and for every N there exists C_N such that

$$|(\partial_t^2 - L)u(t, x; h)| \leq C_N h^N \quad \text{uniformly on } \Omega_\xi.$$

15.4. **Approximate solution for hyperbolic equations.** The development of this ansatz also allows us to construct an "approximate solution" for the general wave equation. Here the notion of "approximate inverse" must be taken with a grain of salt–the approximation here is in the sense of *regularity*, not in terms of size. Thus, the following theorem shows that we can construct a function (for short times) which solves our hyperbolic equation $(\partial_t^2 - L)u = 0$ *up to smooth errors* (which *a priori* could be large in size, but at the very least contains no additional singularities):

Theorem 15.5. *Suppose U and T satisfy Assumption 15.3, and suppose $f_0, f_1 \in L^2(\mathbb{R}^n)$ and are compactly supported in U. Then there exists $u \in C^1([0,T]; \mathcal{E}'(\mathbb{R}^n))$ such that*

$$\begin{aligned}(\partial_t^2 - L)u &\in C_c^\infty([0,T] \times \mathbb{R}^n), \\ \lim_{t\to 0^+} [u(t) - f_0] &\in C_c^\infty(\mathbb{R}^n), \\ \lim_{t\to 0^+} [\partial_t u(t) - f_1] &\in C_c^\infty(\mathbb{R}^n).\end{aligned}$$

Moreover,

$$\text{supp } u \subset \bigcup_{\xi \in \mathbb{R}^n \setminus \{0\}} \Omega_\xi.$$

The motivation is that the solution to the Cauchy problem to the linear wave equation is given by

$$u(t,x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(\cos(t|\xi|) \hat{f}_0(\xi) + \frac{\sin(t|\xi|)}{|\xi|} \hat{f}_1(\xi) \right) d\xi.$$

Writing cos and sin in terms of complex exponentials, and writing ξ in polar coordinates $\xi = \omega/h$, with $h = 1/|\xi|$, we can rewrite the solution as

$$u(t,x) = (2\pi)^{-n} \sum_{k=0,1,\sigma=\pm} \int_{\mathbb{S}^{n-1}} \int_0^\infty e^{i\phi_\pm(t,x;\omega)/h} a_{k,\sigma}(\omega) \hat{f}_k(\omega/h)\, h^{-n-1-j}\, dh\, d\mathbb{S}^{n-1}(\omega)$$

where

$$a_{0,+} = \frac{1}{2}, \quad a_{0,-} = \frac{1}{2}, \quad a_{1,+} = \frac{1}{2i}, \quad , a_{1,-} = -\frac{1}{2i}.$$

Proof. Let $\chi \in C_c^\infty(U)$ satisfy $\chi \equiv 1$ on the support of f_0 and f_1, and let $\rho \in C_c^\infty(\mathbb{R})$ satisfy $\rho(h) = 1$ for all $|h| < 1$. We take

$$u(t,x) = (2\pi)^{-n} \sum_{k=0,1,\sigma=\pm} \int_{\mathbb{S}^{n-1}} \int_0^\infty e^{i\phi_\sigma(t,x;\omega)/h} a_{k,\sigma}(t,x;\omega,h) \rho(h) \hat{f}_k(\omega/h)\, h^{-n-1-j}\, dh\, d\mathbb{S}^{n-1}(\omega)$$

where $\phi_\pm(t,x;\omega)$ solve

$$-(\partial_t\phi_\pm)^2 + \sum_{j,k=1}^n g^{jk}(x) \partial_{x_j}\phi \partial_{x_k}\phi = 0, \quad \phi_\pm(0,x;\omega) = x\cdot\omega, \quad \pm\partial_t\phi_\pm(0,x;\omega) > 0.$$

(we require the differential equation to be satisfied at least on Ω_ω for all $\omega \in \mathbb{S}^{n-1}$ in case it cannot be solved on all of $\mathbb{R}^n$), and $a_{k,\pm}(t,x;\omega,h)$ satisfies an asymptotic expansion

$$a_{k,\pm}(t,x;\omega,h) \sim \sum_{j\geq 0} a_{k,\pm}^{(j)}(t,x;\omega)h^j$$

where

$$i\mathcal{L}_{\phi_\pm} a_{k,\pm}^{(0)} = 0, i\mathcal{L}_{\phi_\pm} a_{k,\pm}^{(j)} + (\partial_t^2 - L)a_{k,\pm}^{(j-1)} = 0 \text{ for } j \geq 1$$

and at $t = 0$ we have

$$a_{0,+}(0,x;\omega) = \frac{1}{2}\chi(x), \quad a_{0,-;\omega} = \frac{1}{2}\chi(x), \quad a_{1,+} = \frac{1}{2i}\chi(x), \quad , a_{1,-} = -\frac{1}{2i}\chi(x).$$

(i.e. $a_{k,\pm}^{(0)}$ are the above values at $t = 0$, and $a_{k,\pm}^{(j)} = 0$ at $t = 0$). Then, by differentiating under the integral sign, we see that

$$(\partial_t^2 - L)u(t,x) = (2\pi)^{-n} \sum_{k=0,1,\sigma=\pm} \int_{\mathbb{S}^{n-1}} \int_0^\infty e^{i\phi_\sigma(t,x;\omega)/h} O(h^\infty)\rho(h)\hat{f}_k(\omega/h)\, h^{-n-1-j}\, dh\, d\mathbb{S}^{n-1}(\omega).$$

We note that this integral converges due to the $\rho(h)$ term controlling behavior for "large h" (i.e. for small ξ), and the "$O(h^\infty)$" error controlling behavior for "small h" (i.e. for ξ large). Moreover, taking any derivatives in t and x will return an expression of the same form, since the worst that could happen is multiplying the integrand by a factor of h^{-1} (which is absorbed by the $O(h^\infty)$ error). Moreover, by construction all $a_{k,\pm}$ are compactly supported, since it solves a transport equation with compactly supported initial data. Thus, we have $(\partial_t^2 - L)u \in C_c^\infty([0,T] \times \mathbb{R}^n)$. Finally, at $t = 0$ we have

$$\begin{aligned} u(0,x) &= (2\pi)^{-n} \sum_{k=0,1,\sigma=\pm} \int_{\mathbb{S}^{n-1}} \int_0^\infty e^{ix\cdot\omega/h} a_{k,\sigma}(0,x;\omega,h)\rho(h)\hat{f}_k(\omega/h)\, h^{-n-1-j}\, dh\, d\mathbb{S}^{n-1}(\omega) \\ &= (2\pi)^{-n} \int_{\mathbb{S}^{n-1}} \int_0^\infty e^{ix\cdot\omega/h}\chi(x)\rho(h)\hat{f}_0(\omega/h)h^{-n-1}\, dh\, d\mathbb{S}^{n-1}(\omega) \\ &= \chi(x)(2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi}\rho(|\xi|^{-1})\hat{f}_0(\xi)\, d\xi. \end{aligned}$$

The integral returns $f_0(x)$ up to an error whose Fourier transform is supported in $|\xi|^{-1} \geq 1$, i.e. whose Fourier transform is compactly supported, and hence equals f_0 plus a smooth error. Multiplying this by $\chi(x)$ returns $f_0(x)$ (since χ is identically one on the support of f_0) plus χ times a smooth error, i.e. a $C_c^\infty(\mathbb{R}^n)$ error. Thus, $u(0,x) - f_0(x) \in C_c^\infty(\mathbb{R}^n)$. Similarly, we have $\partial_t u(0,x) - f_1(x) \in C_c^\infty(\mathbb{R}^n)$. □

Remark 23. A more polished (though characteristically terse) treatment of this is given in Grigis and Sjöstrand's *Microlocal Analysis for Differential Operators* [GS94], Chapter 7.

15.5. **Finite Speed of Propagation.** We end the variable-coefficient discussion with the finite speed of propagation in the variable coefficient case. In the constant coefficient case, we saw that solutions "propagate at finite speed" (specifically speed 1), in that the value of a solution depends only on the values at a previous time at points which can be reached by a geodesic (i.e. straight line) at speed at most 1. The natural generalization one may guess for the variable coefficient case is to replace the above by a geodesic with respect to the metric g_{ij} induced by the dual coefficients g^{ij}, and this is indeed the case here[37]:

More concretely, fix $x_0 \in \mathbb{R}^n$, and let $q(x) = d_g(x, x_0)$ be the geodesic distance from x_0 to $x \in \mathbb{R}^n$. Let $B_r(x_0)$ be the geodesic ball of radius r around x_0, i.e. $\{x : d_g(x, x_0) < r\}$. One can check, as an exercise in the method of characteristics, that if the metric has no *conjugate points* in $B_r(x_0)$ (roughly speaking that a variation in an initial velocity yields a nontrivial variation in the geodesic endpoint), then q is smooth in $B_r(x_0)\backslash\{x_0\}$, and in fact it is the unique smooth solution to the equation

$$\sum_{j,k=1}^{n} g^{jk}(x)\partial_{x_j} q(x)\partial_{x_k} q(x) = 1 \text{ in } B_r(x_0)\backslash\{x_0\}, \quad \lim_{x\to x_0} q(x) = 0.$$

(Note that the assumption of no conjugate points is automatically satisfied if the sectional curvature is non-positive everywhere.)

Now, assume t_0 satisfies the assumption above. Let $K_t = B_{t_0-t}(x_0)$ and $K = \{(t, x) : 0 \le t \le t_0, x \in K_t\}$. We then have the analogue of the finite speed of propagation statement for the standard wave equation:

Theorem 15.6 (Finite speed of propagation)**.** *Suppose $u \in C^\infty([0, t_0] \times \mathbb{R}^n)$ solves $(\partial_t^2 - L)u$ in $[0, t_0] \times \mathbb{R}^n$, and suppose $u \equiv 0$ and $\partial_t u \equiv 0$ on K_0. Then $u \equiv 0$ in K.*

Proof sketch. The proof is similar to the standard case: we define a suitable "local energy"

$$E(t) = \frac{1}{2}\int_{K_t} |\partial_t u(t, x)|^2 + \sum_{j,k=1}^{n} g^{jk}(x)\partial_{x_j} u(t, x)\partial_{x_j} u(t, x)\, dx$$

and aim to show that $E'(t) \le CE(t)$ for some constant $C > 0$. See [Eva10] Section 7.2.4 Theorem 8 for details. □

[37]The material below is taken from [Eva10] Section 7.2.4

16. Lecture 16 (05/19): Introduction to Microlocal Analysis: Symbols and Pseudodifferential Operators

References: Grigis and Sjöstrand, *Microlocal Analysis for Differential Operators* [GS94], and Hörmander, *The Analysis of Linear Partial Differential Operators III* [Hör07], specifically Ch. 18.

Some of the arguments will also be adapted from expository lecture notes of Melrose [Mel] and Wunsch [Wun].

Convention: A capital D denotes derivatives with "a factor of i^{-1} built in", i.e. $D^\alpha = \left(\frac{1}{i}\partial\right)^\alpha = i^{-|\alpha|}\partial^\alpha$. Then with respect to (our conventions on) the Fourier transform we have $\widehat{D^\alpha u}(\xi) = \xi^\alpha \hat{u}(\xi)$.

16.1. **Motivation.** Suppose we have a *constant-coefficient* differential operator P. We can write P in the form

$$P = \sum_{|\alpha| \le m} a_\alpha D^\alpha, \quad a_\alpha \in \mathbb{C}$$

(where D^α is defined as above). We can ask: how can we invert such an operator P?

The answer is *not* by applying another differential operator: note that the composition of two differential operators is another differential operator, whose orders add. Since the orders are always nonnegative, we can never compose two differential orders of positive order to obtain a zeroth order differential operator, such as the identity. However, if there was a way to define a "negative-order" differential operator, then this may be possible–this is one motivation behind a *pseudo*differential operator, as a way of inverting a differential operator.

If u is sufficiently nice (say in $\mathcal{S}'(\mathbb{R}^n)$), and $Pu = f$, then taking the Fourier transform gives

$$p(\xi)\hat{u}(\xi) = \hat{f}(\xi), \quad p(\xi) = \sum_{|\alpha| \le m} a_\alpha \xi^\alpha.$$

That is, on the Fourier side, constant-coefficient differential operators turn into multiplication by a polynomial. In particular, if this polynomial $p(\xi)$ does not vanish for any $\xi \in \mathbb{R}^n$, then we can recover u (uniquely among tempered distributions) via the Fourier transform:

$$\hat{u}(\xi) = \frac{1}{p(\xi)}\hat{f}(\xi).$$

If in turn $f \in L^1(\mathbb{R}^n)$, we can rewrite the Fourier transforms as integrals to obtain

$$\begin{aligned} u(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{u}(\xi)\, d\xi &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \frac{1}{p(\xi)} \hat{f}(\xi)\, d\xi \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \frac{1}{p(\xi)} \left(\int_{\mathbb{R}^n} e^{-i\xi\cdot y} f(y)\, dy \right) d\xi \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} \frac{1}{p(\xi)} f(y)\, dy\, d\xi. \end{aligned}$$

In the case that P does not have constant coefficients, we can still write

$$P = \sum_{|\alpha| \le m} a_\alpha(x) D^\alpha,$$

where now the $a_\alpha(x)$ are functions and not constant numbers. Then we can still write

$$Pu(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} p(x,\xi) f(y)\, dy\, d\xi, \quad p(x,\xi) = \sum_{|\alpha|\le m} a_\alpha(x) \xi^\alpha.$$

One can guess that a way to recover u from $f = Pu$ is to do the same Fourier multiplier approach, except now with this multiplier $p(x,\xi)$ which depends on x. That is, we can guess:

$$u(x) \overset{?}{=} (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} \frac{1}{p(x,\xi)} f(y)\, dy\, d\xi.$$

Perhaps unsurprisingly this does not give us the correct answer, but we can ask how close this gets us. This motivates studying expressions of the form on the RHS above.

We first study a class of "symbols" (i.e. the multiplier $1/p(x,\xi)$ above) before using those symbols to create operators known as *pseudodifferential operators*.

16.2. **Symbol classes.**

Definition 16.1. Let $n, p \in \mathbb{N}$, $0 \le \rho \le 1$, $0 \le \delta < 1$, and $m \in \mathbb{R}$. The space of *symbols* $S^m_{\rho,\delta}(\mathbb{R}^p;\mathbb{R}^n)$ is the set of smooth functions $a : \mathbb{R}^p_z \times \mathbb{R}^n_\xi \to \mathbb{C}$ such that, for any multi-indices $\alpha \in \mathbb{N}^n$ and $\beta \in \mathbb{N}^p$, there exists a constant $C_{\alpha,\beta} > 0$ such that

$$|\partial_z^\beta \partial_\xi^\alpha a(z,\xi)| \le C_{\alpha,\beta}(1+|\xi|)^{m-\rho|\alpha|+\delta|\beta|}.$$

Typically either $p = n$ (in which case we write x instead of z) or $p = 2n$ (in which case we write $z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^n$). One should interpret $\mathbb{R}^p \times \mathbb{R}^n$ as a vector bundle, where $\mathbb{R}^p$ is the "base" space and $\mathbb{R}^n$ is the "fiber" of the bundle.

Note: a commonly used pair of parameters (ρ,δ) is $(\rho,\delta) = (1,0)$.

Example 16.2. In all of these cases, $p = n$.

