LECTURE NOTES FOR MATH 222A

SUNG-JIN OH

These are (evolving) lecture notes for the graduate PDE course (Math 222A) at UC Berkeley in Fall 2023. The principal references are: [Eva10], my previous lecture notes for Math 222A, and [Hö03].

1. Introduction to PDEs

At the most basic level, a Partial Differential Equation (PDE) is a *functional equation*, in the sense that its unknown is a function. What distinguishes a PDE from other functional equations, such as Ordinary Differential Equations (ODEs), is that a PDE involves *partial derivatives* ∂_i of the unknown function. So the unknown function in a PDE necessarily depends on *several variables*.

What makes PDEs interesting and useful is their *ubiquity* in Science and Mathematics. To give a glimpse into the rich world of PDEs, let us begin with a list of some important and interesting PDEs. Among the equations below, those that we will study in detail in this course are boxed.

1.1. A list of PDEs.

We start with the two most fundamental *PDEs for a single real or complex-valued function*, or in short, *scalar PDEs*.

- **The Laplace equation.** For $u : \mathbb{R}^d \to \mathbb{R}$ (or $\mathbb{C}$),

$$\boxed{\Delta u = 0, \qquad \text{where } \Delta = \sum_{i=1}^{d} \partial_i^2.}$$

  The differential operator Δ is called the *Laplacian*.
- **The wave equation.** For $u : \mathbb{R}^{1+d} \to \mathbb{R}$ (or $\mathbb{C}$),

$$\boxed{\Box u = 0, \qquad \text{where } \Box = -\partial_0^2 + \Delta.}$$

  Let us write $x^0 = t$, as the variable t will play the role of *time*. The differential operator $\Box$ is called the *d'Alembertian*.

The Laplace equation arise in the description of numerous "equilibrium states." For instance, it is satisfied by the temperature distribution function in equilibrium; it is also the equation satisfied by the electric potential in electrostatics in regions where there is no charge. The wave equation provides the usual model for wave propagation, such as vibrating string, drums, sound waves, light etc. Needless to say, this list of instances where these PDEs arise is (very much) non-exhaustive.

The Laplace and wave equations are important not only because of their ubiquity, but also because they are archetypical examples of major classes of PDEs, called the *elliptic* and *hyperbolic* PDEs. We will see more examples soon.

Date: March 19, 2024.

I would like to thank Jiahan Du, Mehdi Ouaki and Jianzhi Wang for pointing out errors in the earlier version of the notes.

Let us continue with our list of fundamental scalar PDEs.

- **The transport equation.** For $u : \mathbb{R}^d \to \mathbb{R}$ (or $\mathbb{C}$), and $b : \mathbb{R}^d \to \mathbb{R}^d$,
$$\boxed{\sum_j b^j \partial_j u = 0.}$$
This equation (and its variants) models phenomena that are transported by the vector field b.
- **The heat equation.** For $u : \mathbb{R}^{1+d} \to \mathbb{R}$ (or $\mathbb{C}$),
$$\boxed{(\partial_t - \Delta)u = 0.}$$
The heat equation is the usual model for heat flow in a homogeneous isotropic medium. It is prototypical of *parabolic* PDEs.
- **The (free) Schrödinger equation.** For $u : \mathbb{R}^{1+d} \to \mathbb{C}$ and $V : \mathbb{R}^{1+d} \to \mathbb{R}$,
$$\boxed{(i\partial_t - \Delta + V)u = 0.}$$
The Schödinger equation lies at the heart of non-relativistic quantum mechanics, and describes the free dynamics of a wave function. It is a prototypical *dispersive* PDE.

Although these two equations formally look similar, their solutions exhibit wildly different behaviors. Very roughly speaking, the heat equation has many similarities with the Laplace equation, whereas the Schrödinger equation is more similar to the wave equation.

Next, let us see some important examples of *PDEs for vector-valued functions*, or in short, *systems of PDEs*.

- **The Cauchy–Riemann equations.** For $u : \mathbb{R}^2 \to \mathbb{R}$, $v : \mathbb{R}^2 \to \mathbb{R}$,
$$\begin{cases} \partial_x u - \partial_y v = 0, \\ \partial_y u + \partial_x v = 0. \end{cases}$$
This is the central equation of complex analysis. Indeed, a pair of C^1 functions (u, v) satisfies the Cauchy–Riemann equation if and only if the complex-valued function $u + iv$ is holomorphic in $z = x + iy$. Also note that if (u, v) obeys the Cauchy–Riemann equation, then u and v each satisfy the Laplace equation.
- **The (vacuum) Maxwell equations.** For $\mathbf{E} : \mathbb{R}^{1+3} \to \mathbb{R}^3$ and $\mathbf{B} : \mathbb{R}^{1+3} \to \mathbb{R}^3$,
$$\begin{cases} -\partial_t \mathbf{E} + \nabla \times \mathbf{B} = 0, \\ \partial_t \mathbf{B} + \nabla \times \mathbf{E} = 0, \\ \nabla \cdot \mathbf{E} = 0, \\ \nabla \cdot \mathbf{B} = 0. \end{cases}$$
This is the main equation of electromagnetism and optics. Note that if $(\mathbf{E}, \mathbf{B})$ solves the Maxwell equations, then each component of $\mathbf{E}$ and $\mathbf{B}$ satisfy the wave equation, i.e., $\Box \mathbf{E}^j = 0$ and $\Box \mathbf{B}^j = 0$.

All equations mentioned so far have in common the property that they can be formally[1] written in the form
$$\mathcal{F}[u] = 0,$$

[1]Here, the word *formal* is used because, at the moment, $\mathcal{F}[u]$ makes sense for sufficiently regular functions. We will explore ways to extend this definition later in the course, when we discuss the *theory of distributions*.

where $\mathcal{F}$ is an operator that takes a function and returns a function, which is *linear* in the sense that

$$\mathcal{F}[\alpha u + \beta v] = \alpha \mathcal{F}[u] + \beta \mathcal{F}[v]$$

for any real (or even complex) numbers α, β and functions u, v. For this reason, they are called *linear* PDEs. Given a linear operator $\mathcal{F}[\cdot]$, the equation $\mathcal{F}[u] = 0$ is said to be *homogeneous* associated to $\mathcal{F}$, and any equation of the form $\mathcal{F}[u] = f$ is called *nonhomogeneous* (or *inhomogeneous*). The inhomogeneous Laplace equation,

$$\boxed{\Delta u = f}$$

has a special name; it is called the **Poisson equation**.

It turns out that many important and interesting PDEs are *nonlinear*. Let us see a few key examples from Geometry and Physics. To relate with the previously listed fundamental PDEs, the type of each nonlinear PDE (elliptic/hyperbolic/parabolic/dispersive/transport) will be indicated. However, we will refrain from actually defining what these types are, since it is one of those concepts that become counter-productive to make precise. It is sufficient to interpret the type as an indication of which of the fundamental PDEs the PDE at hand resembles the most.

We start with *nonlinear scalar PDEs.*

- **Burgers equation.** For $u : \mathbb{R}^{1+1} \to \mathbb{R}$ and $\nu \geq 0$,

$$\boxed{\partial_t u + u \partial_x u = \nu \partial_x^2 u,}$$

  This PDE arises from gas dynamics; it is *parabolic* if $\nu > 0$, and similar to the transport equation if $\nu = 0$. This is perhaps the simplest nonlinear PDE. But despite its simplicity, we will see that it already contains many interesting non-linear phenomena! Its study will form a useful guide for understanding nonlinear PDEs in general.
- **Hamilton–Jacobi equation.** For $u : \mathbb{R}^d \to \mathbb{R}$,

$$\boxed{\partial_t u + H(t, x, Du) = 0,}$$

  where $H : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called a *Hamiltonian.* This is the PDE that underlies both classical mechanics and geometric optics. It is similar to the transport equation.
- **Minimal surface equation.** For $u : \mathbb{R}^d \to \mathbb{R}$,

$$\Delta u - \sum_{i,j=1}^{d} \frac{\partial_i u \partial_j u}{1 + |Du|^2} \partial_i \partial_j u = 0.$$

  This is the PDE obeyed by the graph of a soap film, which minimizes the area under smooth, localized perturbations. It is of the *elliptic* type.
- **Korteweg–de Vries (KdV) equation.** For $u : \mathbb{R}^{1+1} \to \mathbb{R}$,

$$\partial_t u + \partial_x^3 u - 6u \partial_x u = 0.$$

  This PDE arises in the study of water waves. It is of the *dispersive* type.

Finally, let us turn to interesting examples of *nonlinear systems of PDEs.*

- **The compressible Euler equations.** For $\rho : \mathbb{R}^{1+d} \to \mathbb{R}$, $\mathbf{u} : \mathbb{R}^{1+d} \to \mathbb{R}^d$ and $E : \mathbb{R}^{1+d} \to \mathbb{R}$,

$$\begin{cases} \partial_t \rho + \sum_{j=1}^{d} \partial_j (\rho \mathbf{u}^j) = 0, \\ \partial_t (\rho \mathbf{u}^j) + \sum_{j=1}^{d} \partial_k (\rho \mathbf{u}^j \mathbf{u}^k + \delta^{jk} p) = 0, \\ \partial_t (\rho E) + \sum_{j=1}^{d} \partial_j (\rho \mathbf{u}^j E + p \mathbf{u}^j) = 0, \end{cases}$$

where δ^{jk} is the Kronecker delta. This is the basic equation of motion for gas (or more generally, compressible fluids) dynamics in the absence of viscosity. Here, ρ is the *gas density*, $\mathbf{u}^j$ is the *velocity*, and p is the pressure. The *specific total energy* E consists of

$$E = \frac{1}{2} |\mathbf{u}|^2 + e,$$

where $\frac{1}{2}|\mathbf{u}|^2$ is the (specific) kinetic energy and e is the specific internal energy. For a single gas, the specific internal energy is given as a function of ρ, p by physical considerations, i.e., $e = e(\rho, p)$. For instance, for an ideal gas,

$$e(\rho, p) = \frac{p}{\rho(\gamma - 1)}, \quad \text{where } \gamma > 1 \text{ is a constant.}$$

- **The Navier–Stokes equations.** For $\mathbf{u} : \mathbb{R}^{1+d} \to \mathbb{R}^d$,

$$\begin{cases} \partial_t \mathbf{u}^j + \sum_{j=1}^{d} \partial_k (\mathbf{u}^j \mathbf{u}^k + \delta^{jk} p) = \Delta \mathbf{u}^j, \\ \sum_{k=1}^{d} \partial_k \mathbf{u}^k = 0. \end{cases}$$

This is the basic equation for incompressible fluids (like water). It may be classified as *parabolic* PDE. The question whether every solution that is smooth at $t = 0$ stays smooth for all time is an (in)famous open problem.

The last two examples require a bit of differential geometry to state properly, but they are very amusing.

- **The Ricci flow.** For a Riemannian metric $\mathbf{g}$ on a smooth manifold,

$$\partial_t \mathbf{g}_{jk} = -2 \mathbf{Ric}_{jk}[\mathbf{g}]$$

where $\mathbf{Ric}_{jk}$ is the Ricci curvature associated with $\mathbf{g}_{jk}$; technically, $\mathbf{Ric}_{jk}$ is given in terms of $\mathbf{g}_{jk}$ by an expression of the form

$$\mathbf{Ric}_{jk}[\mathbf{g}] = (\mathbf{g}^{-1}) \partial^2 \mathbf{g} + (\mathbf{g}^{-1}) \partial \mathbf{g} (\mathbf{g}^{-1}) \partial \mathbf{g},$$

where $(\mathbf{g}^{-1})$ denotes the inverse (as a matrix) of $\mathbf{g}$. The Ricci flow is a *parabolic* PDE, which played a major role in the Hamilton–Perelman proof of the Poincaré conjecture.

- **The (vacuum) Einstein equations.** For a Lorentzian metric $\mathbf{g}$ (i.e., a symmetric 2-tensor that defines a quadratic form of signature $(-,+,\cdots+)$ on each tangent space) on a smooth manifold,

$$\mathbf{Ric}_{jk}[\mathbf{g}] - \frac{1}{2}R\mathbf{g}_{jk} = 0,$$

  where $\mathbf{Ric}_{jk}$ is the Ricci curvature associated with $\mathbf{g}$ and $R = \sum_{j,k=0}^{d} \mathbf{g}^{jk}\mathbf{Ric}_{jk}$. This is the central equation of General Relativity.