- Suppose $a(x,\xi) = \sum_{|\alpha|\le m} a_\alpha(x)\xi^\alpha$ where a_α are smooth, and all derivatives of each a_α are uniformly bounded. Then $a \in S^m_{1,0}(\mathbb{R}^n;\mathbb{R}^n)$.
- Suppose a is smooth and $a(x,\xi) = \frac{P(\xi)}{Q(\xi)}$ when ξ is sufficiently large, where P and Q are polynomials of order m_1 and m_2, with $|Q(\xi)| \ge (1+|\xi|)^{m_2}$ for sufficiently large ξ. Then $a \in S^{m_1-m_2}_{1,0}(\mathbb{R}^n;\mathbb{R}^n)$.
- Suppose $a(x,\xi) = \log(1+|\xi|^2)$. Then $a \in \cap_{m>0} S^m_{1,0}(\mathbb{R}^n;\mathbb{R}^n)$.
- Suppose $a(t,x,\tau,\xi)$ is smooth and independent of (t,x), and for $|(\tau,\xi)|$ sufficiently large we have $a(\tau,\xi) = \frac{1}{|\xi|^2+i\tau}$. Then $a \in S^{-1}_{1/2,0}(\mathbb{R}^n;\mathbb{R}^n)$.
- Suppose $a(x,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^n;\mathbb{R}^n)$, and $\tilde{a}(y,\eta) = a(F(y), G(y)\eta)$, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is smooth with bounded derivatives of all orders, $G : \mathbb{R}^n \to \mathrm{Mat}_{n\times n}(\mathbb{R})$ is a matrix-valued smooth function with bounded derivatives of all orders, such that $\det G(y)$ is bounded from above and also bounded away from zero. Then $\tilde{a} \in S_{\rho,\max(\delta,1-\rho)}(\mathbb{R}^n;\mathbb{R}^n)$, and the parameter $\max(\delta, 1-\rho)$ cannot in general

be improved. (An example of such a pair (F, G) would be: F is an arbitrary diffeomorphism of $\mathbb{R}^n$ such that F and F^{-1} both have bounded derivatives, and $G(y) = (DF(y)^\top)^{-1}$, where in this context DF is the derivative matrix of $F : \mathbb{R}^n \to \mathbb{R}^n$. Note that $(y, \eta) \mapsto (F(y), G(y)\eta)$ is the symplectomorphism on $T^*\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$ corresponding to the diffeomorphism $x = F(y)$.)

- Suppose $a(x, \xi) = e^{ix_0 \cdot \xi}$ where $x_0 \in \mathbb{R}^n$. Then $a \in S^0_{0,0}(\mathbb{R}^n; \mathbb{R}^n)$.

Remark 24. If $\mathbb{R}^p_z = \mathbb{R}^{p_1}_x \times \mathbb{R}^{p_2}_y$, and $a(x, \xi) \in S^m(\mathbb{R}^{p_1}; \mathbb{R}^n)$, then $\tilde{a}(x, y, \xi) = a(x, \xi)$ is also in $S^m(\mathbb{R}^p; \mathbb{R}^n)$, since any derivatives in the additional variables just annihilate $\tilde{a}$ completely.

Proposition 16.3. *Each $S^m_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n)$ is closed under addition, and under pointwise multiplication we have*

$$S^{m_1}_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n) \cdot S^{m_2}_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n) \subset S^{m_1+m_2}_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n)$$

(in fact, it turns out to be an equality).

Proof sketch for the multiplication property. Given $a \in S^{m_1}_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n)$ and $b \in S^{m_2}_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n)$, we need to estimate derivatives on the product ab. By Leibniz rule, this turns into a sum of products of the form (some derivatives on a) times (some derivatives on b); multiplying the estimates on those terms gives the desired estimate. $\square$

We consider a "residual space" consisting of symbols decaying very quickly:

Definition 16.4. The space $S^{-\infty}(\mathbb{R}^p; \mathbb{R}^n)$ consists of smooth functions $a : \mathbb{R}^p \to \mathbb{R}^n$ with the property that, for all multi-indices α and β, and *all* $N \in \mathbb{R}$, there exists $C_{\alpha,\beta,N}$ such that

$$|\partial_z^\beta \partial_\xi^\alpha a(z, \xi)| \leq C_{\alpha,\beta,N}(1 + |\xi|)^{-N}$$

As a quick exercise:

$$S^{-\infty}(\mathbb{R}^p; \mathbb{R}^n) = \cap_{m \in \mathbb{R}} S^m_{\rho,\delta}(\mathbb{R}^p; \mathbb{R}^n)$$

for *any* (ρ, δ).

16.3. **Pseudodifferential Operators.** We now take $p = 2n$, and consider operators which are defined with respect to $a \in S^m_{\rho,\delta}(\mathbb{R}^{2n}; \mathbb{R}^n)$ as follows:

$$Au(x) = (2\pi)^{-n} \int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi} a(x, y, \xi) u(y)\, dy\, d\xi. \tag{14}$$

We can write that the Schwartz kernel of this operator is

$$K(x, y) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x, y, \xi)\, d\xi \tag{15}$$

interpreted as an oscillatory integral. Actually, all of this really means that we want to consider the distribution K satisfying

$$(K, \phi \otimes \psi)_{\mathbb{R}^{2n}} = (2\pi)^{-n} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi} a(x, y, \xi) \phi(x) \psi(y)\, dx\, dy \right) d\xi \tag{16}$$

for $\phi, \psi \in \mathcal{S}(\mathbb{R}^n)$.

Proposition 16.5. *Suppose $u \in \mathcal{S}(\mathbb{R}^n)$. Then the integral in the RHS of (14), interpreted as an iterated integral (first over $y \in \mathbb{R}^n$, then over $\xi \in \mathbb{R}^n$) converges for every x. Moreover, the resulting function $Au(x)$ belongs to $\mathcal{S}(\mathbb{R}^n)$.*

We denote the operator A obtained in (14) through the symbol a as $\mathrm{Op}(a)$, i.e. the operator corresponding to the symbol a. Thus,

$$\mathrm{Op}(a)u(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,y,\xi) u(y)\, dy \right) d\xi.$$

Proof. Since we assume, for fixed ξ, that a is uniformly bounded in (x, y), it follows that the inner integral

$$\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,y,\xi) u(y)\, dy$$

converges for every ξ, as $u \in \mathbb{S}(\mathbb{R}^n)$. The question is whether the value of this integral, considered as a function of ξ, is sufficiently controlled so that it is integrable. To show that this inner integral is integrable as a function of ξ, it suffices to show that $(1+|\xi|^2)^N \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,y,\xi) u(y)\, dy$ satisfies a uniform bound in ξ for sufficiently large N.

This in turn follows from integration by parts. We note that

$$(1+|\xi|^2) e^{i(x-y)\cdot\xi} = (1 - \xi \cdot D_y)(e^{i(x-y)\cdot\xi})$$

from which we have

$$\begin{aligned}
&(1+|\xi|^2)^N \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,y,\xi) u(y)\, dy \\
&= \int_{\mathbb{R}^{2n}} (1 - \xi \cdot D_y)^N (e^{i(x-y)\cdot\xi}) a(x,y,\xi) u(y)\, dy \\
&= \int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi} (1 + \xi \cdot D_y)^N (a(x,y,\xi) u(y))\, dy.
\end{aligned}$$

(the last line following by iterating integration by parts, noting that boundary terms at infinity vanish due to u being Schwartz). I now claim that

$$(1 + \xi \cdot D_y)^N (au) = \sum_{|\alpha| \le N} a_{\alpha,N} D^\alpha u \tag{17}$$

for some choice of symbols $a_{\alpha,N}$ satisfying $a_{\alpha,N} \in S^{m+N+\delta(N-|\alpha|)}_{\rho,\delta}(\mathbb{R}^{2n}; \mathbb{R}^n)$. This follows by induction and from the Leibniz rule: each derivative D_y either falls on u (which just returns another Schwartz function) or on a symbol; in that case it raises its order by δ; furthermore we can absorb the $\xi\cdot$ multiplier on the symbol and raise its order by 1 as well. It follows that

$$\left| \int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi} (1 + \xi \cdot D_y)^N (a(x,y,\xi) u(y))\, dy \right| \le C_N (1 + |\xi|)^{m+N+\delta N},$$

the exponent following since it is the highest order of any of the $a_{\alpha,N}$ in (17), and hence

$$\begin{aligned}\left|\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,y,\xi)u(y)\,dy\right| &= (1+|\xi|^2)^{-N}\left|\int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi}(1+\xi\cdot D_y)^N(a(x,y,\xi)u(y))\,dy\right| \\ &\le C_N(1+|\xi|)^{-2N}(1+|\xi|)^{m+N+\delta N} \\ &= C_N(1+|\xi|)^{m-(1-\delta)N}.\end{aligned}$$

If $\delta < 1$, then taking N sufficiently large gives that the interior integral is at most $(1+|\xi|)^{-n-\epsilon}$ and hence integrable. Thus, the outer integral defining $Au(x)$ converges absolutely, and furthermore by the Dominated Convergence Theorem we have that $Au \in C^0(\mathbb{R}^n)\cap L^\infty(\mathbb{R}^n)$.

It remains to show that in fact $Au \in \mathcal{S}(\mathbb{R}^n)$. To do so, we provisionally consider the space

$$\begin{aligned} S &= \mathrm{Op}\left(\cup_m S^m_{\rho,\delta}(\mathbb{R}^n)\right)\left(\mathcal{S}(\mathbb{R}^{2n};\mathbb{R}^n)\right) \\ &= \{v \,:\, v = \mathrm{Op}(a)u \text{ for some } u\in\mathcal{S}(\mathbb{R}^n) \text{ and some } a \in \cup_m S^m_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)\}.\end{aligned}$$

That is, we consider the space of *all functions* obtainable as $\mathrm{Op}(a)(u)$ for some symbol a and some Schwartz function u. The above paragraph shows that $S \subset C^0(\mathbb{R}^n)\cap L^\infty(\mathbb{R}^n)$. It suffices to show that

$$\partial_{x_j} S \subset S \text{ and } x_j S \subset S$$

to show that

$$S \subset \mathcal{S}(\mathbb{R}^n).$$

To show that S is closed under differentiation, we note that

$$\begin{aligned}&\partial_{x_j}\left((2\pi)^{-n}\int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi}a(x,y,\xi)u(y)\,dy\,d\xi\right) \\ &= (2\pi)^{-n}\int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi}(i\xi_j a + \partial_{x_j}a)(x,y,\xi)u(y)\,dy\,d\xi = \mathrm{Op}(\tilde a)u,\end{aligned}$$

where $\tilde a = i\xi_j a + \partial_{x_j} a \in S^{m+1}_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$ if $a \in S^m_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$. Thus, $\partial_{x_j}\mathrm{Op}(a)u \in S$. Similarly, for multiplication, we rewrite $x_j = (x_j - y_j) + y_j$ to obtain

$$x_j\mathrm{Op}(a)u = \mathrm{Op}((x_j - y_j)a)u + \mathrm{Op}(a)(y_j u(y)).$$

The latter term belongs to S since $y_j u(y) \in \mathcal{S}(\mathbb{R}^n)$ when $u \in \mathcal{S}(\mathbb{R}^n)$. By the former term, we mean the function whose value is the interated integral

$$(2\pi)^{-n}\int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi}(x_j - y_j)a(x,y,\xi)u(y)\,dy\,d\xi,$$

which *a priori* is not of our desired form, as $(x_j - y_j)a$ is unbounded in (x,y). Nonetheless, we can integrate by parts by noting that

$$(x_j - y_j)e^{i(x-y)\cdot\xi} = D_{\xi_j}(e^{i(x-y)\cdot\xi}),$$

so the above integral can be written as

$$\begin{aligned}(2\pi)^{-n}&\int_{\mathbb{R}^{2n}} D_{\xi_j}(e^{i(x-y)\cdot\xi})a(x,y,\xi)u(y)\,dy\,d\xi\\ &= (2\pi)^{-n}\int_{\mathbb{R}^{2n}} e^{i(x-y)\cdot\xi}(-D_{\xi_j}a)(x,y,\xi)u(y)\,dy\,d\xi = \mathrm{Op}(-D_{\xi_j}a)u.\end{aligned}$$

Thus, $x_j\mathrm{Op}(a)u = \mathrm{Op}(-D_{\xi_j}a)u + \mathrm{Op}(a)(y_ju) \in S$, as desired. ☐

Remark 25. The uniform integrability of the function defined by the inner integral in (14) allows one to prove, using Fubini's theorem, that the iterated integral definition coincides with the distributional definition given in (15) or (16).

Definition 16.6. Given $a(x,y,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$, the operator $A = \mathrm{Op}(a) : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ defined in (14) is called the *pseudodifferential operator quantized* by the symbol a (sometimes written ΨDO for short). The set of ΨDO quantized by $a(x,y,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$ is denoted $\Psi^m_{\rho,\delta}(\mathbb{R}^n)$.

Remark 26. If $(\rho,\delta) = (1,0)$, then often the subscript (ρ,δ) is dropped.

Example 16.7. Some examples of ΨDOs:

- If $a(x,y,\xi) = \sum_{|\alpha|\le m} a_\alpha(x)\xi^\alpha$, then $A = \sum_{|\alpha|\le m} a_\alpha(x)D^\alpha$.
- If $a(x,y,\xi) = \sum_{|\alpha|\le m} a_\alpha(y)\xi^\alpha$, then $A = \sum_{|\alpha|\le m} D^\alpha a_\alpha(x)$, i.e. $Au(x) = \sum_{|\alpha|\le m} D^\alpha(a_\alpha u)(x)$.
- If $a = a(\xi)$ is independent of x and y, then A is the Fourier multiplier operator corresponding to multiplier $a(\xi)$.

We now consider the (complex) formal adjoint of a ΨDO, i.e. the operator A^* satisfying $\int_{\mathbb{R}^n} Au(x)\overline{v(x)}\,dx = \int_{\mathbb{R}^n} u(x)\overline{A^*v(x)}\,dx$ for $u,v \in \mathcal{S}(\mathbb{R}^n)$. The Schwartz kernel satisfies[38]

$$K_{A^*}(x,y) = \overline{K_A(y,x)} = \overline{(2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(y-x)\cdot\xi}a(y,x,\xi)\,d\xi} = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi}\overline{a}(y,x,\xi)\,d\xi.$$

It follows that $\mathrm{Op}(a(x,y,\xi)) = \mathrm{Op}(\overline{a}(y,x,\xi))$. Thus

$$A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n) \implies A^* \in \Psi^m_{\rho,\delta}(\mathbb{R}^n).$$

Since A^* maps $\mathcal{S}(\mathbb{R}^n)$ to itself, we can then claim:

Proposition 16.8. *A ΨDO $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$, initially defined as an operator $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$, extends uniquely to an operator $\mathcal{S}'(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$.*

Finally, ΨDOs can be left-reduced, meaning the following: while we initially considered quantizing symbols whose base space is $\mathbb{R}^{2n}$, i.e. symbols $a(x,y,\xi)$ where we allow for dependence both in x and y, it turns out that requiring dependence on just n base variables is sufficient. In particular, we can choose a symbol independent of y (i.e. only depending on the "left" variables x) to give the same operator:

[38] To be rigorous/safe, these manipulations should be done distributionally.

Proposition 16.9. *Suppose $\rho > \delta$. Let $a(x,y,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$. Then there exists a unique $\tilde{a}(x,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^n;\mathbb{R}^n)$ such that $\sigma(a) = \sigma(\tilde{a})$.*

The idea of this proof is by considering a Taylor expansion of $a(x,y,\xi)$ along the diagonal $y = x$:

$$a(x,y,\xi) \sim \sum_\alpha \frac{\partial_y^\alpha a(x,x,\xi)}{\alpha!}(y-x)^\alpha.$$

Putting this into the integral, and recognizing that $(y-x)^\alpha e^{i(x-y)\cdot\xi} = (i\partial_\xi)^\alpha(e^{i(x-y)\cdot\xi})$, we thus have that the Schwartz kernel equals

$$\sum_\alpha (2\pi)^{-n} \int_{\mathbb{R}^n} (i\partial_\xi)^\alpha(e^{i(x-y)\cdot\xi}) \frac{\partial_y^\alpha a(x,x,\xi)}{\alpha!}\, d\xi.$$

Integration by parts gives

$$(2\pi)^{-n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} \sum_\alpha \frac{\partial_y^\alpha D_\xi^\alpha a(x,x,\xi)}{\alpha!}\, d\xi.$$

Thus, we should take

$$\tilde{a}(x,\xi) \sim \sum_\alpha \frac{\partial_y^\alpha D_\xi^\alpha a(x,x,\xi)}{\alpha!}.$$

Note that each of the terms in the sum belongs to $S^{m-(\rho-\delta)|\alpha|}$, so at least we are summing over "lower-order" terms.

The issue with this argument is that the sum above has no reason to converge anywhere (in fact, the Taylor expansion which started the proof is an "asymptotic expansion" and should not be interpreted as a convergence in the sense of series; cf. non-analytic smooth functions). So we would need a notion of asymptotic summation for symbols. Such a notion happens to exist in this case.

Next time: more on left-reduction/asymptotic summation, composition, principal symbol, ellipticity, elliptic regularity.

17. Lecture 17 (05/24): More on the pseudodifferential calculus

For today: always take $0 \leq \delta < \rho \leq 1$.

Remark 27. If $\rho' \leq \rho$ and $\delta' \geq \delta$, then it is easy to see that $S^m_{\rho,\delta}(\mathbb{R}^p;\mathbb{R}^n) \subset S^m_{\rho',\delta'}(\mathbb{R}^p;\mathbb{R}^n)$. In particular, $S^m_{1,0}$ is contained in $S^m_{\rho,\delta}$ for any (ρ,δ) satisfying $0 \leq \delta, \rho \leq 1$, so we can always multiply a symbol (of any parameter (ρ,δ)) by a symbol in $S^m_{1,0}$ (e.g. by polynomials in ξ) and get out another symbol with parameter (ρ,δ) (with the orders adding as expected).

17.1. **More on reduction, asymptotic summations.** Last time, we gave an argument that a ΨDO *a priori* quantized by a symbol $a(x,y,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$ can be quantized by a symbol depending only on n variables $\tilde{a}(x,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^n;\mathbb{R}^n)$. The idea is to use the Taylor expansion

$$a(x,y,\xi) \sim \sum_\alpha \frac{\partial^\alpha a(x,x,\xi)}{\alpha!}(y-x)^\alpha$$

(the sum is taken over all multi-indices α). More specifically, we have

$$a(x,y,\xi) = \sum_{|\alpha|\leq N} \frac{\partial_y^\alpha a(x,x,\xi)}{\alpha!}(y-x)^\alpha + \sum_{|\alpha|=N+1} R_\alpha a(x,y,\xi)(y-x)^\alpha$$

where

$$R_\alpha a(x,y,\xi) = \int_0^1 \frac{|\alpha|(1-t)^{|\alpha|-1}}{\alpha!}\partial_y^\alpha a(x,x+t(y-x),\xi) \in S^{m+\delta|\alpha|}_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n).$$

Noting as well that $(y-x)^\alpha e^{i(x-y)\cdot\xi} = (i\partial_\xi)^\alpha(e^{i(x-y)\cdot\xi})$, we have

$$\begin{aligned}(2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi}\frac{\partial_y^\alpha a(x,x,\xi)}{\alpha!}(y-x)^\alpha\,d\xi &= (2\pi)^{-n}\int_{\mathbb{R}^n}(i\partial_\xi)^\alpha(e^{i(x-y)\cdot\xi})\frac{\partial_y^\alpha a(x,x,\xi)}{\alpha!}\,d\xi \\ &= (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi}\frac{i^{-|\alpha|}\partial_y^\alpha\partial_\xi^\alpha a(x,x,\xi)}{\alpha!}\,d\xi;\end{aligned}$$

similarly

$$(2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi}R_\alpha a(x,y,\xi)(x-y)^\alpha\,d\xi = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi}D_\xi^\alpha R_\alpha a(x,y,\xi)\,d\xi,$$

with $D_\xi^\alpha R_\alpha a \in S^{m-(\rho-\delta)|\alpha|}_{\rho,\delta}(\mathbb{R}^{2n};\mathbb{R}^n)$. It follows that for $\tilde{a}_N(x,\xi) = \sum_{|\alpha|\leq N}\frac{i^{-|\alpha|}\partial_y^\alpha\partial_\xi^\alpha a(x,x,\xi)}{\alpha!}$ we have

$$\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}_N(x,\xi)) \in \Psi^{m-(\rho-\delta)(N+1)}_{\rho,\delta}(\mathbb{R}^n),$$

so with $\rho-\delta > 0$ we see that we can find a left-reduced symbol $\tilde{a}_N$ where the corresponding operator $\mathrm{Op}(\tilde{a}_N)$ agrees with $\mathrm{Op}(a)$ *up to a lower-order operator error* in $\Psi^{m-(\rho-\delta)(N+1)}_{\rho,\delta}(\mathbb{R}^n)$; by choosing N appropriately large we can make the error have arbitrarily small (i.e. negative) order.

We however would like a left-reduced symbol which straight up matches $\mathrm{Op}(a)$. To do this, we'd like to set

$$\tilde{a} \sim \sum_{\alpha} \frac{\partial_y^\alpha D_\xi^\alpha a(x,x,\xi)}{\alpha!},$$

but we'd need a way to make sense of the above sum.

Proposition 17.1. *Suppose $\{m_j\}$ is a decreasing sequence of real numbers, and $a_j \in S^{m_j}_{\rho,\delta}(\mathbb{R}^p;\mathbb{R}^n)$ is* any *sequence of symbols (of decreasing order m_j). Then there exists $a \in S^{m_0}_{\rho,\delta}(\mathbb{R}^p;\mathbb{R}^n)$ such that, for any $N \in \mathbb{R}$,*

$$a - \sum_{m_j > -N} a_j \in S^{-N}_{\rho,\delta}(\mathbb{R}^p;\mathbb{R}^n).$$

We then say that a is an ***asymptotic sum*** *of the symbols a_j.*

Note that any two asymptotic sums differ by a symbol in $S^{-\infty}(\mathbb{R}^p;\mathbb{R}^n)$. In addition, there are no requirements on the a_j (in particular the sum does not need to pointwise converge anywhere) for an asymptotic sum to exist, beyond that the symbols belong to a specified symbol class.

Thus, we can find $\tilde{a} \in S^m_{\rho,\delta}(\mathbb{R}^n;\mathbb{R}^n)$ such that $\tilde{a} \sim \sum_\alpha i^{-|\alpha|}\frac{\partial_y^\alpha \partial_\xi^\alpha a(x,x,\xi)}{\alpha!}$. In that case, we have

$$\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}) = (\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}_N)) - \mathrm{Op}(\tilde{a} - \tilde{a}_N) \in \Psi^{m-(\rho-\delta)(N+1)}_{\rho,\delta}(\mathbb{R}^n).$$

So we once again have $\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}) \in \Psi^{m-(\rho-\delta)(N+1)}_{\rho,\delta}(\mathbb{R}^n)$ for all N, except the LHS does not depend on N. Thus, we have $\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}) \in \cap_N \Psi^{-N}_{\rho,\delta}(\mathbb{R}^n)$. We would like a characterization of the latter space:

Proposition 17.2. *Let A be a ΨDO on $\mathbb{R}^n$. The following statements are equivalent:*

- *$A = Op(\tilde{a})$ for some $\tilde{a}(x,\xi) \in S^{-\infty}(\mathbb{R}^n;\mathbb{R}^n)$.*
- *$A = Op(a)$ for some $a(x,y,\xi) \in S^{-\infty}(\mathbb{R}^{2n};\mathbb{R}^n)$.*
- *$A \in \cap_m \Psi^m_{\rho,\delta}(\mathbb{R}^n)$.*
- *The Schwartz kernel $K(x,y)$ of A is a smooth function on $\mathbb{R}^{2n}$ which is "Schwartz off the diagonal": that is, for any multi-indices $\alpha \in \mathbb{N}^{2n}$ and $\beta \in \mathbb{N}^n$, there exists $C_{\alpha,\beta} > 0$ such that*

$$|(x-y)^\alpha \partial^\beta_{x,y} K(x,y)| \le C_{\alpha,\beta} \text{ for all } (x,y) \in \mathbb{R}^{2n}.$$

Definition 17.3. The space $\Psi^{-\infty}(\mathbb{R}^n)$ consists of operators which satisfy any (and hence all) of the equivalent conditions in Proposition 17.2.

Thus, it follows that $\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}) \in \Psi^{-\infty}(\mathbb{R}^n)$, so $\mathrm{Op}(a) - \mathrm{Op}(\tilde{a}) = \mathrm{Op}(a_\infty)$ for some $a_\infty(x,\xi) \in S^{-\infty}(\mathbb{R}^n;\mathbb{R}^n)$, and hence $\mathrm{Op}(a) = \mathrm{Op}(\tilde{a}(x,\xi) + a_\infty(x,\xi))$. It follows that $\tilde{a}(x,\xi) + a_\infty(x,\xi)$ is our desired left-reduced symbol; note that such a symbol satisfies the same asymptotic sum as $\tilde{a}$ since the addition of a_∞ does not affect our asymptotic sum.

Remark 28. A computation which helps with proving Proposition 17.2: if A is quantized by a left-reduced symbol $a(x,\xi)$, then we have

$$K(x,y) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,\xi)\, d\xi = \mathcal{F}^{-1}_{\xi\to z} a(x, x-y),$$

where $\mathcal{F}^{-1}_{\xi\to z} a(x,z)$ is the inverse Fourier transform of *a in the ξ variables only.* Thus we can recover $a(x,\xi)$ by taking the Fourier transform of $K(x,x-z)$:

$$a(x,\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot z} K(x,x-z)\, dz.$$

This in fact also shows the uniqueness of the left-reduced symbol.

From now on, we denote $S^m(\mathbb{R}^n;\mathbb{R}^n)$ simply by $S^m(\mathbb{R}^n)$, or even S^m if the context is clear; in other words we will by default consider symbols on $\mathbb{R}^n \times \mathbb{R}^n$.

Definition 17.4. Suppose $\rho > \delta$, and let $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$. The *left-reduced symbol* $\sigma_L(A)$ is the unique symbol $a(x,\xi) \in S^m_{\rho,\delta}(\mathbb{R}^n)$ satisfying $A = \mathrm{Op}(a)$. The *principal symbol* (of order m) $\sigma_m(A)$ of A is the equivalence class of $\sigma_L(A)$ in the quotient space $S^m_{\rho,\delta}(\mathbb{R}^n)/S^{m-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$.

In practice, we identify $\sigma_m(A)$ with one of its representatives; this is particularly the case if one of the representatives is homogeneous of degree m.

Example 17.5. Let $A = \sum_{j,k=1}^n g^{jk}(x)\partial_j\partial_k + \sum_{k=1}^n b^k(x)\partial_k + q(x)$. Then $A \in \Psi^2_{1,0}(\mathbb{R}^n)$, and

$$\sigma_L(A)(x,\xi) = -\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k + i\sum_{k=1}^n b^k(x)\partial_k + q(x),$$

and

$$\sigma_2(A)(x,\xi) = -\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k.$$

By the last line, we mean that there exists a representative in the equivalence class of $\sigma_L(A)$ which is homogeneous of degree 2 (in fact a polynomial), and $\sigma_L(A)$ equals the above homogeneous polynomial up to a difference in $S^1_{1,0}(\mathbb{R}^n)$.

Thus, to summarize our left-reduction result:

Proposition 17.6. *Suppose $a(x,y,\xi) \in S^m(\mathbb{R}^{2n};\mathbb{R}^n)$. Then there exists a unique $\tilde{a} \in S^m(\mathbb{R}^n)$ such that $Op(a) = Op(\tilde{a})$. Moreover,*

$$\tilde{a}(x,\xi) \sim \sum_\alpha \frac{i^{-|\alpha|}}{\alpha!} \partial_y^\alpha \partial_\xi^\alpha a(x,x,\xi),$$

so that in particular

$$\sigma_m(Op(a)) = a(x,x,\xi).$$

Corollary 17.7. *Let $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$, and let $a = \sigma_L(A) \in S^m_{\rho,\delta}(\mathbb{R}^n)$. Then*

$$\sigma_L(A^*) \sim \sum_\alpha \frac{i^{-|\alpha|}}{\alpha!} \partial_x^\alpha \partial_\xi^\alpha \overline{a}(x,\xi).$$

In particular, $\sigma_m(A^*)(x,\xi) = \overline{a(x,\xi)}$.

Proof. This follows since if A is left-quantized by $a(x,\xi)$, then A^* is quantized by the full symbol $(x,y,\xi) \mapsto \overline{a(y,\xi)}$. Applying Proposition 17.6 gives the desired asymptotic expansion. □

Corollary 17.8. *Let $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$. Then there exists a unique "right-reduced" symbol $a_R(y,\xi) \in S^m(\mathbb{R}^n)$ such that $A = Op(a_R(y,\xi))$. Moreover, if $a_L(x,\xi) = \sigma_L(A)$ is the left-reduced symbol of A, then*

$$a_R(y,\xi) \sim \sum_\alpha \frac{i^{|\alpha|}}{\alpha!} \partial_x^\alpha \partial_\xi^\alpha a(y,\xi).$$

In particular, $a_R(y,\xi) - a_L(y,\xi) \in S^{m-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$.

Proof. $a_R(y,\xi) = \overline{\sigma_L(A^*)}(y,\xi)$. □

Remark 29. One advantage of using left/right-reduced symbols is that they interact well with the Fourier transform. If $A = \mathrm{Op}(a_L(x,\xi))$, then we can in fact write

$$Au(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} a_L(x,\xi)\hat{u}(\xi)\, d\xi,$$

while if $A = \mathrm{Op}(a_R(y,\xi))$, then

$$Au(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(\int_{\mathbb{R}^n} e^{-iy\cdot\xi} a_R(y,\xi) u(y)\, dy \right) d\xi,$$

and noting that the inner integral does not depend on x, it follows that Au is just the inverse Fourier transform of the inner integral, i.e.

$$\widehat{Au}(\xi) = \int_{\mathbb{R}^n} e^{-iy\cdot\xi} a_R(y,\xi) u(y)\, dy.$$

17.2. **Composition.** The above work leads to:

Theorem 17.9. *Let $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ and $B \in \Psi^{m'}_{\rho,\delta}(\mathbb{R}^n)$. Then $AB \in \Psi^{m+m'}_{\rho,\delta}(\mathbb{R}^n)$. Moreover,*

$$\sigma_L(AB) \sim \sum_\alpha \frac{i^{-|\alpha|}}{\alpha!} \partial_\xi^\alpha \sigma_L(A)(x,\xi) \partial_x^\alpha \sigma_L(B)(x,\xi).$$

In particular,

$$\sigma_{m+m'}(AB) = \sigma_m(A) \cdot \sigma_{m'}(B).$$

Proof. Consider the left-reduced symbol $a_L = \sigma_L(A)$ of A and the right-reduced symbol $b_R = \sigma_R(B)$. Then

$$\begin{aligned} ABu(x) &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} a(x,\xi) \widehat{Bu}(\xi)\, d\xi \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} a(x,\xi) \left(\int_{\mathbb{R}^n} e^{-iy\cdot\xi} b_R(y,\xi) u(y)\, dy \right) d\xi \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n\times\mathbb{R}^n} e^{i(x-y)\cdot\xi} a_L(x,\xi) b_R(y,\xi) u(y)\, dy\, d\xi. \end{aligned}$$

Thus, AB is the ΨDO quantized by the full symbol $(x,y,\xi) \mapsto a_L(x,\xi)b_R(y,\xi)$.