1.2. **Basic terminologies.** Let us cover some often-used terminologies in PDE.

- Multi-index notation: $\alpha = (\alpha_1, \dots, \alpha_n)$. For $y = (y^1, \dots, y^n) \in \mathbb{R}^n$, $y^\alpha = (y^1)^{\alpha_1} \cdots (y^n)^{\alpha_n}$. When $n = d$, $\partial^\alpha = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d}$ (or D^α may also be used). $|\alpha| = \alpha_1 + \dots + \alpha_n$ is the *order* of D^α.
- The *order* of a PDE (resp. differential operator $\mathcal{F}$) is the order of the highest order derivative that occurs in the PDE (resp. differential operator $\mathcal{F}$).
- The following classification of nonlinear PDEs is commonly used (from simple to intricate):
  - Semilinear: If the PDE is linear in the highest order derivative with coefficients that does not depend on u, i.e.,

$$\sum_{|\alpha|=k} a_\alpha(x) D^\alpha u + b(D^{k-1}u, \dots, Du, u, x) = 0.$$

  - Quasilinear: If the PDE is linear in the highest order derivative with coefficients that depends on at most $k-1$ derivatives of u, i.e.,

$$\sum_{|\alpha|=k} a_\alpha(D^{k-1}u, \dots, Du, u, x) D^\alpha u + b(D^{k-1}u, \dots, Du, u, x) = 0.$$

  - Fully nonlinear: If the PDE is nonlinear in the highest order derivative.

1.3. **Basic problems and concepts.** Now that we have seen some examples of important and interesting PDEs, let us discuss the basic problems for PDE and themes that often arise in their study.

When we solve a PDE, we want to find not just any solution, but a meaningful one. To achieve this, we prescribe data for the solution in various ways. Some important examples are:

- **Boundary value problems.** For a PDE posed on a domain U, data are typically prescribed on the boundary ∂U.
  - Dirichlet problem

$$\Delta u = f \text{ in } U, \quad u = g \text{ on } \partial U.$$

  - Neumann problem

$$\Delta u = f \text{ in } U, \quad \nu \cdot Du = g \text{ on } \partial U,$$

    where ν is the unit outward normal to ∂U. By the divergence theorem, we need to require that $\int_U f = \int_{\partial U} g$. Two solutions should be considered equivalent if they differ by a constant.
- **Initial value problem (Cauchy problem).** For evolutionary equation, the basic problem is to start with data at the initial time $t = 0$, and find a solution that agrees with the data at $t = 0$.

In the case of heat and Schrödinger equations, we only need to prescribe the initial data for u, since they are first-order in time.

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

$$(i\partial_t - \Delta + V)u = f \text{ in } \mathbb{R} \times \mathbb{R}^d, \quad u = g \text{ on } \{0\} \times \mathbb{R}^d.$$

In the case of the wave equation, which is second-order in time, we need to prescribe the initial data for both u and $\partial_t u$:

$$(-\partial_t^2 - \Delta)u = f \text{ in } \mathbb{R} \times \mathbb{R}^d, \quad u = g \text{ and } \partial_t u = h \text{ on } \{0\} \times \mathbb{R}^d.$$

The admissible boundary (or initial) data for a PDE is often dictated by its physical/geometric origin.

A boundary (or initial) value problem is said to be *well-posed* if the following three conditions hold:

- **Existence.** For each set of data, there exists a solution.
- **Uniqueness.** For each set of data, there exists at most one solution.
- **Continuous dependence.** The data-to-solution map is continuous.

To precisely formulate a wellposedness statement, we needs to specify the function spaces for data and solutions. When any of the above properties fail, the boundary (or initial) value problem is said to be *ill-posed*.

According to this terminology, the Fundamental Theorem of ODEs (often also referred to as the Picard–Lindelöf theorem) furnishes a general local(-in-time) wellposedness statement for ODEs. Let us recall this theorem as a reminder:

Theorem 1.1 (Fundamental Theorem of ODEs)**.** *Consider the initial value problem for an ODE for $x : \mathbb{R} \to \mathbb{R}^n$ of the form*

$$\begin{cases} \partial_t x = F(t, x(t)) \text{ for } t \in \mathbb{R}, \\ x(0) = x_0, \end{cases}$$

where $F : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n, (t, x) \mapsto F(t, x)$ is continuous in (t, x) and Lipschitz in x (i.e., $|F(t, x) - F(t, y)| < C(t)|x - y|$ for some $0 < C(t) < \infty$). Then there exists an interval $J \ni 0$ such that there exists a unique solution $x : J \to \mathbb{R}^n$ to the initial value problem on J.

Perhaps the most simple notion of "solving a boundary/initial value problem" is to find a closed formula that represents the solution in terms of the data (*representation formula*). However, such a formula is available only for very special PDEs. Even for the four fundamental linear scalar PDEs listed above (the Laplace, wave, heat and Schrödinger equations), we will be able to find closed representation formulas in special cases, and only with ad-hoc arguments.

So often, we ask the following questions for a boundary/initial value problem:

- **Wellposedness.** Is the BVP or IVP wellposed or illposed?
- **Regularity (vs. singularity).** If the data are regular, is the corresponding solution also regular? Failure of regularity (singularity) is very interesting; for PDEs from Science, singularity indicates the breakdown of the model at hand.
- **Asymptotics.** Can the solution be approximated by a simpler object (e.g., solution to a simpler PDE) as some parameter (e.g., time) tends to ∞?
- **Dynamics.** What are the equilibrium solutions (steady states)? Are these stable? Can a solution that is close to one equilibrium approach another equilibrium?

1.4. **The action principle (optional).** Many of the equations stated above come from the *action principle.*

- **Newton's equations.** We start with an ODE example. Let I be an interval, $\mathbf{x} : I \to \mathbb{R}^d$ and $V : \mathbb{R}^d \to \mathbb{R}$. Consider the following functional:

$$\mathcal{S}[\mathbf{x}] = \int_I \left[\frac{1}{2} |\dot{\mathbf{x}}(t)|^2 - V(\mathbf{x}(t)) \right] \mathrm{d}t$$

Here, $\mathbf{x}(t)$ may be thought of as the *position* of a particle at time t, V is the *potential energy*, $L(\mathbf{x})$ is the *Lagrangian*, and $\mathcal{S}$ is the *action.* The *action principle* in classical mechanics says:

> the path $\mathbf{x}(t)$ followed by the particle is that for which the action is minimized, or more precisely, is *stationary.*

Let us derive the condition satisfied by such a path $\mathbf{x}$. Consider deformation of $\mathbf{x}$ by a perturbation $s\boldsymbol{\varphi}$ that is sufficiently regular (i.e., C^1) and compactly supported in I, i.e., s is small and $\boldsymbol{\varphi} \in C^1_c(I; \mathbb{R}^d)$. Demanding that the action is stationary under any such perturbations can be phrased as:

$$0 = \frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[\mathbf{x} + s\boldsymbol{\varphi}] \Big|_{s=0}.$$

The consequence of the above assertion is as follows:

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[\mathbf{x} + s\boldsymbol{\varphi}] \Big|_{s=0} &= \frac{\mathrm{d}}{\mathrm{d}s} \int_I \left[\frac{1}{2} |\partial_t (\mathbf{x} + s\boldsymbol{\varphi})|^2 - V(\mathbf{x} + s\boldsymbol{\varphi}) \right] \mathrm{d}t \Big|_{s=0} \\ &= \int_I \partial_s \left[\frac{1}{2} |\partial_t (\mathbf{x} + s\boldsymbol{\varphi})|^2 - V(\mathbf{x} + s\boldsymbol{\varphi}) \right] \Big|_{s=0} \mathrm{d}t \\ &= \int_I \left[\dot{\mathbf{x}} \cdot \dot{\boldsymbol{\varphi}} - DV(\mathbf{x}) \cdot \boldsymbol{\varphi} \right] \mathrm{d}t \\ &= \int_I \left[-\ddot{\mathbf{x}} - DV(\mathbf{x}) \right] \cdot \boldsymbol{\varphi} \, \mathrm{d}t. \end{aligned}$$

On the second line, we changed the order of differentiation and integration, which is possible if, say, $\mathbf{x}$, V and their derivatives are bounded. On the last line, we performed an integration by parts, in order to move the derivative away from $\boldsymbol{\varphi}$. Since this identity holds for every $\boldsymbol{\varphi} \in C^1_c(I; \mathbb{R}^d)$, it follows that

$$-\ddot{\mathbf{x}} - DV(\mathbf{x}) = 0.$$

This equation is precisely *Newton's second law of motion* with a conservative force given by minus the gradient of the potential energy. In general, the relation for $\mathbf{x}$ that arises as a result of stationarity is called the *Euler–Lagrange equation.*

- **The Laplace equation.** Let U be a bounded domain in $\mathbb{R}^d$. Consider the following functional for a real-valued function u, which is called the *Dirichlet energy*:

$$\mathcal{S}[u] = \frac{1}{2} \int_U |Du|^2 \, \mathrm{d}x.$$

We look for the equation satisfied by a critical point u of the functional $\mathcal{S}[u]$ with respect to sufficiently regular (say, C^1) and compactly supported deformations (Euler–Lagrange equation). That is, for any φ that is C^1 and compactly supported in U,

$$0 = \frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[u + s\varphi] \Big|_{s=0} = \int_U Du \cdot D\varphi \, \mathrm{d}x$$

$$= - \int_U (\Delta u)\, \varphi \,\mathrm{d}x.$$

Since φ is arbitrary, we see that $\Delta u = 0$.

We note that the assumption that u is real-valued was made simply for convenience; an analogous discussion applies to complex-valued functions u.

- **The wave equation.** Consider the formal expression

$$\text{“}\mathcal{S}[u] =\text{”} \frac{1}{2} \int_{\mathbb{R}^{1+d}} \left(-(\partial_t u)^2 + \sum_{j=1}^{d} (\partial_j u)^2 \right) \mathrm{d}t \mathrm{d}x,$$

This integral does not make sense in general since $\mathbb{R}^{1+d}$ is noncompact (hence the disclaimer “formal”). However, for a sufficiently regular (say, C^1) and compactly supported function φ, the expression for the would-be first-order variation in the direction φ makes sense:

$$\text{“}\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[u + s\varphi]\Big|_{s=0} =\text{”} \int_{\mathbb{R}^{1+d}} \left(-\partial_t u \partial_t \varphi + \sum_{j=1}^{d} \partial_j u \partial_j \varphi \right) \mathrm{d}t \mathrm{d}x.$$

After an integration by parts,

$$\text{“}\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[u + s\varphi]\Big|_{s=0} =\text{”} - \int_{\mathbb{R}^{1+d}} \big((-\partial_t^2 + \Delta) u \big)\, \varphi \,\mathrm{d}t \mathrm{d}x,$$

so requiring that the “first-order variation of $\mathcal{S}[u]$ in the direction φ” vanishes for every sufficiently regular (say, C^1) and compactly supported φ leads to the wave equation. In this sense, the wave equation is the formal Euler–Lagrange equation for $\mathcal{S}[u]$ (equivalently, a solution to the wave equation is a formal critical point of $\mathcal{S}[u]$).

The Schödinger equation also turns out to have an action principle formulation, similar to the case of the wave equation. It is more difficult to see, but the Maxwell equations also arise from the action principle.

The heat equation does not come from an action principle formulation, but rather it arises as the *gradient flow* for the Dirichlet energy.

- **The heat equation.** Consider again the Dirichlet energy

$$\mathcal{S}[u] = \frac{1}{2} \int_U |Du|^2 \,\mathrm{d}x.$$

From the previous computation for the Laplace equation, we saw that

$$\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[u + s\varphi]\Big|_{s=0} = - \int (\Delta u) \varphi \,\mathrm{d}x.$$

The LHS can be interpreted as the directional derivative of the functional $\mathcal{S}[u]$ in the direction φ. In analogy with vector calculus, we may then interpret $-\Delta u$ as the *gradient* of the functional $\mathcal{S}[u]$ with respect to the inner product $(u, v) = \int uv \,\mathrm{d}x$. By the Schwarz inequality,

$$\left| \int -\Delta u \varphi \,\mathrm{d}x \right| \leq \|-\Delta u\|_{L^2} \|\varphi\|_{L^2},$$

and the equality is achieved if and only if φ is parallel to $-\Delta u$ (i.e., φ is of the form $c(-\Delta u)$ for some $c \in \mathbb{R}$). Hence, the gradient $-\Delta u$ (resp. the minus of the gradient Δu) of $\mathcal{S}[u]$ represents the direction of steepest ascent (resp. descent) of the functional $\mathcal{S}[u]$. The heat equation is obtained by equating $\partial_t u$ with the minus of the gradient, i.e.,

$$\partial_t u = \Delta u.$$

Many important *nonlinear* PDEs also arise from an action principle. The first example is:

- **The minimal surface equation.** Consider the functional

$$\mathcal{S}[u] = \int_U \sqrt{1 + |Du|^2} \, \mathrm{d}x,$$

which is nothing but the area of the graph of $u : U \to \mathbb{R}$. The Euler–Lagrange equation for $\mathcal{S}[u]$ is the minimal surface equation.

Indeed, for any φ that is sufficiently regular (say, C^1) and compactly supported in U, we have

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}s} \mathcal{S}[u + s\varphi] \Big|_{s=0} \\ &= \int_U \frac{1}{\sqrt{1 + |Du|^2}} Du \cdot D\varphi \, \mathrm{d}x \\ &= - \int_U \left(\sum_{j=1}^d \partial_j \left(\frac{1}{\sqrt{1 + |Du|^2}} \partial_j u \right) \right) \varphi \, \mathrm{d}x. \end{aligned}$$

Since this equation holds for any C^1 and compactly supported φ, we have

$$\sum_{j=1}^d \partial_j \left(\frac{1}{\sqrt{1 + |Du|^2}} \partial_j u \right) = 0.$$

After a simple algebra, we obtain the minimal surface equation that was written down before.

It turns out that the KdV and the vacuum Einstein equations arise from the action principle. It is a very deep fact, due to G. Perelman, that the Ricci flow may be interpreted as a gradient flow.

2. Nonlinear scalar first-order equations

Words on notation. This part of the course is largely based on Chapter 3 of Evans. However, we deviate from the notation in Evans (sorry!) in the following ways:

- Instead of n, we use d to denote the dimension of the underlying space;
- Instead of x_j, we use x^j to denote the j-th coordinate (this is the notation consistent with differential geometry).

2.1. Method of characteristics.

Consider a general nonlinear scalar first-order equation:

$$F(x, u, Du) = 0, \tag{2.1}$$

where $F : U \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. The *method of characteristics* is a way to reduce the PDE (2.1) to a set of ODEs, which is in general simpler.

2.1.1. *Inhomogeneous transport equation.* To motivate the method, let us consider a simple example of (2.1), namely the *inhomogeneous transport equation*:

$$\sum_j b^j(x)\partial_j u = f \text{ in } U, \tag{2.2}$$

where U is an open subset of $\mathbb{R}^d$. Let us assume $b \in C^\infty(U;\mathbb{R}^d)$ and suppose that we are already given with a solution $u : U \to \mathbb{R}$ that is smooth, i.e., $u \in C^\infty(U;\mathbb{R})$ (we will figure out the sharp regularity condition needed later; note that $f \in C^\infty(U;\mathbb{R})$ if u, b are as above).

Intuitively, (2.2) tells us that the rate of change of u in the direction $b(x_0)$ at x_0 is f. To make use of this information, we may integrate the direction field $b(x)$ to obtain a curve $\gamma = \{x(s)\}$ (called an *integral curve* of b):

$$\dot{x}(s) = b(x(s)), \quad x(s_0) = x_0.$$

Then, intuitively, the rate of change of $u(x(s))$ at $s = s_0$ is given by $f(x(s_0))$. Indeed, by the chain rule,

$$\frac{\mathrm{d}}{\mathrm{d}s}u(x(s))|_{s=s_0} = \sum_j \dot{x}^j(s_0)(\partial_j u)(x(s_0)) = \sum_j b^j(x(s_0))(\partial_j u)(x(s_0)) = f(x(s_0)).$$

Observe at this point that the above computation makes sense if $u \in C^1(U)$ and $x \in C^1(I)$ (where I is the domain of x), and the latter condition is ensured by the Cauchy–Lipschitz theorem if $b \in C^1(U)$ (hence, $k = k_0 = 1$); it is then natural to put $f \in C^0(U)$.

In conclusion, let $b \in C^1(U;\mathbb{R}^d)$ and $f \in C^0(U)$, then consider the system of ODEs

$$\dot{x}(s) = b(x(s)), \tag{2.3}$$

$$\dot{z}(s) = f(x(s)). \tag{2.4}$$

If u is a C^1 solution to (2.2) and x is a C^1 curve solving (2.3), then $z(s) := u(x(s))$ obeys (2.4). This gives us a simple way to describe the solution u in terms of solutions (x, z) to the system of ODEs (2.3)–(2.4), which can be solved (with unique C^1 solution) if $b \in C^1(U;\mathbb{R}^d)$ and $f \in C^0(U)$.

2.1.2. *Derivation of characteristic equations.* We now derive the *characteristic equations* for the general equation (2.1). Let us start with a smooth solution u that solves (2.1) in U. The idea, as before, is to find curves γ in U along which the solution is described by a set of ODEs.

To properly begin our discussion, let us assume $F \in C^\infty$, and suppose that we are already given with a solution $u : U \to \mathbb{R}$ with $u \in C^\infty(U;\mathbb{R})$ (again, the correct C^k-regularity will be figured out later). Consider a smooth curve γ in U parametrized by $x : I \to U$, i.e., $\gamma = \{x(s) : s \in I\}$. In the simple case of (2.2), we found the correct curve γ by geometric intuition. In the general case, we will perform a more analytical reasoning. Here, it turns out that we also need to keep track of $\partial_j u$ in addition to u along γ. We write z and p_j for the values of u and $\partial_j u$ on the point $x(s)$ on γ:

$$z(s) = u(x(s)), \quad p_j(s) = \partial_j u(x(s)).$$

On the one hand, the enemy for getting a closed system of ODEs for (x, z, p) is the fact that, in general, we expect $\dot{p}_j(s)$ to involve second order derivatives of u, which cannot be written in terms of (x, z, p). Indeed, by the chain rule,

$$\dot{p}_j(s) = \frac{\mathrm{d}}{\mathrm{d}s}\partial_j u(x(s)) = \sum_k \dot{x}^k \partial_j \partial_k u(x(s)).$$

On the other hand, we have the freedom of choosing the evolution equation for the curve $x(s)$. The idea is to use this freedom to deal with the above problem! To get information about the the second-order derivatives of the solution u, we differentiate (2.1) in ∂_j and obtain, by the chain rule,

$$\begin{aligned} 0 = (\partial_{x^j} F)(x, u(x), Du(x)) &+ \partial_{x^j} u (\partial_z F)(x, u(x), Du(x)) \\ &+ \sum_k \partial_j \partial_k u (\partial_{p_k} F)(x, u(x), Du(x)). \end{aligned}$$

After restricting to γ, we have

$$0 = (\partial_{x^j} F)(x, z, p) + p_j (\partial_z F)(x, z, p) + \sum_k \partial_j \partial_k u(x(s)) (\partial_{p_k} F)(x, z, p).$$

The key observation – which makes the method of characteristics work for (2.1) – is that the last term very much resembles the expression for $\dot{p}_j$ we had above! This motivates us to select the ODE for x to be

$$\dot{x}^k = \partial_{p_k} F(x, z, p),$$

so that

$$\dot{p}_j = -(\partial_{x^j} F)(x, z, p) - p_j (\partial_z F)(x, z, p).$$

Finally, $\dot{z}$ can be computed from the chain rule:

$$\dot{z} = \frac{\mathrm{d}}{\mathrm{d}s} u(x(s)) = \sum_j \dot{x}^j(s) \partial_j u(x(s)) = (\partial_{p_j} F)(x, z, p) p_j.$$

What we have done so far can be summarized as follows. We introduce the *characteristic equations* as follows:

$$\dot{x}^j(s) = \partial_{p_j} F(x(s), z(s), p(s)), \tag{2.5}$$

$$\dot{z}(s) = \sum_j \partial_{p_j} F(x(s), z(s), p(s)) p_j(s), \tag{2.6}$$

$$\dot{p}_j(s) = -\partial_{x^j} F(x(s), z(s), p(s)) - \partial_z F(x(s), z(s), p(s)) p_j(s). \tag{2.7}$$

The above argument, with now attention to how many continuous derivatives are needed, shows:

Theorem 2.1. *Let $u \in C^2(U)$ be a solution to $F(x, u, Du) = 0$ in U, where F is C^1. If $x(s)$, which lies in U for $s \in I$, solves the ODE* (2.5)*, then $z(s) = u(x(s))$ and $p_j(x(s))$ obey* (2.6) *and* (2.7)*, respectively.*

The system (2.5)-(2.7) is called the *characteristic equations* of (2.1). The solutions $(x(s), z(s), p(s)) \in \mathbb{R}^{2d+1}$ are called the *characteristics*, and $x(s)$ is referred to as the *projected characteristic*.

2.1.3. *Examples: finding solution via the method of characteristics.* Let us first recover the special case (2.2) from Theorem 2.1 for the general case (2.1).

- *F is linear.* Consider the equation

$$\sum_j b^j(x)\partial_j u(x) + c(x)u(x) = f$$

Then $F = \sum_j b^j(x)p_j + c(x)z - f(x)$. Thus,

$$\partial_{p_j} F(x, z, p) = b^j(x)$$

and

$$\dot{x}^j(s) = b^j(x(s)), \quad \dot{z}(s) = \sum_j b^j(x(s))p_j(s) = -c(x(s))z(s) + f(x(s)).$$

When $c = 0$, this is precisely the inhomogeneous transport equation (see also the exercise above). Note that the ODE for p is not needed to determine $u(x(s)) = z(s)$. Moreover, observe that these two equations have a hierarchy: The ODE for x is closed by itself, and once we know x, we can solve the ODE for z.

From now on, we consider the boundary value problem associated with (2.1). Given an open subset U of $\mathbb{R}^d$ and an open subset Γ of ∂U, consider

$$\begin{cases} F(x, u, Du) = 0 & \text{in } U, \\ u = g & \text{on } \Gamma. \end{cases}$$

Logically, Theorem 2.1 presupposes the existence of a solution $u \in C^2(U)$ of the above equation. Nevertheless, we can turn the table around and *define* $u(x)$ by:

(1) solving the characteristic equations (2.5)–(2.7) with initial conditions

$$x(0) \in \Gamma, \quad z(0) = g(x(0)),$$

and $p(0)$, if necessary[2], determined implicitly from $F(x(0), z(0), p(0)) = 0$;

(2) declaring $u(x(s)) = z(s)$ for each such solution $(x(s), z(s), p(s))$.

We may then argue that $u(x)$, if well-defined, is a good candidate for the solution to the problem at hand. Later, we will develop a general theory to identify conditions under which u is a well-defined solution (for instance, this will always be the case sufficiently close to Γ). But it is often easy to just verify this by hand, without needing such theoretical considerations. This is the practical way of solving boundary value problems using the method of characteristics!

With the above discussion in mind, let us cover some examples first, then develop the general theory later.

- *F is linear, example.*

$$\begin{cases} x^1 u_{x^2} - x^2 u_{x^1} = u & \text{in } U, \\ u = g & \text{on } \Gamma, \end{cases}$$

where $U = \{x^1 > 0,\, x^2 > 0\}$ and $\Gamma = \{x^1 > 0,\, x^2 = 0\} \subset \partial U$.

The characteristic equations are:

$$\dot{x}^1 = -x^2, \quad \dot{x}^2 = x^1, \quad \dot{z} = z.$$

[2] It turns out that the ODE for p is not necessary as long as the equation is quasilinear; see below.