The asymptotic sum for $\sigma_L(AB)$ follows by applying the left and right-reduction formulas for full symbols in Proposition 17.6 and Corollary 17.8. We check that the formula holds up to the $|\alpha| = 1$ terms, i.e. we check the formula holds modulo $S^{m+m'-2(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$; see [Mel] for a full computation.

For $\alpha = 0$, we just evaluate $a_L(x,\xi)b_R(y,\xi)$ along the diagonal. This is

$$\begin{aligned} a_L(x,\xi)b_R(x,\xi) &= a_L(x,\xi) \left(b_L(x,\xi) + \sum_{|\beta|=1} i\partial_x^\beta \partial_\xi^\beta b_L(x,\xi) + S^{m'-2(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n) \right) \\ &= a_L(x,\xi)b_L(x,\xi) - \sum_{|\beta|=1} i^{-1} a_L(x,\xi) \partial_x^\beta \partial_\xi^\beta b_L(x,\xi) + S^{m+m'-2(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n). \end{aligned}$$

For $|\alpha| = 1$ we have the term

$$\begin{aligned} & i^{-1} \partial_y^\alpha \partial_\xi^\alpha |_{y=x} \left(a_L(x,\xi) b_R(y,\xi) \right) \\ &= i^{-1} \left(a_L(x,\xi) \partial_x^\alpha \partial_\xi^\alpha b_R(x,\xi) + \partial_\xi^\alpha a_L(x,\xi) \partial_x^\alpha b_R(x,\xi) \right) \\ &= i^{-1} \left(a_L(x,\xi) \partial_x^\alpha \partial_\xi^\alpha b_L(x,\xi) + \partial_\xi^\alpha a_L(x,\xi) \partial_x^\alpha b_L(x,\xi) \right) + S^{m+m'-2(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n). \end{aligned}$$

It follows that

$$\begin{aligned} \sigma_L(AB) \mod S^{m+m'-2(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n) &= a_L(x,\xi)b_L(x,\xi) - \sum_{|\beta|=1} i^{-1} a_L(x,\xi) \partial_x^\beta \partial_\xi^\beta b_L(x,\xi) \\ &\quad + \sum_{|\alpha|=1} i^{-1} \left(a_L(x,\xi) \partial_x^\alpha \partial_\xi^\alpha b_L(x,\xi) + \partial_\xi^\alpha a_L(x,\xi) \partial_x^\alpha b_L(x,\xi) \right) \\ &= a_L(x,\xi)b_L(x,\xi) + \sum_{|\alpha|=1} i^{-1} \partial_\xi^\alpha a_L(x,\xi) \partial_x^\alpha b_L(x,\xi), \end{aligned}$$

as desired. □

Remark 30. Note that $\Psi^{-\infty}(\mathbb{R}^n)$ is a two-sided ideal: if $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ and $B \in \Psi^{-\infty}(\mathbb{R}^n)$, then $AB \in \Psi^{-\infty}(\mathbb{R}^n)$ and $BA \in \Psi^{-\infty}(\mathbb{R}^n)$.

Corollary 17.10. *Let* $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ *and* $B \in \Psi^{m'}_{\rho,\delta}(\mathbb{R}^n)$. *Then the commutator* $[A,B] = AB - BA \in \Psi^{m+m'-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$. *Moreover,*

$$\sigma_{m+m'-(\rho-\delta)}([A,B]) = i^{-1} \sum_{|\alpha|=1} \left(\partial_\xi^\alpha a \partial_x^\alpha b - \partial_x^\alpha a \partial_\xi^\alpha b\right), \quad a = \sigma_m(A), b = \sigma_{m'}(B).$$

Remark 31. The above formula can be rewritten as $i^{-1}\{\sigma_m(A), \sigma_{m'}(B)\}$, where

$$\{f,g\} = \sum_{j=1}^n \partial_{\xi_j} f \partial_{x_j} g - \partial_{x_j} f \partial_{\xi_j} g$$

is the *Poisson bracket* on $\mathbb{R}^{2n}$.

Next time: We will study ellipticity, which we begin to define in this lecture:

Definition 17.11. A symbol $a \in S^m_{\rho,\delta}(\mathbb{R}^n)$ is *elliptic* if there exist constants $c, C > 0$ such that

$$|a(x,\xi)| \geq c(1+|\xi|)^m \text{ for all } |\xi| > C.$$

An operator $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is *elliptic* if its principal[39] symbol is elliptic.

Example 17.12. For $A = \sum_{j,k=1}^n g^{jk}(x)\partial_j\partial_k + \sum_{k=1}^n b^k(x)\partial_k + q(x)$, we have

$$\sigma_2(A)(x,\xi) = -\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k.$$

Due to the homogeneity of (our choice of representative of) the principal symbol, we see that[40]

$$\sigma_2(A) \text{ is elliptic} \iff \sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k \neq 0 \text{ for all } \xi \neq 0.$$

If g^{jk} is in turn real-valued, this means that the matrix $(g^{jk}(x))_{j,k}$ is always positive definite or negative definite.

One main feature:

Lemma 17.13. *Suppose* $a \in S^m_{\rho,\delta}(\mathbb{R}^n)$ *is elliptic, and* $b(x,\xi)$ *is a function satisfying* $b(x,\xi) = \frac{1}{a(x,\xi)}$ *for sufficiently large* ξ. *Then* $b \in S^{-m}_{\rho,\delta}(\mathbb{R}^n)$.

Thus, we will consider operators quantized by such symbols b, and consider how they interact with an elliptic operator A.

[39]By this, we mean if *some* (and hence every) representative is elliptic; it is easy to see the notion of ellipticity does not change under lower-order perturbations.

[40]Technically one also needs to arrange the uniformity of the estimates, which can be an issue if the coefficients vary wildly over all of $\mathbb{R}^n$, but locally the nonvanishing of the quadratic form is enough to give locally uniform ellipticity estimates.

18. Lecture 18 (05/26)

18.1. **Elliptic operators and parametrices.** From the composition calculus, we can find "parametrices", or approximate inverses, for elliptic operators. Recall that a ΨDO $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is elliptic if its principal symbol $a = \sigma_m(A)$ is elliptic, meaning that

$$|a(x,\xi)| \geq c(1+|\xi|)^m \text{ for all } |\xi| > C$$

for some $c, C > 0$. In that case, $1/a$ (or more pedantically a smooth function agreeing with $1/a$ for large enough ξ) will also be a symbol, in fact in $S^{-m}_{\rho,\delta}(\mathbb{R}^n)$.

Proposition 18.1. *Suppose $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is elliptic. Then there exists $B \in \Psi^{-m}_{\rho,\delta}(\mathbb{R}^n)$ such that $AB - I, BA - I \in \Psi^{-\infty}(\mathbb{R}^n)$.*

Proof. We first guess $B_0 = \operatorname{Op}(b_0)$, where $b_0 \in S^{-m}_{\rho,\delta}(\mathbb{R}^n)$ agrees with $1/a$ for sufficiently large ξ. Then $A \circ B_0 \in \Psi^0_{\rho,\delta}(\mathbb{R}^n)$, with[41]

$$\sigma_0(A \circ B_0) = \sigma_m(A)\sigma_{-m}(B_0) = 1.$$

This means that $\sigma_0(A \circ B_0 - I) = 0$, i.e. $R_0 := A \circ B_0 - I \in \Psi^{-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$. Thus B_0 does not exactly invert A, but it inverts A up to an error that is a bit better than the identity.

Thus, we modify our guess to $B_0 + B_1$. We want to choose B_1 such that $A(B_0+B_1) = I$, i.e. that

$$0 = A(B_0 + B_1) - I = AB_1 + (AB_0 - I) = AB_1 + R_0.$$

Thus, we want B_1 to satisfy $AB_1 = -R_0$. This suggests we take $B_1 \in \Psi^{-m-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$, with

$$\sigma_m(A)\sigma_{-m-(\rho-\delta)}(B_1) = -\sigma_{-(\rho-\delta)}(R_0).$$

Thus, if $r_0 = \sigma_L(R_0)$, we let $b_1 = -b_0 r_0$ (i.e. heuristically $-r_0/a$), and we consider $A(B_0 + B_1)$ with $B_1 = \operatorname{Op}(b_1)$. Then similar arguments as before give that if $R_1 = A(B_0 + B_1) - I$, then $\sigma_{-(\rho-\delta)}(R_1) = 0$, so $R_1 \in \Psi^{-2(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$.

Iterating this argument, by induction we can find $B_j \in \Psi^{-m-j(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$ such that

$$R_j := A\left(\sum_{k=0}^{j} B_k\right) - I \in \Psi^{-(j+1)(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n).$$

The inductive step is by setting $\sigma_L(B_j) = -b_0\sigma_L(R_j)$.

Finally, if $b_j = \sigma_L(B_j)$, we find a symbol b which asymptotically sums the b_j, i.e.

$$b(x,\xi) \sim \sum_{j\geq 0} b_j(x,\xi).$$

Then one can show that

$$AB - I \in \Psi^{-\infty}(\mathbb{R}^n).$$

[41]Technically, ab_0 equals 1 only outside a sufficiently large ball in ξ, i.e. $ab_0 - 1$ is supported in a region of the form $|\xi| < C$. But that means that $ab_0 - 1 \in S^{-\infty}$, so we get equality after passing to the quotient.

Thus, we have found a *right parametrix* for A; pedantically we should write $AB_R - I \in \Psi^{-\infty}(\mathbb{R}^n)$ for some $B_R \in \Psi^{-m}_{\rho,\delta}(\mathbb{R}^n)$. Similar arguments yield the existence of a left parametrix $B_L A - I \in \Psi^{-\infty}(\mathbb{R}^n)$. *A priori* the two parametrices can be different, but it turns out they differ by a trivial error as well: indeed, note that

$$B_L + B_L(AB_R - I) = B_L A B_R = B_R + (B_L A - I)B_R.$$

Thus $B_L - B_R = (B_L A - I)B_R - B_L(AB_R - I) \in \Psi^{-\infty}(\mathbb{R}^n)$. In particular, the right parametrix B_R is also a left parametrix, since $B_R A - I = (B_R - B_L)A + (B_L A - I) \in \Psi^{-\infty}(\mathbb{R}^n)$; similarly the right parametrix is also a left parametrix. □

18.2. **Mapping properties.** To make the "approximate" part of the approximate inverse more useful, it is helpful to have mapping properties of ΨDOs. Note that if A is a Fourier multiplier, and the corresponding multiplier a is bounded, then by Parseval's theorem we have that A maps boundedly from L^2 to L^2. We show that this is the case for 0th order ΨDOs as well.

Theorem 18.2. *Let $A \in \Psi^0_{\rho,\delta}(\mathbb{R}^n)$. Then A maps boundedly from L^2 to L^2.*

Proof. We first show that $\Psi^m_{\rho,\delta}(\mathbb{R}^n)$ maps L^2 to L^2 when $m < -n$. In that case, the oscillatory integral defining the Schwartz kernel

$$K(x,y) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} a(x,\xi)\, d\xi$$

converges absolutely, so $K(x,y)$ is a bounded function on $\mathbb{R}^{2n}$. Moreover,

$$(x-y)^\alpha K(x,y) = (2\pi)^{-n}\int_{\mathbb{R}^n} D_\xi^\alpha (e^{i(x-y)\cdot\xi}) a(x,\xi)\, d\xi = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi}(-D_\xi)^\alpha a(x,\xi)\, d\xi,$$

with $(-D_\xi)^\alpha a(x,\xi) \in S^{m-|\alpha|(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n)$; hence $(x-y)^\alpha K(x,y)$ is uniformly bounded for every α. By taking enough multi-indices, we can conclude that $K(x,y)$ is uniformly integrable in x or y, i.e. that there exists $C > 0$ such that

$$\int_{\mathbb{R}^n} |K(x,y)|\, dy < C \text{ for all } x, \quad \int_{\mathbb{R}^n} |K(x,y)|\, dx < C \text{ for all } y. \tag{18}$$

We then use:

Lemma 18.3 (Schur's criterion)**.** *Suppose $K \in C^0(\mathbb{R}^{2n})$ satisfies* (18)*. Then the corresponding operator $Au(x) = \int_{\mathbb{R}^n} K(x,y)u(y)\, dy$ is bounded from L^2 to L^2.*

(Idea: estimate $|Au(x)|^2 \le \int_{\mathbb{R}^n} |K(x,y)||u(y)|^2\, dy \int_{\mathbb{R}^n} |K(x,y)|\, dy$ for each x.)

Thus, we see that $\Psi^m_{\rho,\delta}(\mathbb{R}^n)$ maps L^2 to L^2 when $m < -n$.

Next, we show that the same conclusion holds for all $m < 0$ (not just $m < -n$). This follows from the observation that

$$A \text{ is bounded } L^2 \to L^2 \iff A^*A \text{ is bounded } L^2 \to L^2.$$

This essentially follows from the observation

$$\|Au\|^2_{L^2} = \langle Au, Au\rangle_{L^2} = \langle A^*Au, u\rangle_{L^2}.$$

From this, we see that

$$\Psi^m_{\rho,\delta}(\mathbb{R}^n) \text{ maps } L^2 \text{ to } L^2 \iff \Psi^{2m}_{\rho,\delta}(\mathbb{R}^n) \text{ maps } L^2 \text{ to } L^2.$$

In particular, for any k we have that $\Psi^m_{\rho,\delta}(\mathbb{R}^n)$ maps L^2 to L^2 if and only if $\Psi^{2^k m}_{\rho,\delta}(\mathbb{R}^n)$ maps L^2 to L^2. For any $m < 0$, we have $2^k m < -n$ for k large enough, from which we get the result.

Finally, suppose $A \in \Psi^0_{\rho,\delta}(\mathbb{R}^n)$. Let $a \in \sigma_L(A)$, and let $C > 0$ satisfy $C > |a(x,\xi)| + 1$ for all (x,ξ). The heuristic argument goes as follows: we should be able to find $B \in \Psi^0_{\rho,\delta}(\mathbb{R}^n)$ such that

$$C\mathrm{Id} = A^*A + B^*B.$$

In that case, we have

$$C\|u\|_{L^2} = \langle (A^*A + B^*B)u, u\rangle = \|Au\|^2_{L^2} + \|Bu\|^2_{L^2} \geq \|Au\|^2_{L^2}.$$

In fact, for the above equation to hold, we should have

$$C^2 = |a(x,\xi)|^2 + |b(x,\xi)|^2 \implies b(x,\xi) = \sqrt{C^2 - |a(x,\xi)|^2}.$$

It turns out that b, as defined above, is indeed a symbol in $S^0_{\rho,\delta}(\mathbb{R}^n)$.