Thus,

$$x^1(s) = y\cos s, \quad x^2(s) = y\sin s, \quad z(s) = g(y)e^s,$$

where $(y,0) \in \Gamma$, $0 \leq s \leq \frac{\pi}{2}$. Given a point $(x^1, x^2) \in U$, we need to find y and s so that $(x^1, x^2)(s) = (x^1, x^2)$. By elementary geometry, we see that

$$y = ((x^1)^2 + (x^2)^2)^{\frac{1}{2}}, \quad s = \arctan\left(\frac{x^2}{x^1}\right).$$

Thus,

$$u(x) = g\left(((x^1)^2 + (x^2)^2)^{\frac{1}{2}}\right)\exp\left(\arctan\left(\frac{x^2}{x^1}\right)\right).$$

Once we arrive at this expression, we can check that $u(x)$ is indeed a solution we want; the above argument shows that it is unique.

- *F is quasilinear.* In this case, F must be of the form

$$F(x, u, Du) = \sum_j b^j(x,u) \cdot Du + c(x,u).$$

Then the characteristic equations are

$$\dot{x}^j(s) = b^j(x(s), z(s)), \quad \dot{z}(s) = -c(x, z(s)).$$

Again, the ODE for p is not needed to determine $u(x(s)) = z(s)$. However, the ODEs for x and z are now coupled.

- *F is quasilinear, example.*

$$\begin{cases} u_{x^1} + u_{x^2} = u^2 & \text{in } U, \\ u = g & \text{on } \Gamma, \end{cases}$$

where U is the half-space $\{x^2 > 0\}$ and $\Gamma = \{x^2 = 0\}$. Then

$$\dot{x}^1 = 1, \quad \dot{x}^2 = 1, \quad \dot{z} = z^2.$$

Thus

$$x^1 = y + s, \quad x^2 = s, \quad z(s) = \frac{g(y)}{1 - sg(y)},$$

where $(y,0) \in \Gamma$ and $s \geq 0$, provided that $z(s)$ is well-defined. For $x \in U$, we choose $x^0 = x^1 - x^2$ and $s = x^2$. Thus,

$$u(x) = \frac{g(x^1 - x^2)}{1 - x^2 g(x^1 - x^2)}.$$

Once we arrive at this expression, we can check that $u(x)$ is indeed a solution we want; the above argument shows that it is unique.

- *F is quasilinear, example 2 (Burgers equation).*

$$\begin{cases} u_{x^0} + uu_{x^1} = 0 & \text{in } U, \\ u = g & \text{on } \Gamma, \end{cases}$$

where U is the half-space $\{x^0 > 0\}$ and $\Gamma = \{x^0 = 0\}$. Then

$$\dot{x}^0 = 1, \quad \dot{x}^1 = z, \quad \dot{z} = 0.$$

Thus

$$x^0 = s, \quad x^1 = y + sg(y), \quad z(s) = g(y)$$

where $(0,y) \in \Gamma$ and $s \geq 0$. Note, however, that the projected characteristics may now collide!

In the quasilinear case, observe that we did not have to consider the characteristic ODE for p, and hence its initial data. In the fully nonlinear case, consideration of p may be necessary, in which case the initial data for p is determined implicitly from $F(x, z, p) = 0$ on Γ.

- *F is fully nonlinear, example.*

$$\begin{cases} u_{x^1} u_{x^2} = u & \text{in } U, \\ u = g & \text{on } \Gamma, \end{cases}$$

where $U = \{x^1 > 0\}$ and $\Gamma = \{x^1 = 0\}$. Here, the characteristic equations are

$$\dot{x}^1 = p_2, \quad \dot{x}^2 = p_1, \quad \dot{z} = 2p_1 p_2, \quad \dot{p}_1 = p_1, \quad \dot{p}_2 = p_2.$$

We integrate these equations to find

$$\begin{aligned} x^1(s) &= (p_0)_2(e^s - 1), \quad x^2(s) = y + (p_0)_1(e^s - 1), \\ z(s) &= z_0 + (p_0)_1(p_0)_2(e^{2s} - 1), \\ p_1(s) &= (p_0)_1 e^s, \quad p_2(s) = (p_0)_2 e^s. \end{aligned}$$

where $(0, y) \in \Gamma$, $s \in \mathbb{R}$. Note that $z_0 = g(y)$ and $(p_0)_2 = g'(y)$. Moreover, by the PDE itself,

$$(p_0)_1 = \frac{z_0}{(p_0)_2} = \frac{g(y)}{g'(y)}.$$

Let us further restrict to

$$g(y) = y^2,$$

and perform the exercise of determining the solution. We have

$$z_0 = y^2, \quad (p_0)_1 = \frac{y}{2}, \quad (p_0)_2 = 2y.$$

Given $(x^1, x^2) \in U$, we need to determine y, s such that

$$(x^1, x^2) = ((p_0)_2(e^s - 1), y + (p_0)_1(e^s - 1)) = (2y(e^s - 1), y + \frac{y}{2}(e^s - 1)).$$

The answer is

$$y = \frac{4x^2 - x^1}{4}, \quad e^s = \frac{x^1 + 4x^2}{4x^2 - x^1}$$

and thus

$$u(x) = z(s) = y^2 e^{2s} = \frac{(x^1 + 4x^2)^2}{16}.$$

Once we arrive at this expression, we can check that $u(x)$ is indeed a solution we want; the above argument shows that it is unique.

- *F is fully nonlinear, example 2.* Another example of a fully nonlinear equation is *the Hamilton–Jacobi equation*,

$$\partial_t u + H(x, \partial_x u) = 0.$$

The characteristic equations are:

$$\dot{x}^j = \partial_{p_j} H(x, p), \quad \dot{p}_j = -\partial_{x^j} H(x, p).$$

These are, in fact, *Hamilton's equations* in classical mechanics. We explore this topic deeper later.

2.2. **Local existence and uniqueness for boundary value problems.** Let us now develop the general theory and prove theorems concerning the local existence and uniqueness for boundary value problems for (2.1) (for a discussion of continuous dependence, see Proposition 2.11 below).

Let U be an open subset of $\mathbb{R}^d$, and $\Gamma \subseteq \partial U$. We will be interested in the BVP

$$\begin{cases} F(x, u, Du) = 0 \text{ in } U, \\ \qquad\qquad u = g \text{ on } \Gamma. \end{cases} \tag{2.8}$$

We make the following definitions. For any subset S of $\mathbb{R}^d$, $C(S)$ is simply the set of all continuous functions on S. We define $C^k(S)$ as the collection of all functions $u \in C(S)$ that extends to a C^k function $\widetilde{u}$ in an open neighborhood U of S such that $\partial^\alpha \widetilde{u} \in C(S)$ for all $|\alpha| \leq k$. The reason for the somewhat longwinded definition of $C^k(S)$ is because we wish to avoid taking derivatives on the boundary.

The regularity assumptions we need on F, g, etc. will be made precise as we develop the theory.

2.2.1. *Local existence theory, when* Γ *is flat.* Our next goal is to formulate an existence theory for (2.8), at least in a neighborhood of a boundary point in Γ. For simplicity, we first consider the special case when

$$\Gamma \text{ is an open subset of } \{x^d = 0\}. \tag{2.9}$$

We may now parametrize Γ by $(y, 0)$ with $y = (y^1, \dots, y^{d-1}) \in \mathbb{R}^{d-1}$.

Our strategy will again be thinking about all characteristics $(x(s), z(s), p(s))$ such that $x(0) \in \Gamma$. We first need to ask: *what are the compatibility conditions that the initial conditions* $x_0 := x(0)$, $z_0 := z(0)$ *and* $p_0 := p(0)$ *must satisfy if it comes from a solution to* (2.8)*?* The answer is provided by the following definition:

Definition 2.2 (Admissible boundary data, when Γ is flat)**.** We say that the triple (x_0, z_0, p_0) is *admissible* if $x_0 = (y_0, 0) \in \Gamma$ and the following holds:

$$\begin{cases} \qquad\quad z_0 = g(y_0), \\ \quad (p_0)_j = \partial_j g(y_0) \qquad \text{for } j = 1, \dots, d-1, \\ F(x_0, z_0, p_0) = 0. \end{cases}$$

As intended, if (x_0, z_0, p_0) arises from an actual solution u to (2.8), then clearly it must be admissible. Note that the above constitute $d+1$ equations for $d+1$ variables $(z_0, (p_0)_1, \dots, (p_0)_d)$, but the solution need not exist nor be unique.

Next, we ask: *given an admissible boundary data* (x_0, z_0, p_0)*, what is the condition that ensures that admissible boundary data can be found for nearby points?* That is, for $y = (y^1, \dots, y^{d-1})$ close to y_0 (i.e., $x_0 = (y_0, 0)$), we wish to find $(q_1(y), \dots, q_d(y))$ such that

$$(p_0)_j = q_j(y_0)$$

and

$$\begin{cases} \qquad q_j(y) = \partial_j g(y) \text{ for } j = 1, \dots, d-1, \\ F(y, g(y), q(y)) = 0. \end{cases}$$

The general tool we need is the implicit function theorem:

Theorem 2.3 (Implicit function theorem)**.** *Let* $\mathbf{F} : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ *be a* C^1 *function, and let* $\mathbf{x}_0 \in \mathbb{R}^d$, $\mathbf{y}_0 \in \mathbb{R}^n$ *satisfy:*

- $\mathbf{F}(\mathbf{x}_0, \mathbf{y}_0) = 0$,

· $\det \partial_{\mathbf{y}^j} \mathbf{F}^k(\mathbf{x}_0, \mathbf{y}_0) \neq 0$.

Then there exist neighborhoods $U \ni \mathbf{x}$ *and* $V \ni \mathbf{y}$*, and a* C^1 *function* $U \to V$, $\mathbf{x} \mapsto \mathbf{y}(\mathbf{x})$ *such that*

· $\mathbf{y}_0 = \mathbf{y}(\mathbf{x}_0)$ *and* $F(\mathbf{x}, \mathbf{y}(\mathbf{x})) = 0$*;*
· *If* $(\mathbf{x}, \mathbf{y}) \in U \times V$ *satisfy* $F(\mathbf{x}, \mathbf{y}) = 0$*, then* $\mathbf{y} = \mathbf{y}(\mathbf{x})$.

Moreover, if $\mathbf{F} \in C^k$*, then so is* $\mathbf{y}$ *on* U.

We remark that $\partial_{\mathbf{x}^j} \mathbf{y}^k$ can be computed by implicit differentiation.
We are now ready to provide the answer to the question:

Definition 2.4 (Noncharacteristic boundary data, when Γ is flat)**.** We say that an admissible triple (x_0, z_0, p_0) is *noncharacteristic* if

$$\partial_{p_d} F(x_0, z_0, p_0) \neq 0.$$

By the implicit function theorem, this assumption allows us to solve for $(p_0)_d$ as a function of $(y^1, \dots, y^{d-1})$ provided that F is C^2.

- *Example:* Consider the case when F is quasilinear.

$$F(x, z, p) = \sum_j b^j(x, z) p_j + c(x, z).$$

Then the noncharacteristic condition is $b^d(x_0, z_0) \neq 0$ regardless of the choice of p_0. Moreover, it allows us to determine $(p_0)_d$ uniquely by writing

$$(p_0)_d = -\frac{1}{b^d(x_0, z_0)} \left(\sum_{j=1}^{d-1} b^j(x_0, z_0)(p_0)_j + c(x_0, z_0) \right).$$

So for a quasilinear first-order scalar PDE, the noncharacteristic condition allows us to uniquely determine $(p_0)_d$ from the data.

- *Example:* To motivate the general formulation of the theorem, consider the simple fully nonlinear first-order scalar equation

$$\begin{cases} (\partial_x u)^2 = 1 \in \text{ in } (0, \infty) \\ u = g \text{ at } x = 0. \end{cases}$$

or $F(x, u, Du) = 0$ with

$$F(x, z, p) = p^2 - 1.$$

With $x_0 = 0$, there are two choices of admissible boundary values: $(x_0, z_0, p_0) = (0, g, 1)$ and $(x_0, z_0, p_0) = (0, g, -1)$. Both admissible triples are noncharacteristic.

We now return to the development of a local existence theory for (2.8). Assume that $F \in C^2(U \times \mathbb{R} \times \mathbb{R}^d)$ and $g \in C^2(\Gamma)$. Given $x_0 \in \Gamma$, let (x_0, z_0, p_0) be a noncharacteristic boundary data for (2.8). Abusing the notation a bit, we find, by the implicit function theorem, a neighborhood W of y in Γ, a C^1 function $p_0(y)$ on W and $z_0(y) = g(y)$ (which is C^2) such that (i) $((y, 0), z_0(y), p_0(y))$ is admissible for all $y \in W$ and (ii) $((y_0, 0), z_0(y_0), p_0(y_0)) = (x_0, z_0, p_0)$. For each $y \in W$, denote by $(x(y, s), z(y, s), p(y, s))$ the unique solution to (2.5)–(2.7) with initial conditions

$$(x(y, 0), z(y, 0), p(y, 0)) = ((y_0, 0), z_0(y), p_0(y)),$$

defined on the maximal interval $(S_-(y), S_+(y))$ (that contains 0).

To construct the solution u, the key step is to invert the map $(y^1, \dots, y^{d-1}, s) \mapsto (x^1, \dots, x^d)$. In some neighborhood W of $x_0 = (y_0, 0)$. To show this, we use the inverse function theorem:

Theorem 2.5 (Inverse function theorem)**.** *Let $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a C^1 function such that $\mathrm{D}\mathbf{F}(\mathbf{x}_0)$, viewed as a linear map, is invertible. Then there exists a neighborhood V of $\mathbf{x}_0$ such that $\mathbf{F}$ is invertible on V. Moreover, if $\mathbf{F} \in C^k$ for $k \geq 1$, so is the inverse on V.*

We remark that $\mathrm{D}\mathbf{F}^{-1}$ can be computed by implicit differentiation.

To apply the inverse function theorem, we need to compute the matrix

$$\begin{pmatrix} \partial_{y^1} x^1 & \cdots & \partial_{y^{d-1}} x^1 & \partial_s x^1 \\ \vdots & & \vdots & \vdots \\ \partial_{y^1} x^d & \cdots & \partial_{y^{d-1}} x^d & \partial_s x^d \end{pmatrix} (x_0) = \begin{pmatrix} & & & \partial_{p^1} F(x_0, z_0, p_0) \\ & I_{(d-1)\times(d-1)} & & \vdots \\ & & & \\ 0 & \cdots & 0 & \partial_{p^d} F(x_0, z_0, p_0), \end{pmatrix}$$

which is clearly invertible if and only if $\partial_{p^d} F(x_0, z_0, p_0) \neq 0$. Therefore, there exists an C^1-inverse $x \mapsto (y, s)$ for x in some neighborhood V of x_0.

Finally, we are ready to construct u by defining

$$u(x) := z(y(x), s(x)) \text{ for } x \in V.$$

The claim is that this gives a local solution:

Theorem 2.6 (Local existence theorem)**.** *Assume that $F \in C^2(U \times \mathbb{R} \times \mathbb{R}^d)$ and $g \in C^2(\Gamma)$. Then the function u defined above is C^2 and solves the PDE*

$$F(x, u(x), Du(x)) = 0 \text{ in } U \cap V,$$

with the boundary condition

$$u = g \text{ on } \Gamma \cap V.$$

Proof. Given $y \in W$ (i.e., a neighborhood of x_0 in Γ), solve the characteristics equations and define $(x, z, p)(y, s)$. Denote by $(y(x), s(x))$ the inverse of $(y, s) \mapsto x(y, s)$ defined for $x \in V$ (i.e., a neighborhood of x_0 in $\mathbb{R}^d$). Define

$$u(x) = z(y(x), s(x)), \quad v_j(x) = p_j(y(x), s(x)).$$

That u satisfies the boundary condition is clear from the construction. We will split the proof that u solves the PDE into establishing the following two claims:

$$F(x, u(x), v(x)) = 0 \quad \text{for all } x \in V, \tag{2.10}$$

$$\partial_j u(x) = v_j(x) \quad \text{for all } x \in V. \tag{2.11}$$

Step 1: Proof of (2.10). It suffices to prove that, for every $y \in W$,

$$f(y, s) := F(x(y, s), z(y, s), p(y, s))$$

vanishes for all possible s. To see this, we simply compute

$$\partial_s f(y, s) = 0,$$

using the characteristic equations, and also note that $f(y, 0) = 0$ by our choice of the initial conditions $(x, z, p)(y, 0)$.

Step 2: Proof of (2.11). We begin by computing

$$\partial_{x^j} u(x) = \partial_{x^j} (z(y(x), s(x))$$

$$= \sum_{k=1}^{d-1} \partial_{x^j} y^k(x) \partial_{y^k} z(y(x), s(x)) + \partial_{x^j} s(x) \partial_s z(y(x), s(x)).$$

We need to know $\partial_{y^k} z(y(x), s(x))$ and $\partial_s z(y(x), s(x))$. We claim that

$$\partial_s z(y, s) = \sum_\ell p_\ell(y, s) \partial_s x^\ell(y, s), \tag{2.12}$$

$$\partial_{y^k} z(y, s) = \sum_\ell p_\ell(y, s) \partial_{y^k} x^\ell(y, s), \tag{2.13}$$

Let us first show how these claims imply (2.11). Continuing the preceding computation using (2.12)–(2.13), we have

$$\begin{aligned}
\partial_{x^j} u(x) = & \sum_{k=1}^{d-1} \partial_{x^j} y^k(x) \sum_\ell p_\ell(y(x), s(x)) \partial_{y^k} x^\ell(y(x), s(x)) \\
& + \partial_{x^j} s(x) \sum_\ell p_\ell(y(x), s(x)) \partial_s x^\ell(y(x), s(x)) \\
= & \sum_\ell v_\ell(x) \left(\sum_{k=1}^{d-1} \partial_{x^j} y^k(x) \partial_{y^k} x^\ell(y(x), s(x)) + \partial_{x^j} s(x) \partial_s x^\ell(y(x), s(x)) \right) \\
= & \sum_\ell v_\ell(x) \partial_{x^j} x^\ell = v_j(x),
\end{aligned}$$

as desired.

It remains to verify (2.12) and (2.13) from the characteristic equations (2.5)–(2.7). In fact, (2.12) is simply the combination of (2.5) and (2.6). To prove (2.13), we fix $y \in W$ and track the quantity

$$r_k(s) := \partial_{y^k} z(y, s) - \sum_\ell p_\ell(y, s) \partial_{y^k} x^\ell(y, s).$$

Using (2.5), computing $\partial_s \partial_{y^k} z$ and trying to move ∂_s away from the highest order term (in order to be able to apply (2.5)–(2.7)), we compute

$$\begin{aligned}
\partial_s \partial_{y^k} z = \partial_{y^k} \partial_s z = & \partial_{y^k} \left[\sum_j p_j \partial_s x \right] \\
= & \sum_j \left(\partial_{y^k} p_j \partial_s x^j + p_j \partial_{y^k} \partial_s x^j \right) \\
= & \partial_s \left(\sum_j p_j \partial_{y^k} x^j \right) + \sum_j \left(\partial_{y^k} p_j \partial_s x^j - \partial_s p_j \partial_{y^k} x^j \right).
\end{aligned}$$

Note that the terms inside $\partial_s(\cdots)$ on both sides combine to give r_k. Using (2.5) and (2.7), we have

$$\begin{aligned}
\partial_s r_k = & \sum_j \left(\partial_{y^k} p_j \partial_s x^j - \partial_s p_j \partial_{y^k} x^j \right) \\
= & \sum_j \left(\partial_{y^k} p_j \partial_{p_j} F(x, z, p) + \partial_{x^j} F(x, z, p) \partial_{y^k} x^j + p_j \partial_z F(x, z, p) \partial_{y^k} x^j \right)
\end{aligned}$$

$$= -\partial_z F(x,z,p)\partial_{y^k} z + \sum_j p_j \partial_z F(x,z,p)\partial_{y^k} x^j,$$

where, in the last equality, we used

$$0 = \partial_{y^k}\left(F(x,z,p)\right) = \sum_j \partial_{y^k} x^j \partial_{x^j} F(x,z,p) + \partial_{y^k} z \partial_z F(x,z,p) + \sum_\ell \partial_{y^k} p_\ell \partial_{p_\ell} F(x,z,p),$$

which follows by taking ∂_{y^k} to the conclusion of Step 1. Again, after factoring out $-\partial_z F(x,z,p)$, we get r_k. In conclusion, we arrive at

$$\dot{r}_k(s) = -(\partial_z F)(x(y,s), z(y,s), p(y,s))r^i(s).$$

Since $r_k(0) = 0$ by our choice of $(x,z,p)(y,0)$, it follows that $r_k = 0$ for all s, which proves (2.13). □

Remark 2.7. Formally speaking, it is the proof of this theorem guarantees that the $u(x)$ defined using the method of characteristics in the example above indeed solves $F(x,u,Du) = 0$ at the points where u is defined.

2.2.2. *Local existence theory - general case.* By a local coordinate change, the general case can be put in to the flat hyperplane case. First, let us be precise about the regularity of the boundary of a domain.

Definition 2.8. We say that *the boundary ∂U is C^k* if for every point $x_0 \in \partial U$, there exists $r > 0$ and a C^k function $\gamma : \mathbb{R}^{d-1} \to \mathbb{R}$ such that, after relabeling and reorienting the coordinate axes if necessary, we have

$$U \cap B(x_0, r) = \{x \in B(x_0, r) : x^d > \gamma(x^1, \dots, x^{d-1})\}.$$

For a C^1 boundary ∂U, we can associate the notion of a *outward normal vector field* $\nu_{\partial U} = (\nu^1, \dots, \nu^d)$.

Let ∂U be a C^k boundary. Near $x_0 \in U$, we can make a change of coordinates to "flatten out" ∂U. Define

$$\widetilde{x}^i = x^i \quad \text{for } i = 1, \dots, d-1,$$
$$\widetilde{x}^d = x^d - \gamma(x^1, \dots, x^{d-1}).$$

where $B(x_0, r)$ and γ are from the definition of a C^k boundary. Note that in $(\widetilde{x}^1, \dots, \widetilde{x}^d)$, the boundary is now a subset of the hyperplane $\{\widetilde{x}^d = 0\}$, and $U \subset \{\widetilde{x}^d > 0\}$. We often write this change of coordinates as a map $\widetilde{x}^i = \Phi^i(x)$ $(i = 1, \dots d)$. Its inverse can be easily found:

$$x^i = (\Phi^{-1})^i(\widetilde{x}) = \widetilde{x}^i \quad \text{for } i = 1, \dots, d-1,$$
$$x^d = (\Phi^{-1})^d(\widetilde{x}) = \widetilde{x}^d + \gamma(\widetilde{x}^1, \dots, \widetilde{x}^{d-1}).$$

Observe that finding a solution u to $F(x,u,Du) = 0$ is the same as finding a solution $\widetilde{u}(\widetilde{x}) := u(x(\widetilde{x}))$ (or $\widetilde{u} := u \circ (\Phi^{-1})$) to $\widetilde{F}(\widetilde{x}, \widetilde{u}, D_{\widetilde{x}}\widetilde{u}) = 0$, where

$$\widetilde{F}(\widetilde{x}, \widetilde{u}(\widetilde{x}), D_{\widetilde{x}}\widetilde{u}(\widetilde{x})) = F(x,z,p) = 0,$$

with $x = \Phi^{-1}(\widetilde{x})$, $z = \widetilde{u}(\widetilde{x})$ and $p_j = \sum_k \partial_{x^j}\widetilde{x}^k(\Phi^{-1}(\widetilde{x}))\partial_{\widetilde{x}^k}\widetilde{u}(\widetilde{x})$. Therefore, the general situation is reduced to the flat boundary case!

In the general case, the noncharacteristic condition now reads

$$\sum_j \nu^j_{\partial U} \partial_{p_j} F(x_0, z_0, p_0) \neq 0.$$

where $\nu_{\partial U}$ is the outer unit normal to ∂U at x_0.

We have a local existence theorem analogous to Theorem 2.9 for a noncharacteristic admissible triple (x_0, z_0, p_0).