However, since the symbol of the composition is not exactly the product of the symbols, the best we can say is that

$$\sigma_0(A^*A + B^*B - C) = 0 \implies A^*A + B^*B - C = R \in \Psi^{-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n).$$

But this turns out to be okay, since the remainder R is still some negative-order ΨDO, and hence bounded from L^2 to L^2 by our discussion above. Thus

$$\|Au\|^2_{L^2} = \langle A^*Au, u\rangle_{L^2} = C\|u\|^2_{L^2} - \|Bu\|^2_{L^2} + \langle Ru, u\rangle_{L^2} \leq (C + \|R\|_{L^2 \to L^2})\|u\|_{L^2},$$

i.e. A is bounded as desired. □

Corollary 18.4. *$A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is bounded H^{s+m} to H^s for any $s \in \mathbb{R}$.*

Proof. Let Λ^s denote the Fourier multiplier operator with multiplier $(1 + |\xi|^2)^{s/2}$. Then, by definition,

$$\|u\|_{H^s(\mathbb{R}^n)} = \|\Lambda^s\|_{L^2(\mathbb{R}^n)}.$$

Moreover, $\Lambda^s \in \Psi^s_{\rho,\delta}(\mathbb{R}^n)$ for any (ρ,δ). It follows that $\Lambda^s A \Lambda^{-(s+m)} \in \Psi^0_{\rho,\delta}(\mathbb{R}^n)$ and is thus bounded L^2 to L^2, so

$$\begin{aligned}\|Au\|_{H^s(\mathbb{R}^n)} = \|\Lambda^s Au\|_{L^2(\mathbb{R}^n)} &= \|\Lambda^s A \Lambda^{-(s+m)}(\Lambda^{s+m}u)\|_{L^2(\mathbb{R}^n)} \\ &\leq C\|\Lambda^{s+m}u\|_{L^2(\mathbb{R}^n)} = C\|u\|_{H^{s+m}(\mathbb{R}^n)},\end{aligned}$$

as desired. □

A related mapping property:

Proposition 18.5. *Let $A \in \Psi^{-\infty}(\mathbb{R}^n)$ and $u \in \mathcal{S}'(\mathbb{R}^n)$. Then $Au \in C^\infty(\mathbb{R}^n)$.*

18.3. **Applications to Elliptic Regularity.**

Theorem 18.6. *Suppose $P \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is elliptic, and suppose $u \in \mathcal{S}'(\mathbb{R}^n)$ satisfies $Pu = 0$. Then $u \in C^\infty(\mathbb{R}^n)$.*

Proof. Since P is elliptic, there exists a parametrix $Q \in \Psi^{-m}_{\rho,\delta}$ such that $R = QP - I \in \Psi^{-\infty}(\mathbb{R}^n)$. Then

$$QPu = u + Ru.$$

But $Pu = 0$, so $u = -Ru \in C^\infty(\mathbb{R}^n)$ by Proposition 18.5. □

Theorem 18.7. *Suppose $P \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is elliptic, $Pu \in H^s(\mathbb{R}^n)$, and u has some initial Sobolev regularity, say $u \in H^{-N}(\mathbb{R}^n)$. Then $u \in H^{s+m}(\mathbb{R}^n)$, and for any $s' < s + m$ (in particular for very negative s') we have*

$$\|u\|_{H^{s+m}(\mathbb{R}^n)} \le C_{s'}(\|Pu\|_{H^s(\mathbb{R}^n)} + \|u\|_{H^{s'}(\mathbb{R}^n)}).$$

Proof. Similar to above, the existence of the parametrix $Q \in \Psi^{-m}_{\rho,\delta}$ allows us to conclude

$$u = Q(Pu) - Ru.$$

Since $Pu \in H^s$ and Q is order $-m$, i.e. it maps boundedly from H^s to H^{s+m}, we have $Q(Pu) \in H^{s+m}$. On the other hand, R maps *any* Sobolev space (in particular the initial H^{-N} space to which u belongs) to any other Sobolev space. Hence, $u \in H^{s+m}$. Furthermore,

$$\|u\|_{H^{s+m}} \le \|Q\|_{H^s \to H^{s+m}} \|Pu\|_{H^s} + \|R\|_{H^{s'} \to H^s} \|u\|_{H^{s'}},,$$

giving the desired estimate. □

As an application, if $u \in L^2(\mathbb{R}^n)$, P is a second order differential operator which is elliptic as an element of $\Psi^2_{1,0}(\mathbb{R}^n)$[42], and $Pu \in L^2(\mathbb{R}^n)$, then $u \in H^2(\mathbb{R}^n)$.

18.4. **Pseudolocality.** Differential operators are *local*, heuristically meaning that the behavior of the output on an open set depends only on the behavior of the input on that open set. More precisely, if P is a differential operator, then supp $(Pu) \subset$ supp u, so if $u = \tilde{u}$ on some set U, then $Pu = P\tilde{u}$ on U as well.

Pseudodifferential operators do **not** have this property. For example, if $a(x, \xi) = a(\xi)$ is independent of x, then $\mathrm{Op}(a)$ is a Fourier multiplier operator, which is equivalent to convolution with the inverse Fourier transform of $a(\xi)$. This operator is not local unless $\mathcal{F}^{-1}(a)$ is supported at the origin.

Nonetheless, pseudodifferential operators are *pseudolocal*, in that they do locally preserve smoothness property. Recall that a distribution u is *smooth at $x \in \mathbb{R}^n$* if there exists a neighborhood $V \ni x$ such that $u|_V$ is smooth (i.e. agrees with the

[42] This is equivalent to $P = \sum g^{jk}\partial_j\partial_k$ where $(g^{jk}(x))$ is *uniformly elliptic*, i.e.

$$\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k \ge \theta|\xi|^2, \quad \theta > 0 \text{ independent of } x.$$

restriction of a smooth function on V), and that the singular support sing supp u of u is the set of points where u is not smooth.

Proposition 18.8. *Let $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$. Then, for any $u \in \mathcal{S}'(\mathbb{R}^n)$, we have that sing supp $(Au) \subset$ sing supp (u).*

In particular, if u is smooth on U, then Au is smooth on U as well.

Proof. This is equivalent to noting that the Schwartz kernel K of A is smooth away from the diagonal... $\square$

Corollary 18.9. *Suppose $P \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$ is elliptic. Then, for any $u \in \mathcal{S}'(\mathbb{R}^n)$, we have that sing supp $(Pu) =$ sing supp u.*

Proof. Since P admits a parametrix $Q \in \Psi^{-m}_{\rho,\delta}(\mathbb{R}^n)$, we have

$$u = QPu - Ru$$

where $R \in \Psi^{-\infty}(\mathbb{R}^n)$. Then R is smoothing, i.e. $Ru \in C^\infty(\mathbb{R}^n)$, so sing supp $u =$ sing supp (QPu). But $Q \in \Psi^{-m}_{\rho,\delta}(\mathbb{R}^n)$, so sing supp $(QPu) \subset$ sing supp (Pu), as desired. $\square$

19. Lecture 19 (05/31)

19.1. **Coordinate Invariance.** Suppose we have a ΨDO

$$Au(y) = (2\pi)^{-n} \int_{\mathbb{R}^{2n}} e^{i(y-y')\cdot\eta} a(y,\eta)\, dy'\, d\eta$$

and a diffeomorphism $F : \mathbb{R}^n_x \to \mathbb{R}^n_y$. We now want to study the operator $\tilde{A} = F^* A (F^{-1})^*$. That is, given $v(x)$, if $v(x) = u(y)$ when $y = F(x)$ (i.e. $u(y) = v(F^{-1}(y))$), then

$$\tilde{A}v(x) = Au(y).$$

Proposition 19.1. *Let $A \in \Psi^m_{\rho,\delta}(\mathbb{R}^n)$, and suppose, for technical convenience, that*

$$\text{supp } K_A \subset \mathbb{R}^{2n}_{y,y'} \text{ is compact.}$$

Then $\tilde{A} = F^* A (F^{-1})^* \in \Psi^m_{\rho,\delta'}(\mathbb{R}^n)$, where $\delta' = \max(\delta, 1-\rho)$. Moreover, if $\rho < 1/2$, i.e. $\rho > \delta'$, then

$$\sigma_m(\tilde{A})(x,\xi) = \sigma_m(A)(F(x), (DF^\top)^{-1}(x)\xi).$$

Proof Sketch. We have

$$\begin{aligned}
\tilde{A}v(x) &= (2\pi)^{-n} \int_{\mathbb{R}^{2n}} e^{i(F(x)-y')\cdot\eta} a(F(x),\eta) u(F^{-1}(y'))\, dy'\, d\eta \\
&\overset{y'=F(x')}{=} (2\pi)^{-n} \int_{\mathbb{R}^{2n}} e^{i(F(x)-F(x'))\cdot\eta} a(F(x),\eta) u(x')\, |\det DF(x')|\, dx'\, d\eta.
\end{aligned}$$

We can write

$$F(x) - F(x') = G(x,x')(x-x')$$

for some smooth matrix-valued function G. Explicitly

$$G_{ij}(x,x') = \int_0^1 \partial_j F_i(tx + (1-t)x')\, dt,$$

so in particular $G(x,x) = DF(x)$. It follows that if we let $\xi = G(x,x')^T\eta$, so that

$$(F(x) - F(x'))\cdot\eta = (x-x')\cdot G(x,x')^T\eta = (x-x')\cdot\xi,$$

then

$$\tilde{A}v(x) = (2\pi)^{-n} \int_{\mathbb{R}^{2n}} e^{i(x-x')\cdot\xi} a(F(x), (G(x,x')^T)^{-1}\xi) u(x')\, \frac{|\det DF(x')|}{|\det G(x,x')|}\, dx'\, d\xi.$$

Thus, it suffices to show that

$$\tilde{a}(x,x',\xi) = a(F(x), (G(x,x')^\top)^{-1}\xi) \frac{|\det DF(x')|}{|\det G(x,x')|}$$

is a symbol in $S^m_{\rho,\delta'}(\mathbb{R}^{2n};\mathbb{R}^n)$. We leave this calculation as an exercise; we only comment that taking derivatives in x' on $\tilde{a}$ may end up taking derivatives in η on a and then subsequently multiplying by η, thus leading to an overall loss of $1-\rho$ orders of decay; hence the need to increase δ to δ'.

In particular, since $G(x,x) = DF(x)$, it follows that

$$\tilde{a}(x,x,\xi) = a(F(x), (DF(x)^\top)^{-1}\xi),$$

so if $\rho > \delta'$, then we get the desired principal symbol statement. $\square$

Thus, we can view the (principal) symbol a as naturally living on $T^*\mathbb{R}^n$. In general, the coordinate invariance result allows us to define pseudodifferential operators $\Psi^m_{\rho,1-\rho}(M)$ (with $\rho > 1/2$) on smooth manifolds M, essentially by considering operators whose restrictions to local charts look like ΨDOs on $\mathbb{R}^n$.

19.2. **Symbolic iterated regularity and radial compactification.** For now, we consider $(\rho,\delta) = (1,0)$.

The symbolic estimates for $S^m_{1,0}(\mathbb{R}^n)$ can be thought of as saying: there is an "initial" growth rate of $(1+|\xi|)^m$, with every derivative in x preserving this rate, and every derivative in ξ improving the rate of decay by 1. Thus there is an asymmetry in the roles of x and ξ. One way to have the two variables serve somewhat more similar roles is as follows: viewing the derivatives ∂_{x_j} and ∂_{ξ_j} as vector fields on $T^*\mathbb{R}^n$, instead of considering the vector fields ∂_{ξ_j}, we consider the vector fields

$$\xi_i \partial_{\xi_j}, \quad 1 \le i,j \le n.$$

Then, applying a vector field of the above form to a symbol will *keep* the rate of growth: the derivative lowers the rate by one, but the multiplication by ξ_i raises it back.

This is also true for a product V^α of vector fields of the above form as well. Here, V^α denotes a product $V^\alpha = V^{\alpha_1}V^{\alpha_2}\dots V^{\alpha_{|\alpha|}}$, where each V^{α_k} is either ∂_{x_j} or $\xi_i\partial_{\xi_j}$ for some $1 \le i,j \le n$. In fact, we have:

Lemma 19.2. *Let $a : T^*\mathbb{R}^n \to \mathbb{C}$ be smooth. Then $a \in S^m_{1,0}(\mathbb{R}^n)$ if and only if, for every product V^α of vector fields of the form ∂_{x_j} or $\xi_i\partial_{\xi_j}$, $1 \le i,j \le n$, there exists a constant C_α (depending on the choice of vector fields) such that*

$$|V^\alpha a(x,\xi)| \le C_\alpha (1+|\xi|)^m \text{ for all } (x,\xi) \in T^*\mathbb{R}^n.$$

Thus, we can say that a satisfies "iterative regularity" with respect to the set of vector fields given by

$$\partial_{x_j}, \xi_i\partial_{\xi_j}, \quad 1 \le i,j \le n.$$

There is a geometric way to think about this set of vector fields. We can "radially compactify" the fibers of $T^*\mathbb{R}^n$ to form a fiber bundle $\overline{T^*\mathbb{R}^n}$ where each fiber is a *compact manifold with boundary*, namely a closed half-sphere. For each fiber $\mathbb{R}^n$, this is done via the radial compactification map $\mathbb{R}^n \to \mathbb{S}^n_+$, where

$$\mathbb{S}^n_+ = \left\{ (z_1, \dots, z_{n+1}) \in \mathbb{R}^{n+1} \,:\, \sum_{i=1}^{n+1} z_i^2 = 1, z_{n+1} > 0 \right\}$$

defined by

$$\varphi(x) = \left(\frac{x}{\sqrt{1+|x|^2}}, \frac{1}{\sqrt{1+|x|^2}} \right).$$

Geometrically, this is done by mapping $x \in \mathbb{R}^n$ to $(x, 1) \in \mathbb{R}^{n+1}$, and then setting $z = \varphi(x)$ to be the point on $\mathbb{S}^n$ on the ray from the origin to $(x, 1)$. We then define

$$\overline{\mathbb{R}^n} := \overline{\mathbb{S}^n_+} = \left\{ (z_1, \dots, z_{n+1}) \in \mathbb{R}^{n+1} \ : \ \sum_{i=1}^{n+1} z_i^2 = 1, z_{n+1} \geq 0 \right\},$$

where we identify the interior of $\overline{\mathbb{R}^n}$ with $\mathbb{R}^n$ via the identification of $\mathbb{R}^n$ with the interior of $\overline{\mathbb{S}^n_+}$, i.e. $\mathbb{S}^n_+$. We define $\overline{T^*\mathbb{R}^n}$ by radially compactifying each fiber; in other words $\overline{T^*\mathbb{R}^n} = \mathbb{R}^n \times \overline{\mathbb{R}^n}$.

Remark 32. An alternative way to construct the radial compactification is as follows: we define the compactification map "near infinity" by the "inverse polar coordinates" map $\tilde{\varphi} : \mathbb{R}^n \backslash \{0\} \to (0, \infty) \times \mathbb{S}^{n-1}$,

$$\tilde{\varphi}(x) = \left(\frac{1}{|x|}, \frac{x}{|x|} \right).$$

In polar coordinates, this is the map $(r, \omega) \mapsto (r^{-1}, \omega)$. We then identify $[0, \infty) \times \mathbb{S}^{n-1}$ with $\overline{\mathbb{S}^n_+} \backslash \{(0, \dots, 0, 1)\}$ via a "collar neighborhood" map viewing $\mathbb{S}^{n-1} = \partial \overline{\mathbb{S}^n_+}$, before adding back in the origin which was originally left out, identifying it with the point $(0, \dots, 0, 1) \in \mathbb{S}^n_+$.

Note that if $\rho = 1/r = 1/|x|$, then

$$|dx|^2 = dr^2 + r^2 \, d\omega^2 = \frac{d\rho^2}{\rho^4} + \frac{d\omega^2}{\rho^2}.$$

Let $X = \overline{T^*\mathbb{R}^n}$, and consider the map $\iota : T^*\mathbb{R}^n \to X$, $\iota(x, \xi) = (x, \varphi(\xi))$ where φ is the radial compactification map.

Proposition 19.3. *Let $\mathcal{V}(X; \partial X)$ denote the collection of vector fields on X (with coefficients in $C^\infty(X)$, i.e. functions smooth up to the boundary) which are tangent to ∂X. Then the vector fields*

$$\iota_*(\partial_{x_j}), \iota_*(\xi_i \partial_{\xi_j}), \iota_*(\partial_{\xi_j}), \quad 1 \leq i, j \leq n,$$

initially defined as vector fields in the interior of X, extend to smooth vector fields in $\mathcal{V}(X; \partial X)$. Moreover, $\mathcal{V}(X; \partial X)$ is generated, over $C^\infty(X)$ by these vector fields. Here, ι_* denotes the pushforward of vector fields (colloquially the vector field in the interior viewed as a vector field on X).