Theorem 2.9 (Local existence theorem)**.** *Assume that ∂U is C^2, $F \in C^2(U \times \mathbb{R} \times \mathbb{R}^d)$ and $g \in C^2(\Gamma)$. Let (x_0, z_0, p_0) be a noncharacteristic admissible triple, where $x_0 \in \Gamma$. Then there exists a neighborhood V of x_0 and a function $u \in C^2(\overline{U} \cap V)$ that solves the PDE*

$$F(x, u(x), Du(x)) = 0 \text{ in } U \cap V,$$

with the boundary condition

$$u = g \text{ on } \Gamma \cap V.$$

2.2.3. *Uniqueness for* (2.8). Next, we turn to uniqueness, which is more straightforward to study using the method of characteristics.

Assume that we are given a sufficiently regular solution u to (2.8). For each $x_0 \in \Gamma$, consider a triple $(x(x_0, s), z(x_0, s), p(x_0, s))$ with $z(x_0, s) := u(x(x_0, s))$, $p(x_0, s) := Du(x(x_0, s))$ and $x(x_0, s)$ solving (2.5) with the initial condition $x(x_0, 0) = x_0 \in \Gamma$. By Theorem 2.1, we know that $(x(x_0, s), z(x_0, s), p(x_0, s))$ solves the characteristic ODEs (2.5)–(2.7).

Now, assume that $F \in C^2$, and for each $x_0 \in \Gamma$, denote by $[0, S_+(x_0))$ the maximal interval for which Theorem 1.1 (the fundamental theorem of ODEs) applies to the characteristic ODEs (2.5)–(2.7) (note that $F \in C^2$, or $C^{1,1}$ with initial conditions $(x_0, z(x_0, 0), p(x_0, 0))$; note also that $S_+(x_0)$ may be $+\infty$). Define

$$V_u := \{x(x_0, s) : x_0 \in \Gamma,\ s \in [0, S_+(x_0))\}.$$

Geometrically, V_u is the union of the image of all projected characteristics associated with u emanating from Γ. Observe that V_u is open if $(x_0, u(x_0), Du(x_0))$ is noncharacteristic at every $x_0 \in \Gamma$, and Γ is an open subset of ∂U. In view of the uniqueness assertion in Theorem 1.1, we now arrive at the following result:

Theorem 2.10 (Uniqueness for (2.8))**.** *Let $F \in C^2(U \times \mathbb{R} \times \mathbb{R}^d)$ and $u \in C^2(\overline{U})$, and let V_u be defined as above. If $v \in C^2(\overline{U})$ is a solution to* (2.8) *such that $\nu_\Gamma \cdot Du = \nu_\Gamma \cdot Dv$ on Γ, then $u = v$ on V_u.*

Here, ν_Γ is the outward unit normal to Γ. The reason why we need to assume $\nu_\Gamma \cdot Du = \nu_\Gamma \cdot Dv$ is because $F(x, u, Du) = 0$ might not uniquely determine the normal derivative of the solution, as we have seen in the discussion of the local existence theory. We leave filling in the details of the proof as an exercise.

2.2.4. *Additional remarks.*

- What about continuous dependence for (2.8)? It does hold in the reasonable setting, but in a subtle way!

  Proposition 2.11. *Let $g \in C^2(\Gamma)$ with $\|g\|_{C^2(\Gamma)} \leq A$. Suppose that at all points $y \in \Gamma$, we have a continuous choice of noncharacteristic triples, which also depends continuously on g. Then the solution to the boundary value problem with boundary data g given by Theorem 2.9 exists in a neighborhood V of x_0 that is independent of g; let us call this solution $u[g]$. The map $g \mapsto u[g]$ from $C^2(\Gamma)$ to $C^2(V)$ is continuous.*

However, there exists an example for which

$$\|u[g^{(n)}] - u[h^{(n)}]\|_{C^2(V)} > n\|g^{(n)} - h^{(n)}\|_{C^2(\Gamma)}$$

for $n \nearrow \infty$; i.e., the solution map is not Lipschitz in general. For more on this topic, see: https://terrytao.wordpress.com/2010/02/21/quasilinear-wellposedness/.

- **Applications.** *F linear.* When $F(x, u, Du) = \sum_j b^j(x)\partial_j u(x) + c(x)u(x) = 0$, the noncharacteristic assumption at a point $x_0 \in \Gamma$ becomes

$$\sum_j b^j \nu_\Gamma^j = 0,$$

which does not involve x_0 or p_0 at all.

Example. Consider the boundary value problem

$$\begin{cases} \sum_j b^j(x)\partial_j u = 0 \text{ in } U \\ u = g \text{ in } \Gamma. \end{cases}$$

– Case 1 from Evans (flow to an attracting point). If we take $\Gamma = \partial U$, then u is obtained by setting the solution to be constant on each projected characteristic; u is not well-defined at the attracting point, unless $g = const$.

– Case 2 from Evans (flow across a domain). If we take $\Gamma = \{x \in \partial U : \sum b^j \nu_{\partial U}^j < 0\}$ (i.e., points on $x \in U$ at which b^j points inside U), then the smooth solution u can be found by setting the solution to be constant on each projected characteristic.

– Case 3 from Evans (flow with characteristic points). If we define u to be constant on each projected characteristic, then u is discontinuous (at D, according to the labels in Evans).

Example. Consider the boundary value problem

$$\begin{cases} \partial_{x^1} u + u\partial_{x^2} u = 2u \text{ in } \mathbb{R}^2, \\ u(x^1, 0) = x^1 \text{ in } \Gamma = \mathbb{R} \times \{0\}. \end{cases}$$

Find the smooth solution u on the maximal domain of existence U.

Here, the characteristic equations are

$$\dot{x}^1 = 1, \quad \dot{x}^2 = z, \quad \dot{z} = 2.$$

The characteristic with initial data $(x_0, 0, x_0)$ is given by

$$x^1(s) = x_0 + s, \quad x^2(s) = x_0 s + s^2, \quad z(s) = x_0 + 2s.$$

Let us study the projected characteristics carefully. Substituting $s = x^1 - s$ in the equation for $x^2(s)$, we see that the graphs of the projected characteristics are given by

$$\begin{aligned} x^2 &= x_0(x^1 - x_0) + (x^1 - x_0)^2 \\ &= -x_0 x^1 + (x^1)^2. \end{aligned}$$

When we formally solve for (x_0, s) in terms of (x^1, x^2), then we obtain

$$x_0 = \frac{(x^1)^2 - x^2}{x^1}, \quad s = \frac{x^2}{x^1},$$

which only makes sense when $x^1 \neq 0$. Moreover, when $x^1 = 0$, $x^2 = 0$ on the projected characteristic regardless of what x_0 is. In other words, when $x_0 \neq 0$, the projected characteristics are parabola that are concave upward and passes through Γ at points $(0,0)$ and $(x_0, 0)$. When $x_0 = 0$, the projected characteristic is tangent to Γ; a related observation is that the boundary data at the point $(0,0)$ is characteristic.

The projected characteristic is free of crossing when

$$x_0 \geq 0, \text{ and } x^1(s) > 0, \quad \text{ or } \quad x_0 \leq 0 \text{ and } x^1(s) < 0,$$

which is equivalent to

$$(x^1, x^2) \in U := \{(x^1, x^2 \in \mathbb{R}^2 : (x^1)^2 \geq x^2, \quad x^1 \neq 0\}.$$

So for $(x^1, x^2) \in U$, the unique smooth solution u is given by

$$u(x^1, x^2) = z(s) = \frac{(x^1)^2 + x^2}{x^1}.$$

2.3. **A remark on existence: Lewy–Nirenberg example for nonexistence (optional).** We will discuss an example of a linear *system* of first-order PDEs, for which even the local existence property fails. For $u : \mathbb{R}^{1+1} \to \mathbb{C}$ and $f : \mathbb{R}^{1+1} \to \mathbb{C}$, consider the equation

$$\partial_t u + it\partial_x u = f(t, x). \tag{2.14}$$

If it were not for the coefficient i in front of t, this equation will be covered by the previous subsection, and many solutions would exist.

Theorem 2.12. *There exists a smooth function f with the property that no(!) C^1 solution to* (2.14) *exists in any neighborhood of* $(0,0)$.

The following argument is from L. Simon's lecture notes [Sim15], and is attributed to L. Nirenberg and H. Lewy.

The proof of Theorem 2.12 involves Complex Analysis, so you will not be required to know the proof for the homework and exams.

Proof. Given $r > 0$, we introduce the notation

$$B_r = \{(t, x) \in \mathbb{R}^{1+1} : t^2 + x^2 < r^2\}, \quad \partial B_r = \{(t, x) \in \mathbb{R}^{1+1} : t^2 + x^2 = r^2\}.$$

Take $f : \mathbb{R}^{1+1} \to \mathbb{C}$ to be any function with the following properties:

- $f(t, x) = f(-t, x)$;
- for some sequences $r_n \searrow 0$ (say, $r_n = 2^{-n}$) and $0 < \delta_n < \frac{1}{2} r_n$, we have

$$f = 0 \quad \text{in } B_{r_n + \delta_n} \setminus B_{r_n - \delta_n}, \quad \text{ while } \quad \int_{B_{r_n}} f \, \mathrm{d}x \neq 0. \tag{2.15}$$

Assume, for the sake of contradiction, that a C^1 solution u to (2.14) exists on B_{r_0} for some $r_0 > 0$ with such an f. Replacing u by $\frac{1}{2}(u(t, x) - u(-t, x))$, we may assume that u is odd with respect to t, i.e., $u(t, x) = -u(-t, x)$. Moreover, fix n sufficiently large so that $r_n < r_0$.

On the one hand, by the divergence theorem and the second property in (2.15), we have

$$\begin{aligned} 0 \neq \int_{B_{r_n}} f \, \mathrm{d}t\mathrm{d}x &= \int_{B_{r_n}} (\partial_t u + it\partial_x u) \, \mathrm{d}t\mathrm{d}x \\ &= \int_{\partial B_{r_n}} \begin{pmatrix} \frac{t}{r_n} \\ \frac{x}{r_n} \end{pmatrix} \cdot \begin{pmatrix} u \\ itu \end{pmatrix} \mathrm{d}s, \end{aligned} \tag{2.16}$$

where $\mathrm{d}s$ refers to the integration with respect to the arc length. In particular, u must be nontrivial on ∂B_{r_n}. On the other hand, we claim that

$$u = 0 \quad \text{on } \partial B_{r_n}, \tag{2.17}$$

which contradicts (2.16).

To prove (2.17), we use a bit of Complex Analysis. Consider the half-plane $\mathbb{H}^+ = \{(t,x) \in \mathbb{R}^{1+1} : t > 0\}$ and its boundary $\partial\mathbb{H}^+ = \{(0,x) \in \mathbb{R}^{1+1}\}$. On the half-ball

$$B_{r_0}^+ = B_{r_0} \cap \mathbb{H}^+,$$

we make the change of variables

$$(t,x) \mapsto (s,y) = (\frac{1}{2}t^2, x).$$

Then

$$(\partial_s + i\partial_y)u(s,y) = \frac{1}{\sqrt{2s}} f(s,y) \quad \text{ in } \{(s,y) \in \mathbb{R}^{1+1} : 2s + x^2 < r_0^2\}.$$

Note that the operator on the LHS is the Cauchy–Riemann operator for the pair $(\operatorname{Re} u, \operatorname{Im} u)$, so u is holomorphic in $s+iy$ in the domain where $f = 0$. In particular, by the first property in (2.15),

$$u \text{ is holomorphic in } s+iy \text{ in } U^+,$$

where

$$U := \{(s,y) \in \mathbb{R}^{1+1} : (r_n - \delta_n)^2 < 2s + x^2 < (r_n + \delta_n)^2\}, \quad U^+ := U \cap \mathbb{H}^+.$$

Moreover, since u is odd in t, it follows that u extends continuously to 0 on $U \cap \partial\mathbb{H}^+ = \{(0,y) \in \mathbb{R}^{1+1} : (r_n - \delta_n)^2 < y^2 < (r_n + \delta_n)^2\}$. By the Schwarz reflection principle, the continuous extension

$$u(s,y) = \begin{cases} u(s,y) & (s,y) \in U \cap \mathbb{H}^+, \\ \overline{u(-s,y)} & (s,y) \in U \cap \{s \leq 0\}, \end{cases}$$

defines a holomorphic function on the reflected domain U. But since $u = 0$ on $U \cap \partial\mathbb{H}^+$, it follows that $u = 0$ on the whole domain U by analytic continuation. The desired statement (2.17) follows by making the inverse change of variables $(s,y) \mapsto (t,x)$. □

Theorem 2.12 shows that we cannot expect an all encompassing local existence (i.e., existence of a locally defined solution) result for even linear systems of first-order PDEs, in the class of smooth solutions.