Proof sketch. Viewing the fiber as a subset of $\mathbb{R}^{n+1}$ with coordinates $\eta_1, \dots, \eta_{n+1}$, we have

$$\partial_{\xi_j} \varphi_k = \begin{cases} -\frac{\xi_j \xi_k}{(1+|\xi|^2)^{3/2}} + \delta_{jk} \frac{1}{(1+|\xi|^2)^{1/2}} & 1 \leq k \leq n \\ -\frac{\xi_j}{(1+|\xi|^2)^{3/2}} & k = n+1 \end{cases},$$

so

$$\iota_*(\partial_{\xi_j}) = \sum_{k=1}^{n} (\delta_{jk} - \eta_j \eta_k) \eta_{n+1} \partial_{\eta_k} - \eta_j \eta_{n+1}^2 \partial_{\eta_{n+1}}$$

and, using that $(\iota^{-1})^*(\xi_i) = \frac{\eta_i}{\eta_{n+1}}$, we have

$$\iota_*(\xi_i \partial_{\xi_j}) = \sum_{k=1}^{n} \eta_i(\delta_{jk} - \eta_j\eta_k)\partial_{\eta_k} - \eta_i\eta_j\eta_{n+1}\partial_{\eta_{n+1}}.$$

To show these generate all tangent vector fields, we work with local coordinates. Note that $\mathbb{S}^n_+$ can be covered with open sets of the form $\{|\eta_i| > \epsilon\}$, $1 \le i \le n+1$ for sufficiently small $\epsilon > 0$ (explicitly $\epsilon < \frac{1}{\sqrt{n+1}}$ would work).

Suppose first we are in the open set $\{|\eta_i| > \epsilon\}$ for some $1 \le i \le n$; w.l.o.g. we can take $i = n$. Consider

$$z_i = \frac{\xi_i}{\xi_n}, 1 \le i \le n-1, z_n = \frac{1}{\xi_n}.$$

Then $(z_1, \dots, z_n)$ give local coordinates on $\{|\eta_n| > \epsilon\}$, since

$$z_i = \frac{1}{\eta_n}\eta_i, 1 \le i \le n-1, z_n = \frac{1}{\eta_n}\eta_{n+1}.$$

We then have $\xi_i = z_i/z_n$, $1 \le i \le n-1$¡ and $\xi_n = \frac{1}{z_n}$. It follows that

$$\partial_{z_i} = \xi_n\partial_{\xi_i}, \quad 1 \le i \le n-1, \quad z_n\partial_{z_n} = -\sum_{i=1}^{n} \xi_i\partial_{\xi_i}.$$

Thus, we have that ∂_{z_i}, $1 \le i \le n-1$, together with $z_n\partial_{z_n}$, are locally generated by vector fields of the form $\xi_i\partial_{\xi_j}$. Note that in this open set we have that z_n is a boundary defining function for ∂X, so vector fields tangent to ∂X are generated by ∂_{z_i}, $1 \le i \le n-1$, together with $z_n\partial_{z_n}$.

Otherwise, suppose we are in the open set $\{|\eta_{n+1}| > \epsilon\}$. In that case, the ξ coordinates $\{\xi_1, \dots, \xi_n\}$ themselves give local coordinates, so $\{\partial_{\xi_j}\}$ form a local basis for smooth vector fields on X (here the boundary ∂X plays no role). $\square$

Corollary 19.4. *Let $a \in T^*\mathbb{R}^n \to \mathbb{C}$, and for technical convenience assume that*

$$\overline{\Pi_x(\text{supp } a)} \text{ is compact},$$

where $\Pi_x : T^*\mathbb{R}^n \to \mathbb{R}^n$ is the projection onto the base. Then the following are equivalent:

- $a \in S^m_{1,0}(\mathbb{R}^n)$
- *For any product V^α of vector fields in $\mathcal{V}(X, \partial X)$, we have a uniform estimate of the form*
$$V^\alpha\left((1 + |\xi|)^{-m}a\right) \le C_\alpha.$$
- *a is smooth, and for any product V^α of vector fields which are homogeneous of degree 0, we have a uniform estimate of the form*
$$V^\alpha\left((1 + |\xi|)^{-m}a\right) \le C_\alpha.$$

Remark 33. The second condition above is sometimes written as $u \in \mathcal{A}(\overline{T^*\mathbb{R}^n})$, or that u is (L^∞-) conormal to the boundary ∂X of X.

Thus, we can view the symbolic requirements on a as iterated regularity statements with respect to the class of vector fields tangent to a particular compactification of the cotangent bundle.

19.3. **Parametrices for parabolic operators: the hypoelliptic calculus of Boutet de Monvel.** We now turn to a seemingly different topic. Last lecture we showed that elliptic operators admitted parametrices; this crucially used the fact that

$$a \in S^m_{\rho,\delta}(\mathbb{R}^n) \text{ elliptic } \implies \exists b \in S^{-m}_{\rho,\delta}(\mathbb{R}^n) \text{ such that } ab - 1 \in S^{-(\rho-\delta)}_{\rho,\delta}(\mathbb{R}^n).$$

A non-example of an elliptic operator is the heat operator $P = \partial_t - \Delta$ on $\mathbb{R}^{n+1}$. Its full symbol is $p(\tau,\xi) = i\tau + |\xi|^2 \in S^2_{1,0}(\mathbb{R}^n)$, and while $\frac{1-\chi(\tau,\xi)}{i\tau+|\xi|^2}$ is well-defined if χ is identically one for small (τ,ξ), this function does not belong to $S^{-2}_{1,0}(\mathbb{R}^n)$.

Nonetheless, we do have

$$q := \frac{1-\chi(\tau,\xi)}{i\tau + |\xi|^2} \in S^{-1}_{1/2,0}(\mathbb{R}^n),$$

and furthermore since p, q are independent of (t,x), i.e. the corresponding operators are Fourier multiplier operators, we have

$$\mathrm{Op}(q)P - I = P\mathrm{Op}(q) - I = \mathrm{Op}(qp-1) \in \Psi^{-\infty}(\mathbb{R}^n)$$

since $qp-1$ is compactly supported in (τ,ξ). Thus, P admits a parametrix $Q = \mathrm{Op}(q)$, which is a ΨDO, albeit not of the "ideal" order (at best we can say it's order -1, despite P itself being order 2). Nonetheless, it still has nice mapping properties; in particular it is pseudolocal. Thus we have:

Proposition 19.5. *Suppose $u \in \mathcal{S}'(\mathbb{R}^{n+1})$. Then*

$$\textit{sing supp } (u) = \textit{sing supp } ((\partial_t - \Delta)u).$$

In particular, if $f = (\partial_t - \Delta)u \in C^\infty(\mathbb{R}^{n+1})$, then $u \in C^\infty(\mathbb{R}^{n+1})$.

Thus, we can say that the heat operator $\partial_t - \Delta$ is *hypoelliptic*, in that like elliptic operators we have the property that solutions u do not have additional singularities beyond what's present in the output Pu.

There is, however, a more refined pseudodifferential calculus we can consider. This is a construction of Boutet de Monvel, in 1974 (see [Bou74]).

20. Lecture 20 (06/02)

20.1. **The Boutet de Monvel hypoelliptic calculus.** We consider the calculus first introduced by Boutet de Monvel in 1974 to provide parametrices for hypoelliptic operators. We are particularly interested in

$$P = \partial_t - \left(\sum_{j,k} a^{jk}(t,x)\partial_j\partial_k + \sum_k b^k(t,x)\partial_k + c(t,x) \right),$$

where we assume for convenience that all coefficients above are smooth and have bounded derivatives of all orders, and that the matrix $(a^{jk}(x))_{j,k}$ is uniformly elliptic with bound independent of x. Then

$$\sigma_L(P) = i\tau + \sum_{j,k} a^{jk}(x)\xi_j\xi_k - i\sum_k b^k(x)\xi_k + c(x).$$

The calculus is in general defined for any manifold M and any conic subset $\Sigma \subset T^*M$; we'll focus on the specific case $M = \mathbb{R}^{n+1}_{t,x}$ and

$$\Sigma = \{(t,x,\tau,\xi) \in T^*\mathbb{R}^{n+1} \,:\, \xi = 0\}.$$

In this case, let

$$d_\Sigma = \left(\frac{|\xi|^2}{|(\xi,\tau)|^2} + \frac{1}{|(\xi,\tau)|} \right)^{1/2}.$$

Note that d_Σ is bounded from above in the region $|(\tau,\xi)| > 1$, and also bounded from below by $(|\xi|/|(\xi,\tau)|)^2$; on the other hand at $\xi = 0$ we have that d_Σ decays as $|\tau|^{-1/2}$.

Definition 20.1. Let $m,k \in \mathbb{R}$. The space $S^{m,k}(M,\Sigma)$ (where $M = \mathbb{R}^{n+1}$ and $\Sigma = \{\xi = 0\}$) is the space of smooth functions $a : T^*\mathbb{R}^{n+1} \to \mathbb{C}$ such that we have estimates of the form

$$|W^\beta V^\alpha a(t,x,\tau,\xi)| \leq C_{\alpha,\beta}|(\tau,\xi)|^m d_\Sigma^{k-|\beta|} \quad \text{for all } |(\tau,\xi)| > 1$$

whenever V^α is a product of vector fields homogeneous of degree 0 *tangent to* Σ, and W^β is a product of vector fields homogeneous of degree 0 (not necessarily tangent to Σ).

For this lecture, we write the space as $S^{m,k}$ for short.

Thus, in our iterated regularity requirement, we distinguish between "tangent" vector fields tangent to Σ (which do not affect the growth/decay of the symbol) and "non-tangent" vector fields which affects the behavior when d_Σ is small, i.e. near Σ.

Remark 34. In our specific case, in testing the symbolic requirement it suffices to consider the tangent vector fields

$$\partial_{x_j}, \xi_i\partial_{\xi_j}, \xi_i\partial_\tau, \tau\partial_\tau, 1 \leq i \leq n$$

and the non-tangent vector fields

$$\tau\partial_{\xi_j}, 1 \leq j \leq n.$$

This since the tangent (resp. non-tangent) vector fields homogeneous of degree 0 can be generated (say over $S^0(\mathbb{R}^{n+1})$ away from the origin) by the above vector fields.

Example 20.2. Let P be our parabolic operator. Then

$$\sigma_L(P) = i\tau + \sum_{j,k} a^{jk}(x)\xi_j\xi_k - i\sum_k b^k(x)\xi_k + c(x) \in S^{2,2}(\mathbb{R}^{n+1}, \Sigma).$$

To see this, we note that $\sigma_L(P)$ is a sum of terms which are monomials in (τ, ξ) with coefficients in (t, x). If we consider just the vector fields in Remark 34, then each vector field applied to a monomial in (τ, ξ) with coefficients in (t, x) will return a term of the exact same form and of the same degree. Thus, the terms of degree at most 1 all remain monomials of degree at most 1 under arbitrary application of the relevant vector fields (regardless of whether they are tangent to Σ or not). Noting that

$$|(\xi, \tau)|^2 d_\Sigma^2 = |\xi|^2 + |(\xi, \tau)|,$$

we see that monomials of degree 1 are bounded by $|(\xi, \tau)|$, and hence always by $|(\xi, \tau)|^2 d_\Sigma^2$.

Thus, it remains to check the sum $\sum_{j,k} a^{jk}(x)\xi_j\xi_k$. We see that we always obtain a monomial of degree 2 upon applying one of the vector fields in the Remark **??**, and in fact applying a tangent vector field will return a monomial where the terms are both ξ's (no τs). The only issue is in applying the non-tangent vector fields $\tau\partial_{\xi_j}$. One application returns terms of the form $\tau\xi_k$, and one additional application returns terms of the form $|\tau|^2$. (Applying it a third time actually annihilates the term, but the fact that the behavior changes under application of one or two of these vector fields already indicates an interesting behvaior). We now note that

$$|\tau\xi_k| \le |(\xi, \tau)| \cdot |(\xi, \tau)| d_\Sigma = |(\xi, \tau)|^2 d_\Sigma, \quad |\tau|^2 \le |(\xi, \tau)|^2$$

to conclude that

$$|W^\beta V^\alpha(\sigma_L(P))(t, x, \xi, \tau)| \le C_{\alpha,\beta}|(\xi, \tau)|^2 d_\Sigma^{2-|\beta|}$$

whenever V and W are products of vector fields in Remark 34.

Proposition 20.3. *The symbol classes $S^{m,k}$ satisfy the following properties:*

- *Sums and products: $S^{m,k}$ is closed under addition, and*
$$S^{m,k} \cdot S^{m',k'} \subset S^{m+m',k+k'}.$$
- *Inclusion: we have $S^{m,k} \subset S^{m',k'}$ if and only if*
$$m \le m' \quad \text{and} \quad m - k/2 \le m' - k'/2.$$
This follow by noting that if $k' \ge k$, then
$$d_\Sigma^k = d_\Sigma^{k'} d_\Sigma^{-(k'-k)} \le d_\Sigma^{k'}|(\xi, \tau)|^{(k'-k)/2}.$$
- *We have $S^m_{1,0}(\mathbb{R}^{n+1}) \subset S^{m,0} \subset S^m_{1/2,0}(\mathbb{R}^{n+1})$. Consequently, we also have*
$$S^{m,k} \subset S^{m+(k_-)/2}_{1/2,0}(\mathbb{R}^{n+1}), \quad k_- = \max(0, -k).$$
- *Vector fields: we have $\partial_t, \partial_{x_j} : S^{m,k} \to S^{m,k}$, and $\partial_{\xi_k} : S^{m,k} \to S^{m-1,k-1}$. We also have $\partial_\tau : S^{m,k} \to S^{m-1,k}$.*

- *Ellipticity: suppose $a \in S^{m,k}$ also satisfied*
$$|a(t,x,\tau,\xi)| \geq c|(\xi,\tau)|^m d_\Sigma^k \text{ for all } |(\xi,\tau)| > 1.$$
Then, if b is smooth and agrees with $1/a$ for sufficiently large $|(\xi,\tau)|$, then $b \in S^{-m,-k}$. We will then say that a is an elliptic $S^{m,k}$ symbol.
- *Residual class: for any m and k, we have*
$$\cap_{j\geq 0} S^{m-j,k} = \cap_{j\geq 0} S^{m-j,k-j} = S^{-\infty}(\mathbb{R}^{n+1}).$$
(For the second space, note that $S^{m-j,k-j} \subset S^{m-j/2,k}$.)
- *Asymptotic sum: if $a_j \in S^{m-j,k-j}$, then there exists $a \in S^{m,k}$ such that*
$$a - \sum_{j=0}^{N-1} a_j \in S^{m-N,k-N}.$$

Example 20.4. Let $p = \sigma_L(P)$ where P is as above. Then, if $\chi(\tau,\xi) \in C_c^\infty(\mathbb{R}^{n+1})$ is identically one on a sufficiently large ball, we have
$$\frac{1-\chi(\tau,\xi)}{p(t,x,\tau,\xi)} \in S^{-2,-2} \subset S^{-2+(-2)_-/2}_{1/2,0}(\mathbb{R}^n) = S^{-1}_{1/2,0}(\mathbb{R}^n).$$

We now quantize such symbols. Since, for any m, k, we have that $S^{m,k} \subset S^{m'}_{1/2,0}(\mathbb{R}^{n+1})$ for some m', it follows that we can quantize symbols in $S^{m,k}$ to get operators that at the very least belong to $\Psi^{m'}_{1/2,0}(\mathbb{R}^n)$.