2.4. **Introduction to Hamilton–Jacobi equation.** In this section we study the *Hamilton–Jacobi equation*, which is an important example of a *fully nonlinear first-order PDE.* The equations reads as follows:

$$\partial_t u + H(x, Du) = 0, \tag{2.18}$$

where $u : (0,\infty)_t \times \mathbb{R}^d_x \to \mathbb{R}$ is the unknown, and $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called the *Hamiltonian.* Here, $Du = (\partial_{x^1} u, \dots, \partial_{x^d} u)$ is the *spatial* part of the differential of u.

2.4.1. *Method of characteristics: local classical solution and its breakdown.* We study the initial value problem

$$\begin{cases} \partial_t u + H(x, Du) = 0 & \text{in } (0,\infty)_t \times \mathbb{R}^d, \\ u = g & \text{on } \{t = 0\}. \end{cases} \tag{2.19}$$

The characteristic equations are:

$$\begin{cases} \dot{x}^j = \partial_{p_j} H(x,p), \\ \dot{p}_j = -\partial_{x^j} H(x,p), \end{cases} \tag{2.20}$$

which are completely decoupled from the equation for z:

$$\dot{z} = \sum_j p_j \partial_{p^j} H. \tag{2.21}$$

Equation (2.20) is a celebrated equation in classical mechanics, namely, *Hamilton's equations* – we will study the deeper reason why this connection arises in what follows. For the moment, we observe that the method of characteristics gives us a unique regular local (in time) solution $u : [0,T) \times \mathbb{R}^d \to \mathbb{R}$ to (2.19) for some $T > 0$. The following example shows that, in general, the regular solution u *cannot* be continued globally (i.e., for all times $t > 0$).

Example 2.13. Consider $d = 1$ and $H(x,p) = \frac{1}{2}p^2$. Then (2.18) becomes

$$\partial_t u + \frac{1}{2}(\partial_x u)^2 = 0.$$

Differentiating this equation, we see that $v(t,x) := \partial_x u(t,x)$ obeys the equation

$$\partial_t v + v \partial_x v = 0,$$

which is the (inviscid) Burgers equation. We now reproduce the well-known proof that, in fact, the solution to the inviscid Burgers equation with *any* initial function $v(t=0) \in C^2(\mathbb{R})$ with a point with negative slope leads to a finite time breakdown (singularity).

Assume that $v \in C^2([0,T) \times \mathbb{R})$ solves the Burgers equation. Note that the characteristic equations are:

$$\dot{t} = 1, \quad \dot{x} = z, \quad \dot{z} = 0, \quad \dot{p}_t = -p_x p_t, \quad \dot{p}_x = -p_x^2.$$

Thanks to $\dot{t} = 1$, we may use t in place of the parameter s for the characteristic curves. By separation of variables, $p_x = \frac{(p_x)_0}{1+t(p_x)_0}$, which blows up in finite time if $(p_x)_0 = \partial_x u(x_0) < 0$. It follows that v fails to be C^1 in finite time if $v(t=0)$ has negative slope at some point.

Returning to the Hamilton–Jacobi equation, we conclude that the unique regular solution u to (2.19) with $d = 1$ and $H = \frac{1}{2}p^2$ necessarily breaks down in finite time if $g''(x_0) < 0$ at some point $x_0 \in \mathbb{R}$.

Despite the possible spontaneous breakdown of the regular, or *classical*, solution u, it is often of interest to find a notion of a *generalized*, or *weak*, *solution* that remains valid afterward and is unique (more ambitiously, leading to well-posedness of the IVP). Unfortunately, there is no known general method for obtaining a satisfactory notion of weak solutions; this task depends highly on the problem at hand.

The goal of this lecture is to exhibit an exemplar case, where we can write down a formula for an appropriate weak solution using ideas from *calculus of variations* or *optimization*, which is an important tool for studying nonlinear PDEs in general. More precisely, we will study the origin of the Hamilton–Jacobi equations in classical mechanics, which would involve some discussion of calculus of variations. Then we will discover that this story provides an inspiration for the formula for an appropriate weak solution, called the *Hopf–Lax formula*, which will be useful later.

(This is, in fact, a part of a bigger story that goes under the name of *viscosity solutions*, but we will have to wait until we pick up adequate tools – maximum principle and theory of parabolic PDEs – to properly develop the theory. We will return to this topic in Math 222B.)

2.4.2. *Review of Lagrangian mechanics: calculus of variations (or optimization).* This is a good place to go back and read about the action principle formulation for Newton's equations in Section 1.4. There, we discussed Lagrange's refomulation of Newton's second law with a conservative force. More generally, consider the following *action functional*:

$$\mathcal{S}[q] = \int_I L(x(t), \dot{x}(t))\,\mathrm{d}t$$

where I is an interval, $x : I \to \mathbb{R}^d$ is the unknown (generalized position variable) and $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called the *Lagrangian*. As stated earlier, the *action principle* in classical mechanics says:

> the path $x(t)$ followed by the particle is that for which the action is minimized, or more precisely, is *stationary*.

The necessary condition for a path $x(t)$ to be a local minimizer of $\mathcal{S}[x]$ is that it is stationary under all compactly supported variations, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{S}[x+\epsilon\varphi]\Big|_{\epsilon=0} = 0 \text{ for all } \varphi \in C^2_c(I).$$

This leads to the *Euler–Lagrange equation* for the functional $\mathcal{S}$:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial v^j}\right)(x,\dot{x}) + \left(\frac{\partial L}{\partial x^j}\right) = 0. \tag{2.22}$$

Indeed, when $L(x,v) = \frac{1}{2}v^2 - V(x)$, we obtain Newton's equations $\ddot{x}^j = -\partial_j V(x)$ for $j = 1, \dots, d$, as we have seen in Section 1.4.

2.4.3. *Lagrangian to Hamilton–Jacobi: dynamical programming.* Hamilton (and also Jacobi) tried to reformulate the d-many second-order ODEs (2.22) in terms of a single scalar first-order PDE. The idea was to consider the following: Given $(t,x) \in \mathbb{R} \times \mathbb{R}^d$, define

$$\mathcal{A}(t,x) := \{\mathbf{x} \in C^2([0,t];\mathbb{R}^d) : \mathbf{x}(t) = x\}.$$

For the purpose of this derivation, let us assume that:

$(*)$ there exists a neighborhood U of $(0,0)$ such that for each $(t,x) \in U$, there exists a unique minimizer $\mathbf{x}_{(t,x)} \in \mathcal{A}(t,x)$ of $\int_0^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s))\,\mathrm{d}s$, which furthermore depends smoothly on (t,x).

In fact, by using the Euler–Lagrange equation and the Fundamental Theorem of ODEs, this can be proved rigorously, but we shall not discuss the proof.

For $(t, x) \in U$, define

$$u(t,x) = \int_0^t L(\mathbf{x}_{(t,x)}(s), \dot{\mathbf{x}}_{(t,x)}(s)) \, \mathrm{d}s.$$

This function is called *Hamilton's principal action.* An alternative way to characterize u is:

$$u(t,x) = \inf_{\mathbf{x} \in \mathcal{A}(t,x)} \left\{ \int_0^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s)) \, \mathrm{d}s \right\}.$$

To proceed, we use note that, for each $0 < h < t$, we must have

$$u(t,x) = \inf_{\mathbf{x} \in \mathcal{A}(t,x)} \left\{ \int_{t-h}^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s)) \, \mathrm{d}s + u(t-h, \mathbf{x}(t-h)) \right\}.$$

or equivalently,

$$0 = \inf_{\mathbf{x} \in \mathcal{A}(t,x)} \left\{ \int_{t-h}^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s)) \, \mathrm{d}s + u(t-h, \mathbf{x}(t-h)) - u(t,x) \right\}. \tag{2.23}$$

This is an example of *dynamical programming*; since $u(t, x)$ is the minimal action from time 0 to t, its restriction to $[t - h, t]$ must also be minimal. For each $\mathbf{x} \in \mathcal{A}(t, x)$, note that

$$\begin{aligned} &\int_{t-h}^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s)) \, \mathrm{d}s + u(t-h, \mathbf{x}(t-h)) - u(t,x) \\ &= h \left[L(x, \dot{\mathbf{x}}(t)) - \partial_t u(t,x) - \sum_j \dot{\mathbf{x}}^j(t) \partial_{x^j} u(t,x) \right] + O(h^2) \end{aligned}$$

by Taylor expansion for u and $\mathbf{x}$. We now claim that:

$$\inf_{v \in \mathbb{R}^d} \left[L(x, v) - \partial_t u(t,x) - \sum_j v^j \partial_{x^j} u(t,x) \right] = 0. \tag{2.24}$$

Indeed, this claim can be motivated by dividing the previous identity by h and taking the infinimum over all $\mathbf{x} \in \mathcal{A}(t, x)$, observing that this expression now depends only on $\dot{\mathbf{x}}(t) =: v$.

Proof of (2.24). First, using the optimal trajectory $\mathbf{x}_{(t,x)}$, we see that

$$L(x, \dot{\mathbf{x}}_{(t,x)}(t)) - \partial_t u(t,x) - \sum_j \dot{\mathbf{x}}^j_{(t,x)}(t) \partial_{x^j} u(t,x) = 0.$$

Thus, (LHS of (2.24)) ≤ 0. Now, for the purpose of contradiction, suppose that it is strictly negative. Then there exists $\eta > 0$ and $v \in \mathbb{R}^d$ such that

$$L(x, v) - \partial_t u(t,x) - \sum_j v^j \partial_{x^j} u(t,x) = -\eta < 0.$$

Consider the trajectory $\mathbf{x}(s) = x - (t - s)v$ in $\mathcal{A}(t, x)$. For any $h > 0$, By Taylor expansion,

$$\int_{t-h}^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s)) \mathrm{d}s + u(t-h, \mathbf{x}(t-h)) - u(t,x)$$

$$= h(L(x,v) - \partial_t u(t,x) - \sum_j v^j \partial_{x^j} u(t,x)) + O(h^2),$$

where the implicit constant is independent of h. By choosing h sufficiently small, we can ensure that

$$\int_{t-h}^{t} L(\mathbf{x}(s), \dot{\mathbf{x}}(s))\mathrm{d}s + u(t-h, \mathbf{x}(t-h)) - u(t,x) \leq -\frac{1}{2}h\eta < 0,$$

which contradicts (2.23). □

Since $\partial_t u$ is independent of v, we can move it to the LHS and conclude that

$$\partial_t u = \inf_{v \in \mathbb{R}^d} \left\{ L(x,v) - \sum_j v^j \partial_j u \right\} =: -H(x, \partial_j u).$$

This completes the derivation of the Hamilton–Jacobi equation under (∗).

Remark 2.14. A similar computation holds if

$$u(t,x) = \min_{\mathbf{x} \in \mathcal{A}(t,x)} \left\{ \int_0^t L(\mathbf{x}(s), \dot{\mathbf{x}}(s))\, \mathrm{d}s + g(\mathbf{x}(0)) \right\}.$$

When L is C^1 in (x,v) and convex in v for each x,

$$H(x,p) = -\inf_{v \in \mathbb{R}^d} \left\{ L(x,v) - \sum_j v^j p_j \right\} = -L(x,v) + v \cdot p \Big|_{v = v(x,p)},$$

where $\partial_v L(x, v(x,p)) = p$ by calculus. By convexity, $v^*(x,p)$ is unique. In the special case $L = \frac{1}{2}|v|^2 - V(x)$, we see that $v(x,p) = p$, so

$$H(x,p) = \frac{1}{2}|p|^2 + V(x).$$

2.4.4. *Duality between Hamiltonian and Lagrangian: the Legendre transform.* The preceding derivation motivates the study of the *Legendre transformation*

$$L^*(p) := \sup_{v \in \mathbb{R}^d} \{-L(x,v) + v \cdot p\}.$$

Recall that when $L = \frac{1}{2}|v|^2 - V(x)$, then $H = L^* = \frac{1}{2}p^2 + V(x)$ (here and below, the variable x is frozen when taking the Legendre transform). Moreover, by the same computation, $H^* = L^{**} = L$! This relation is not a coincidence; in fact, $L^{**} = L$ provided that L is a *convex* function with $\frac{L(v)}{|v|} \to \infty$ as $|v| \to \infty$, and there are good geometric reasons why.