Definition 20.5. The space $\Psi^{m,k}(\mathbb{R}^{n+1},\Sigma)$ is the collection of operators $\mathrm{Op}(a)$ for $a \in S^{m,k}(\mathbb{R}^{n+1},\Sigma)$. (We can view them as mapping $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ or $\mathcal{S}'(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$.

We write this space of operators as $\Psi^{m,k}$ for short. Some properties about these operators:

Proposition 20.6. *We have:*

- *Composition: $\Psi^{m,k} \cdot \Psi^{m',k'} \subset \Psi^{m+m',k+k'}$, and*
$$\sigma_L(AB) \sim \sum_\alpha \frac{i^{-|\alpha|}}{\alpha!} \partial^\alpha_{\tau,\xi} \sigma_L(a)(t,x,\tau,\xi) \partial^\alpha_{t,x} \sigma_L(b)(t,x,\tau,\xi).$$
Note that each term in the asymptotic sum belongs to $S^{m+m'-|\alpha|,k+k'-|\alpha|}$.
- *Principal Symbol: As such, we can define a principal symbol $\sigma_{m,k} : \Psi^{m,k} \to S^{m,k}/S^{m-1,k-1}$, in which case*
$$\sigma_{m+m',k+k'}(AB) = \sigma_{m,k}(A)\sigma_{m',k'}(B).$$
- *Mapping: we have $\Psi^{0,0}$ maps boundedly from $L^2(\mathbb{R}^{n+1})$ to $L^2(\mathbb{R}^{n+1})$ (note that mapping properties for other $\Psi^{m,k}$ can be obtained via appropriate inclusions).*
- *Parametrix: If $\sigma_{m',k'}(A)$ is an elliptic $S^{m,k}$ symbol, then there exists $B \in \Psi^{-m,-k}$ such that $AB - I$ and $BA - I$ both belong to $\Psi^{-\infty}(\mathbb{R}^n)$.*

Indeed, the third item regarding parametrices follows from the composition and principal symbol results, using *exactly* the same arguments as in the standard case.

We apply the Boutet de Monvel hypoelliptic calculus to parabolic regularity results. Thus, let

$$P = \partial_t - \left(\sum_{j,k} a^{jk}(t,x)\partial_j\partial_k + \sum_k b^k(t,x)\partial_k + c(t,x) \right)$$

be our parabolic operator of interest. Note that our assumptions on the coefficients imply that P is an elliptic $\Psi^{2,2}$ operator. As such, we have:

Theorem 20.7. *Let $u \in \mathcal{S}'(\mathbb{R}^n)$, and let $f = Pu$. Then:*

- *sing supp u = sing supp f.*
- *If $f \in L^2(\mathbb{R}^{n+1})$, and u has some initial regularity $u \in H^{-N}(\mathbb{R}^{n+1})$, then u, $\partial_t u$, $\partial_{x_j} u$, and $\partial^2_{x_j x_k} u$ are all in $L^2(\mathbb{R}^{n+1})$, with L^2 norm bounded by a multiple of $\|f\|_{L^2(\mathbb{R}^{n+1})} + \|u\|_{H^{-N}(\mathbb{R}^{n+1})}$.*

Proof Sketch. All of this follows from the equation

$$u = Qf - Ru$$

where $Q \in \Psi^{-2,2}$ is a parametrix for P, and $R = QP - I \in \Psi^{-\infty}(\mathbb{R}^n)$. To obtain the L^2 bounds on the derivatives of u, we note that

$$\partial_t u = \partial_t Q f - \partial_t R u,$$

with $\partial_t \circ R \in \Psi^{-\infty}(\mathbb{R}^n)$, and $\partial_t \circ Q \in \Psi^{1,0} \circ \Psi^{-2,-2} \subset \Psi^{-1,-2} \subset \Psi^{0,0}$; in particular it maps boundedly from L^2 to L^2. A similar argument holds for ∂_{x_j}. Finally, for $\partial^2_{x_j x_k}$, we note that

$$\sigma_L(\partial^2_{x_j x_k}) = -\xi_j \xi_k \in S^{2,2}.$$

That is, not is the symbol of order 2, but the fact that it vanishes (quadratically) on Σ means that we can also capture its behavior with respect to d_Σ. Consequently, $\partial^2_{x_j x_k} \circ Q \subset \Psi^{2,2} \circ \Psi^{-2,-2} \subset \Psi^{0,0}$, so it also maps boundedly from L^2 to L^2. ☐

Remark 35. The property that $u, \partial_{x_j} u$, and $\partial^2_{x_j x_k} u$ all belong to L^2 can be rephrased as saying that $u \in L^2(\mathbb{R}_t; H^2(\mathbb{R}^n_x))$; furthermore in the above case we have

$$\|u\|_{L^2(\mathbb{R}_t; H^2(\mathbb{R}^n_x))} \leq C(\|f\|_{L^2(\mathbb{R}^{n+1})} + \|u\|_{H^{-N}(\mathbb{R}^{N+1})}).$$

Remark 36. The symbol calculus can also be defined by iterative regularity (with *no* change in growth/decay) with respect to a certain class of vector fields. We can no longer take the vector fields tangent to the boundary of radial compactification of $T^*\mathbb{R}^{n+1}$. However, if we take this compactification, and perform a *parabolic blow-up* of $\Sigma \cap \partial\overline{T^*\mathbb{R}^{n+1}}$, thus creating a manifold with corners, then it turns out that the symbolic requirements to be in $S^{0,0}$ are equivalent to requiring iterative uniform bounds with respect to vector fields tangent to the boundary of this blown-up space.

20.2. **Wavefront sets.**

Definition 20.8. Let $u \in \mathcal{D}'(\mathbb{R}^n)$, and let $(x_0, \xi_0) \in \mathbb{R}^{2n}$, with $\xi_0 \neq 0$. We say that u is *microlocally smooth* at (x_0, ξ_0) if there exists a cutoff $\chi \in C_c^\infty(\mathbb{R}^n)$ with $\chi(x_0) \neq 0$ and some $\epsilon > 0$ such that, for every $N > 0$, we have

$$|\widehat{\chi u}(\xi)| \leq C_N (1 + |\xi|)^{-N}$$

for all ξ which point "roughly in the same direction as ξ_0", i.e. more precisely for all $\xi \neq 0$ such that $\left|\frac{\xi}{|\xi|} - \frac{\xi_0}{|\xi_0|}\right| < \epsilon$. We let $WF(u)$ denote the collection of all (x, ξ) with $\xi \neq 0$ where u is *not* microlocally smooth.

The motivation is the following: recall that the Fourier transform interchanges regularity (in x) for decay (in ξ), and in fact if u is a compactly supported distribution, then u is smooth if and only if its Fourier transform decays faster than any power of ξ as $\xi \to \infty$. We employ a similar notion here, except we also want to distinguish the directions[43] where a singularity may occur.

Note by definition that $WF(u)$ is a conic set[44] which is closed in $\mathbb{R}^n \times (\mathbb{R}^n \backslash \{0\})$. As an exercise, one may show that

$$\Pi_x(WF(u)) = \text{sing supp } u,$$

where $\Pi_x : \mathbb{R}^n \times (\mathbb{R}^n \backslash \{0\}) \to \mathbb{R}^n$ is the projection onto the x-variables.

We proceed to define some quantities associated to operators. Here, we take $(\rho, \delta) = (1, 0)$ for convenience.

Definition 20.9. Let $(x_0, \xi_0) \in (\mathbb{R}^n) \times (\mathbb{R}^n \backslash \{0\})$. A *conical neighborhood* Γ of (x_0, ξ_0) is a set of the form

$$\Gamma = \left\{ (x, \xi) \in (\mathbb{R}^n) \times (\mathbb{R}^n \backslash \{0\}) \,:\, |x - x_0| < \epsilon, \left| \frac{\xi}{|\xi|} - \frac{\xi_0}{|\xi_0|} \right| < \epsilon \right\}.$$

Definition 20.10. Let $A \in \Psi^m(\mathbb{R}^n)$. We say that A is *microlocally trivial* at (x_0, ξ_0) if there exists a conical neighborhood Γ of (x_0, ξ_0) such that, for every $N > 0$, we have

$$|\sigma_L A(x, \xi)| \leq C_N (1 + |\xi|)^{-N} \text{ for all } (x, \xi) \in \Gamma.$$

We let the *essential support* $WF'(A)$ of A denote the set of points $(x, \xi) \in (\mathbb{R}^n) \times (\mathbb{R}^n \backslash \{0\})$ where A is *not* microlocally trivial.

Definition 20.11. Let $A \in \Psi^m(\mathbb{R}^n)$. We say that A is *elliptic at* (x_0, ξ_0) (of order m) if there exists a conical neighborhood Γ of (x_0, ξ_0) such that

$$|\sigma_L A(x, \xi)| \geq c(1 + |\xi|)^m \text{ for all } (x, \xi) \in \Gamma, |\xi| \geq C$$

for some $c, C > 0$. The set of $(x, \xi) \in (\mathbb{R}^n) \times (\mathbb{R}^n \backslash \{0\})$ where A is elliptic is denoted $\text{Ell}(A)$. The complement of $\text{ell}(A)$ (in $(\mathbb{R}^n) \times (\mathbb{R}^n \backslash \{0\})$) is denoted $\text{Char}(A)$, the characteristic set.

---

[43]More naturally the co-directions

[44]In the sense that $(x, \xi) \in WF(u) \implies (x, \lambda\xi) \in WF(u)$ for all $\lambda > 0$.

Example 20.12. Suppose $P = \sum_{|\alpha|\le m} a_\alpha(x) D^\alpha$. Then

$$WF'(P) = \cup_{|\alpha|\le m} \text{supp } a_\alpha$$

(this is essentially due to the fact that polynomials cannot vanish on a nonempty open set). On the other hand, ell(A) may be smaller, and we can in fact describe its complement by

$$\text{Char}(A) = p_m^{-1}(\{0\}),$$

where $p_m(x,\xi) : (\mathbb{R}^n) \times (\mathbb{R}^n \backslash \{0\}) \to \mathbb{C}$,

$$p_m(x,\xi) = \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha.$$

Proposition 20.13. *Let $A \in \Psi^m(\mathbb{R}^n)$ and $u \in \mathcal{S}'(\mathbb{R}^n)$. Then*

$$WF(Au) \subset WF'(A) \cap WF(u).$$

Furthermore,

$$WF(u) \subset WF(Au) \cup \mathit{Char}(A).$$

See [Mel] for proofs.

Corollary 20.14. *If A is elliptic, then $WF(u) = WF(Au)$.*

Corollary 20.15. *If $Au \in C^\infty(\mathbb{R}^n)$, then $WF(u) \subset \mathit{Char}(A)$.*

Using the above properties, we can give an equivalent definition of wavefront set which is often more microlocally practical:

Proposition 20.16. *Given $u \in \mathcal{S}'(\mathbb{R}^n)$, we have that (x_0,ξ_0) is **not** in $WF(u)$ if and only if there exists some ΨDO A which is elliptic at (x_0,ξ_0) such that $Au \in C^\infty(\mathbb{R}^n)$.*

Thus, if A is a differential operator, and we want to study solutions to $Au = 0$, we have some precision in describing how singularities of u can look; namely that the wavefront set must be part of the characteristic set of A, or roughly speaking the zero set of the principal symbol. We now mention one of the most important results in this topic, namely that of the propagation of singularities:

Theorem 20.17 (Hörmander's Propagation of Singularities)**.** *Let $P \in \Psi^m(\mathbb{R}^n)$ satisfy the property that the principal symbol $p = \sigma_m(P)$ is real-valued and homogeneous of degree m. Then, if $Pu \in C^\infty(\mathbb{R}^n)$, we have that $WF(u)$ is invariant under the Hamilton flow of p. That is, if (x,ξ) and (x',ξ') lie on a common Hamiltonian trajectory with respect to p, then $(x,\xi) \in WF(u) \iff (x',\xi') \in WF(u)$.*

Example 20.18. Let g be a Riemannian metric, and consider the wave equation $Pu = 0$, $P = \partial_t^2 - \Delta_g$ with respect to the Laplace-Beltrami operator Δ_g of g. Note that the leading order term of Δ_g is $\sum_{j,k} g^{jk}\partial_{x_j}\partial_{x_k}$, where $(g^{jk}) = (g_{jk})^{-1}$. Then

$$\sigma_2(P) = -\tau^2 + \sum_{j,k} g^{jk}(x)\xi_j\xi_k,$$

and Hamiltonian trajectories $(t(s), x(s), \tau(s), \xi(s))$ satisfy that the (t,x) components trace out a curve that can be described by $(t, x(t))$ where $x(t)$ is a geodesic of speed

1. It follows that, on physical space, singularities of u appear to propagate along unit-speed geodesics, with $WF(u)$, i.e. the "phase-space singularities", propagating along Hamiltonian trajectories.

Recall that the Hamilton vector field appears in the symbol for the commutator of operators: if $P \in \Psi^m$ and $A \in \Psi^{m'}$, with $p = \sigma_m(P)$ and $a = \sigma_{m'}(A)$, for $[P, A] := PA - AP$ we have

$$\sigma_{m+m'-1}([P, A]) = i^{-1}\{p, a\} = i^{-1} H_p a$$

where $\{,\}$ is the Poisson bracket, and $H_p = \sum_{j=1}^n \partial_{\xi_j} p \partial_{x_j} - \partial_{x_j} p \partial_{\xi_j}$ is the Hamilton vector field of p; note that $H_p a(x, \xi) = \frac{d}{ds}|_{s=0}(a(x(s), \xi(s)))$ where $(x(s), \xi(s))$ is (the) Hamilton trajectory with $(x(0), \xi(0)) = (x, \xi)$.

Below is a very rough sketch of the proof (see Section 4.2, Theorem 4.11 in [Wun] or Ch. 5 in [Mel] for more details):

- Assume for additional simplicity that $Pu = 0$, and that $P^* = P$ (this is certainly consist with the condition that $\sigma_m(P)$ is real-valued, since $\sigma_m(P^*) = \overline{\sigma_m(P)}$. Then we have
$$\langle i[P, A]u, u\rangle = 0$$
for any ΨDO A, since
$$\begin{aligned}\langle [P, A]u, u\rangle &= \langle PAu, u\rangle - \langle APu, u\rangle \\ &\overset{P^*=P}{=} \langle Au, Pu\rangle - \langle APu, u\rangle \overset{Pu=0}{=} \langle Au, 0\rangle - \langle A(0), u\rangle.\end{aligned}$$
- Suppose (x, ξ) is not in $WF(u)$, and suppose (x', ξ') is along the same Hamiltonian trajectory of (x, ξ). We want to show that (x', ξ') is also not in $WF(u)$. It turns out that the former condition implies that if $C \in \Psi^s$ satisfies $WF'(C)$ is supported sufficiently close to (x, ξ), then $Cu \in L^2$.
- We now claim that, for any s, that there exist $a \in S^{2s-m+1}$ and $b, c \in S^s$ such that
$$H_p a = b^2 - c^2,$$
where b is elliptic near (x', ξ') and c is supported sufficiently close to (x, ξ). The idea is that the above equation is a *first-order linear* equation in a, which we can solve with the method of characteristics; in this case the relevant characteristics are precisely the Hamiltonian trajectories. Thus, the above equation reduces essentially to an ODE.
- In that case, we have
$$i[P, A] = B^*B - C^*C + R, \quad R \in \Psi^{2s-1}.$$
If $R = 0$, we then have
$$0 = \langle i[P, A]u, u\rangle = \langle (B^*B - C^*C)u, u\rangle = \|Bu\|_{L^2}^2 - \|Cu\|_{L^2}^2,$$
i.e. $\|Bu\|_{L^2} = \|Cu\|_{L^2}$. Since $Cu \in L^2$ as c is supported close enough to (x, ξ), it follows that $Bu \in L^2$ as well. Hence, for any s, there is $B \in \Psi^s$,

elliptic at (x', ξ'), such that $Bu \in L^2$. This is morally the same as saying $(x', \xi') \notin WF(u)$: in fact, we can make a more precise definition

$$WF_s(u) = \{(x, \xi) : \nexists A \in \Psi^s \text{ elliptic at } (x, \xi) \text{ s.t. } Au \in L^2\},$$

in which case the above argument gives $(x, \xi) \notin WF_s(u) \implies (x', \xi') \notin WF_s(u)$. It turns out $WF(u) = \cap_s WF_s(u)$, giving the desired result.

- In general, $R \neq 0$. Hence we have to work iteratively, and assume u has some *a priori* regularity. For example, if we want to establish that $(x', \xi') \notin WF_s(u)$, we can assume that u a priori is (microlocally) $H^{s-1/2}$. Then $\langle Ru, u \rangle$ is finite as well, so the above argument still follows through.

Appendix A. Geodesic and Hamiltonian dynamics

Let $g^{jk}(x)$ $(j,k=1,\dots,n)$ be real-valued smooth functions on $\mathbb{R}^n$ such that $(g^{jk}(x))$ is a positive-definite symmetric matrix for each $x\in\mathbb{R}^n$, and let $g_{jk}(x)$ be functions such that, as matrices, we have $(g_{jk}(x))^{-1}=(g^{jk}(x))$. The goal here is to prove the relationship between geodesic flow and Hamiltonian flow of $\frac{1}{2}G$ where

$$G(x,\xi)=\sum_{j,k=1}^n g^{jk}(x)\xi_j\xi_k.$$

Following geometric conventions on sub/superscripts, we will write spatial coordinates with superscripts (x^j), coordinates of tangent vectors with superscripts (v^j), and coordinates of cotangent vectors with subscripts (ξ_j).

A convenient way to represent vectors and covectors is to represent vectors as column vectors, i.e. $\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ corresponds to the vector $\sum_j v^j\partial_{x^j}$, and to represent covectors as row vectors, i.e. $\begin{pmatrix}\xi_1 & \dots & \xi_n\end{pmatrix}$ corresponds to the covector $\sum_j \xi_j\,dx^j$. In that case, viewing $g=(g_{jk})$ and $G=(g^{jk})$ as matrices, we have $G=g^{-1}$, and

$$g(v,w)=v^\top gw,\quad G(\xi)=\xi G\xi^\top.$$

Given the metric g, we can define so-called *musical isomorphisms* $\flat: T\mathbb{R}^n\to T^*\mathbb{R}^n$ and $\sharp: T^*\mathbb{R}^n\to\mathbb{R}^n$, also known as *raising and lowering indices*, which give fiber-wise isomorphisms between the tangent and cotangent bundles via the metric. More concretely, $\flat$ is defined by

$$v^\flat=v^\top g$$

when we view the vectors and covectors as column/row vectors; in terms of coefficients we have

$$(v^\flat)_j=\sum_{k=1}^n g_{jk}v^k,$$

and

$$\xi^\sharp=G\xi^\top,\quad \text{i.e. } (\xi^\sharp)^j=\sum_{k=1}^n g^{jk}\xi_k.$$

(Note that $\flat$ and $\sharp$ are inverses of each other.) Moreover, the musical isomorphisms satisfy

$$v^\flat\cdot w=g(v,w),\quad g(\xi^\sharp,w)=\xi\cdot w.$$

(Here, the $\cdot$ represents the natural pairing of covectors with vectors.) In addition, we have

$$g(v,v)=G(\xi)\quad \text{if } \xi=v^\flat\quad (\text{equivalently } v=\xi^\sharp).$$

Finally, we note that matrix-valued functions satisfy a matrix form of the product rule

$$\partial_i(AB)=\partial_iAB+A\partial_iB$$

(note that here the order of multiplication matters), from which we obtain that

$$0 = \partial_i(AA^{-1}) = A\partial_i A^{-1} + (\partial_i A)A^{-1} \implies \partial_i A^{-1} = -A^{-1}\partial_i A A^{-1}.$$

Thus, $\partial_i g = -g\partial_i G g$, so that in particular if $\xi = v^\flat$ (or equivalently $v = \xi^\sharp$), then

$$v^\top \partial_i g v = -\xi \partial_i G \xi^\top.$$

We now recall that given a Riemannian metric g, there is an object called the *Levi-Civita* connection ∇ which measures how vector fields are changing along curves; in particular it gives a notion of when a vector field is "parallel" along a curve. This object takes in two vector fields X and Y and returns another vector field $\nabla_X Y$ which measures "how Y is changing along an integral curve of X". The connection satisfies

$$\nabla_{fX_1+X_2} Y = f\nabla_{X_1} Y + \nabla_{X_2} Y$$

and

$$\nabla_X fY_1 + Y_2 = X(f)Y_1 + f\nabla_X Y_1 + \nabla_X Y_2$$

for all smooth functions f and vector fields X_1, X_2, Y_1, and Y_2, where $X(f)$ is the function which gives the directional derivative of f in the direction of X (note that *all* connection satisfy the above properties). In particular, if we let Γ_{ij}^k be the functions satisfying

$$\nabla_{\partial_{x^i}} \partial_{x^j} = \sum_k \Gamma_{ij}^k \partial_{x^k},$$

then one can compute $\nabla_X Y$ solely in terms of the derivatives of the component functions of X and Y as well as these functions Γ_{ij}^k. The Γ_{ij}^k are called the *Christoffel symbols.*

Remark 37. For a fixed x, if γ is the integral curve of X through x, then the value of $\nabla_X Y$ at x depends only on the values of Y along γ, i.e. if $Y = \tilde{Y}$ along γ, then $\nabla_X Y = \nabla_X \tilde{Y}$ at x. It follows that, for any curve γ, it makes sense to discuss $\nabla_{\dot\gamma} Y$ where Y is a vector field only defined along γ, by (locally) extending Y arbitrarily away from γ. In coordinates, if $\dot\gamma(s) = \sum_i a^i(s)\partial_{x^i}$, and $Y(\gamma(s)) = \sum_j b^j(s)\partial_{x^j}$, then

$$\nabla_{\dot\gamma} Y = \sum_{j=1}^n \dot{b^j}\partial_{x^j} + \sum_{i,j,k=1}^n a^i b^j \Gamma_{ij}^k \partial_{x^k} = \sum_{k=1}^n \left(\dot{b^k} + \sum_{j,k=1}^n \Gamma_{ij}^k a^i b^j \right) \partial_{x^k}.$$

The Levi-Civita connection satisfies two additional conditions (which uniquely specifies this choice of connection), namely that it is torsion-free, meaning

$$\nabla_X Y - \nabla_Y X = [X, Y]$$

where $[,]$ is the Lie bracket of vector fields,

$$\left[\sum_i a^i \partial_{x^i}, \sum_j b^j \partial_{x^j}\right] = \sum_{i,j=1}^n \left(a^i \partial_{x^i} b^j \partial_{x^j} - b^j \partial_{x^j} a^i \partial_{x^i}\right) = \sum_{i=1}^n \left(\sum_{j=1}^n a^j \partial_{x^j} b^i - b^j \partial_{x^j} a^i\right) \partial_{x^i},$$

as well as a compatibility with the metric, meaning

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

for all vector fields X, Y, and Z. Note that the first condition gives $\Gamma^k_{ij} = \Gamma^k_{ji}$ by applying the condition to $X = \partial_i$ and $Y = \partial_j$ (noting that $[\partial_i, \partial_j] = 0$), while the second condition can be used to solve for Γ^k_{ij} in terms of the components of the metric g and its derivatives, by noting that plugging in $X = \partial_i$, $Y = \partial_j$, and $Z = \partial_k$ to give

$$\partial_i g_{jk} = \sum_{l=1}^n \Gamma^l_{ij} g_{kl} + \Gamma^l_{ik} g_{jl}$$

(recall that $g_{jk} = g(\partial_j, \partial_k)$). Changing the order of i, j, and k, and taking advantage of the symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$ derived above gives, after some cancellation,

$$\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij} = 2 \sum_{l=1}^n \Gamma^l_{ij} g_{kl}.$$

This can be used to solve for Γ^k_{ij} by noting that, in matrix notation, if we form a vector whose kth entry is as above, we have

$$(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})_k = 2g(\Gamma^l_{ij})_l \implies (\Gamma^l_{ij})_l = \frac{1}{2} G(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})_k,$$

i.e.

$$\Gamma^l_{ij} = \sum_{k=1}^n \frac{1}{2} g^{lk} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}).$$

Finally, a *geodesic* with respect to the Riemannian metric g is a curve $x(s)$ which satisfies the property that $\nabla_{\dot{x}(s)} \dot{x}(s) = 0$. In coordinates, if $v = \dot{x}$, then we have

$$\dot{v^k}(s) + \sum_{i,j=1}^n \Gamma^k_{ij}(x(s)) v^i(s) v^j(s) = 0, \quad , k = 1, \ldots, n.$$

We now aim to show:

Theorem A.1. *Let $G(x, \xi) = \sum_{i,j=1}^n g^{jk}(x) \xi_j \xi_k$, and let g be the corresponding Riemannian metric satisfying $(g_{jk}(x)) = (g^{jk}(x))^{-1}$. Suppose $(x(s), \xi(s))$ satisfies Hamilton's equations with respect to the Hamiltonian $\frac{1}{2}G$, i.e.*

$$\dot{x^i}(s) = \partial_{\xi_i} \left(\frac{1}{2} G\right)(x(s), \xi(s)), \quad \dot{\xi_i}(s) = -\partial_{x^i} \left(\frac{1}{2} G\right)(x(s), \xi(s)), \quad i = 1, \ldots, n.$$

Then $x(s)$ is a geodesic with respect to the Riemannian metric g. Moreover, viewing $\xi(s)$ as a covector in $T^*_{x(s)}\mathbb{R}^n$, i.e.

$$\xi(s) = \left(\sum_{i=1}^n \xi_i(s)\, dx^i\right)\Big|_{x(s)},$$

we have that $\xi(s) = (\dot{x}(s))^\flat$. Finally, if τ^2 is the constant value of G along the trajectory, then the geodesic has speed $|\tau|$ with respect to the metric g, i.e. that $g_{x(s)}(\dot{x}(s), \dot{x}(s)) = \tau^2$.

Proof. Note that

$$\partial_{\xi_j}\left(\frac{1}{2}G\right)(x,\xi)=\sum_{k=1}^{n}g^{jk}(x)\xi_k,$$

and hence $x(s)$ satisfies

$$\dot{x^j}(s)=\sum_{k=1}^{n}g^{jk}(x(s))\xi_k(s).$$

Letting $v=\dot{x}$, we see that $v(s)=\xi(s)^\sharp$, i.e. that $\xi(s)=v(s)^\flat$. Moreover, we have

$$\partial_{x^i}\left(\frac{1}{2}G\right)(x,\xi)=\frac{1}{2}\sum_{j,k=1}^{n}\partial_i g^{jk}(x)\xi_j\xi_k;$$

since $v(s)=\xi(s)^\sharp$, we see that

$$\partial_{x^i}\left(\frac{1}{2}G\right)(x(s),\xi(s))=-\frac{1}{2}\sum_{j,k=1}^{n}\partial_i g_{jk}(x(s))v^j(s)v^k(s).$$

As such, we have

$$\dot{\xi}_i(s)=-\partial_{x^i}\left(\frac{1}{2}G\right)(x(s),\xi(s))=\frac{1}{2}\sum_{j,k=1}^{n}\partial_i g_{jk}(x)v^j(s)v^k(s)$$

and hence

$$\begin{aligned}\dot{v^j}(s)&=\sum_{k=1}^{n}\frac{d}{ds}((g^{jk}(s)))\xi_k(s)+\sum_{k=1}^{n}g^{jk}(x(s))\dot{\xi}_k(s)\\&=\sum_{i,k=1}^{n}\partial_i g^{jk}(x(s))v^i(s)\xi_k(s)+\frac{1}{2}\sum_{i,k,l=1}^{n}g^{jk}(x(s))\partial_k g_{il}(x(s))v^i(s)v^l(s).\end{aligned}$$

Using that $\partial_i G=-G\partial_i gG$, or in terms of entries that $\partial_i g^{jk}(x)=-\sum_{l,m=1}^{n}g^{jl}(x)\partial_i g_{lm}(x)g^{mk}(x)$, and then using that $\sum_{k=1}^{n}g^{mk}(x)\xi_k=v^m$, we have

$$\sum_{i,k=1}^{n}\partial_i g^{jk}v^i\xi_k=-\sum_{i,k,l,m=1}^{n}g^{jl}\partial_i g_{lm}g^{mk}v^i\xi_k=-\sum_{i,l,m=1}^{n}g^{jl}\partial_i g_{lm}v^iv^m.$$

In the sum on the RHS, we relabel m to l and l to k, and note that the resulting sum is symmetric after interchanging i and l due to the symmetry of g_{kl}, and hence

$$\sum_{i,l,m=1}^{n}g^{jl}\partial_i g_{lm}v^iv^m=\frac{1}{2}\left(\sum_{i,k,l=1}^{n}g^{jk}(\partial_i g_{kl}+\partial_l g_{ki})v^iv^l\right).$$

Hence

$$\dot{v^j}(s)=-\frac{1}{2}\sum_{i,k,l=1}^{n}g^{jk}(x(s))\left(\partial_i g_{kl}(x(s))+\partial_l g_{ki}(x(s))-\partial_k g_{il}(x(s))\right)v^i(s)v^l(s).$$

Using

$$\Gamma^j_{il}(x) = \frac{1}{2}\sum_{k=1}^{n} g^{jk}(x)\,(\partial_i g_{kl}(x) + \partial_l g_{ki}(x) - \partial_k g_{il}(x)),$$

it follows that

$$\dot{v}^j(s) = -\sum_{i,l=1}^{n} \Gamma^j_{il}(x(s))v^i(s)v^j(s).$$

These are precisely the geodesic equations, as desired. $\square$

References

[Bou74] Louis Boutet de Monvel. Hypoelliptic operators with double characteristics and related pseudo-differential operators. *Comm. Pure Appl. Math.*, 27:585–639, 1974. `https://doi.org/10.1002/cpa.3160270502`.

[Cha84] Isaac Chavel. *Eigenvalues in Riemannian Geometry*. Elsevier Science, 1984.

[Eva10] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 2nd edition, 2010.

[GS94] Alain Grigis and Johannes Sjöstrand. *Microlocal Analysis for Differential Operators: An Introduction*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1994.

[Hör90] Lars Hörmander. *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis*. Springer-Verlag, Berlin, 2nd edition, 1990.

[Hör07] Lars Hörmander. *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators*. Springer-Verlag, Berlin, 2nd edition, 2007.

[Law10] Gregory F. Lawler. *Random Walk and the Heat Equation*. American Mathematical Society, 2010. Also available at `https://www.math.uchicago.edu/~lawler/reu.pdf`.

[Luk] Jonathan Luk. Introduction to nonlinear wave equations. Available at `https://web.stanford.edu/~jluk/NWnotes.pdf`.

[Mel] Richard Melrose. From microlocal to global analysis. Available at `https://math.mit.edu/~rbm/18-157-S14/iml.pdf`.

[Oh] Sung-Jin Oh. Lectures on wave equation. Available at `https://math.berkeley.edu/~sjoh/pdfs/linearWave.pdf`.

[Ste] Plamen Stefanov. Lecture notes on geometric optics. Available at `https://www.math.purdue.edu/~stefanop/publications/Geometric_Optics.pdf`.

[Wun] Jared Wunsch. Microlocal analysis and evolution equations: Lecture notes from 2008 cmi/eth summer school. Available at `https://arxiv.org/pdf/0812.3181.pdf`.