We begin with a review of convex functions.

Definition 2.15. We say that $f : \mathbb{R}^d \to \mathbb{R}$ is *convex* if for every $x \neq y \in \mathbb{R}^d$ and $0 \leq \tau \leq 1$,

$$f(\tau x + (1-\tau)y) \leq \tau f(x) + (1-\tau) f(y).$$

We say that f is *strictly convex* if the inequality is strict for all $x, y \in \mathbb{R}^d$ and $0 < \tau < 1$.

Some important properties of convex functions are:

- *Closure under taking sum or multiplying by a positive number.* Let f, g be convex functions, and $\lambda > 0$. Then λf and $f + g$ are convex.

- *Closure under taking supremum.* The supremum of any collection of convex functions, as long as it is finite, is convex.
- *Continuity.* Any convex function $f : \mathbb{R}^d \to \mathbb{R}$ is continuous[3].
- *Jensen's inequality.* Let $f : \mathbb{R} \to \mathbb{R}$ be convex. For any open bounded subset U and $u \in L^1(U; \mathbb{R})$, we have

$$f\left(\frac{1}{|U|} \int u(x) \, \mathrm{d}x\right) \leq \frac{1}{|U|} \int f(u(x)) \, \mathrm{d}x.$$

- *Supporting hyperplanes.* Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex. Then for each $x \in \mathbb{R}^d$, there exists $p \in \mathbb{R}^d$ such that

$$f(y) \geq f(x) + p \cdot (y - x) \quad \text{ for all } y \in \mathbb{R}^d.$$

  In particular, if f is differentiable at x, then $r = Df(x)$.
- *Existence of a unique minimizer.* If $f : \mathbb{R}^d \to \mathbb{R}$ is strictly convex, then there exists a unique minimizer of f.

Let $L : \mathbb{R}^d \to \mathbb{R}$ be a convex function such that $\lim_{|v| \to \infty} \frac{L(v)}{|v|} = +\infty$, and recall the definition of the Legendre transformation,

$$L^*(p) := \sup_{v \in \mathbb{R}^d} \{p \cdot v - L(v)\}.$$

Observe that the set we are taking the supremum over is nonempty thanks to $\lim_{|v| \to \infty} \frac{L(v)}{|v|} = +\infty$, and finite by the convexity of L. We have seen one interpretation of this transformation, in the special case when L is differentiable and strictly convex:

(0) If L is differentiable and strictly convex, then $L^*(p) = p \cdot v(p) - L(v(p))$ where $v(p)$ is the critical point $DL(v(p)) = p$, which is unique thanks to strict convexity.

We also have the following additional, more geometric, characterizations:

(1) L^* finds (minus) the z-intercept of the hyperplane with normal $(p, 0)$ that touches the graph of L from below in $\mathbb{R}^d_v \times \mathbb{R}_z$.
(2) On the other hand, $L^*(p)$ is the supremum of (trivially) convex functions $v \cdot p - L(x, v)$.

Using these two geometric observation, we see that $L^{**}(v)$ is the supremum of all hyperplanes that lie below the graph of $L(v)$ in $\mathbb{R}^d_v$.

At this point, it is not difficult to show the following result:

Proposition 2.16. *If L is convex and $\lim_{|v| \to \infty} \frac{L(v)}{|v|} = +\infty$, then L^* is convex and $\lim_{|v| \to \infty} \frac{L^*(p)}{|p|} = +\infty$. Moreover, $L = L^{**}$.*

Proof. **Proof of the first assertion.** Indeed, L^* is the supremum of convex functions on $\mathbb{R}^d_p$, so it is convex. Moreover, given any $\lambda > 0$, we have

$$L^*(p) = \sup_{v \in \mathbb{R}^d} \{-L(v) + p \cdot v\} \geq \lambda |p| - L(\lambda \tfrac{p}{|p|}) \geq \lambda |p| - \max_{B_\lambda(0)} L$$

where we have chosen $v = \lambda \frac{p}{|p|}$ in the first inequality. Hence $\liminf_{|p| \to \infty} \frac{L^*(p)}{|p|} \geq \lambda$. Since $\lambda > 0$ was arbitrary, $\lim_{|p| \to \infty} \frac{L^*(p)}{|p|} = \infty$.

[3]This strongly uses the fact that f is finite everywhere.

Proof of the second assertion. Since $L^{**}(v)$ is the supremum of all hyperplanes that lie below the graph of $L(v)$ in $\mathbb{R}^d_v$, clearly $L^{**} \leq L$. On the other hand, given any point (v, z) below the graph of L, we can always find a supporting hyperplane at $(v, L(v))$ which is above (v, z); so L^{**} has the same graph as L (i.e., $L^{**} = L$). □

2.4.5. *From Hamilton–Jacobi to Hamiltonian and Lagrangian mechanics.* With these knowledges, we may now discuss how solving the Hamilton–Jacobi equation leads to an easy integration of the ODEs (2.22). In what follows, we work with L such that, for each $x \in \mathbb{R}^d$, $L(x, \cdot)$ is differentiable, strictly convex and $\lim_{|v|\to\infty} \frac{L(x,v)}{|v|} = +\infty$.

We observe that the projected characteristics necessarily coincide with the minimizing trajectory $x(t)$. Indeed, if x_* is the minimizing trajectory, then

$$u(t, x_*(t)) = \int_0^t L(x_*(s), \dot{x}_*(s))\,\mathrm{d}s$$

$$\partial_t u(t, x_*(t)) + \sum_j \dot{x}_*^j(t)\partial_{x^j} u(t, x_*(t)) = L(x_*(t), \dot{x}_*(t))$$

Let $x = x_*(t)$. Since $L = H^*$,

$$-H(x, Du(t,x)) + \sum_j \dot{x}_*^j(t)\partial_{x^j} u(t,x) = -\min_{p\in\mathbb{R}^d} \{H(x,p) - \dot{x}_*(t)\cdot p\}.$$

We may check that H is strictly convex. Then, by the uniqueness of the minimizer, it follows that

$$\partial_{x^j} u(t,x) = p_j, \quad \dot{x}_*^j = \partial_{p^j} H(x,p),$$

which is the desired conclusion.

(**Exercise**: Prove that $H = L^*$ is strictly convex if L is everywhere differentiable. It is useful to introduce the notion of a *subdifferential* of L: $p \in \partial L(v)$ if and only if there exists a supporting hyperplane with normal p at v. Note that (i) $p \in \partial L(v)$ if and only if $v \in \partial H(p)$, and (ii) L is differentiable at v if and only if $\partial L(v)$ consists of exactly one element.)

To close the full circle, note that the same relations show that the x-component of Hamilton's equations coincide with the minimizing trajectory $x(t)$ for $\mathcal{S}$. We have, in fact, just shown the equivalence of three formulations of classical mechanics (Lagrangian, Hamiltonian, Hamilton–Jacobi)!

2.4.6. *Hopf–Lax formula.* Now, start from $H = H(p)$ that is convex in p and $\frac{H(p)}{|p|} \to \infty$ as $|p| \to \infty$. Define $L(v) = H^*(v)$.

$$u(t,x) = \inf_{\mathbf{x}\in\mathcal{A}(t,x)} \left\{ \int_0^t L(\dot{\mathbf{x}}(s)) + g(\mathbf{x}(0)) \right\}.$$

By convexity,

$$u(t,x) = \min_{y\in\mathbb{R}^d} \left\{ tL(\tfrac{x-y}{t}) + g(y) \right\}. \tag{2.25}$$

This is called the *Hopf–Lax formula.*

It turns out that $u(t,x)$ defined by (2.25) obeys the following nice properties:

- as long as g is Lipschitz, u is a well-defined Lipschitz function such that $u(0,x) = g(x)$;
- at every point (t,x) where u is differentiable, the Hamilton–Jacobi equation is satisfied (in this sense, u may be thought as a weak solution);

- $u(t, x)$ turns out to be *unique* among an adequate class of weak solutions.

All these facts are proved in Section 3.3.3 in [Eva10]. However, with the little tools we have at this point, all these (especially the uniqueness assertion) would require seemingly ad-hoc arguments. We will postpone the proper discussion of this topic until Math 222B, when we talk about *viscosity solutions* to (2.19). If you are interested, you may consult Chapter 10 of [Eva10].

Intermission

The goal of the remainder of the course is to introduce three fundamental tools for studying PDEs: *distribution theory*, *Fourier transform* and *Sobolev spaces*. At the same time, using these tools, we will also study the following basic *second-order linear scalar PDEs*:

- **The Laplace equation.** For $u : \mathbb{R}^d \to \mathbb{R}$ or $\mathbb{C}$,

$$\Delta u = 0.$$

- **The wave equation.** For $u : \mathbb{R}^{1+d} \to \mathbb{R}$ or $\mathbb{C}$,

$$\Box u = (-\partial_t^2 + \Delta) u = 0.$$

- **The heat equation.** For $u : \mathbb{R}^{1+d} \to \mathbb{R}$ or $\mathbb{C}$,

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

- **The Schrödinger equation.** For $u : \mathbb{R}^{1+d} \to \mathbb{C}$,

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

Distribution theory (Section 3) provides a unified and natural framework for studying PDEs. In this theory, the concept of a function is generalized so as to allow for meaningful differentiation of otherwise non-differentiable functions (e.g., think of $f(x) = |x|$). Distribution theory furnishes a natural way to formulate the notion of a generalized solution to a PDE, as well as important concepts for linear PDEs such as the *fundamental solution* and *Green's function* (both will be discussed in more detail below).

The *Fourier transform* (Section 8) is particularly useful for analyzing constant-coefficient linear PDEs like the ones above, since it simultaneously diagonalizes all constant-coefficient linear partial differential operators. As we will see, distribution theory also provides a natural framework for studying Fourier transforms in general as well.

Finally, the last theme that will be pointed out is the miraculous cancellations that happen when we multiply each equation with a suitable function and perform an integration by parts. These cancellations form the basis of the so-called *energy method* (Sections 10 and **??**), which turns out to be the most reliable method to study *variable-coefficient* and/or *nonlinear* PDEs. The desire to develop wellposedness theory based on the energy method will motivate us to study the theory of *Sobolev spaces* (Section 11).

3. Introduction to the theory of distributions

Distribution theory allows us to meaningfully differentiate functions that are not classically differentiable. In a sense, it is a completion of differential calculus. The theory was pioneered by L. Schwartz in the mid-20th century, but actually the related ideas have already been used by physicists and engineers.

3.1. An example from electrostatics.

To motivate the theory, let us discuss an example from physics, or more specifically, electrostatics.

The subject of electrostatics deals with the relationship between an electric field $\mathbf{E} : U \to \mathbb{R}^3$ and an electric charge distribution $\rho : U \to \mathbb{R}$ (i.e., $\int_V \rho \, \mathrm{d}x$ gives the total electric charge inside the domain V) in a domain $U \subseteq \mathbb{R}^3$ when nothing varies in time. The main equations of the theory are:

- **The Gauss law**
$$\nabla \cdot \mathbf{E} = \rho \quad \text{in } U,$$
- **The electrostatic law**
$$\nabla \times \mathbf{E} = 0 \quad \text{in } U.$$

If the electrostatic law is satisfied on $U = \mathbb{R}^3$, then as we learn in vector calculus, there exists an *electric potential* $\phi : \mathbb{R}^3 \to \mathbb{R}$ such that $\mathbf{E} = -\nabla\phi$. The Gauss law for the electric potential becomes the Poisson equation:
$$-\Delta\phi = \rho \quad \text{in } \mathbb{R}^3.$$

A basic problem in electrostatics is to determine the electric potential ϕ from a given charge distribution ρ. Here is the physicist's way of solving this problem:

- When the charge distribution ρ consists of a finite sum of point charges q_k placed at $y_k \in \mathbb{R}^3$ for $k \in \{1, \dots, K\}$, i.e., $\rho = \sum_{k=1}^{K} q_k$(point charge at y_k) then by linearity, ϕ must be given by the sum of the electric potentials of the point charges. Again by linearity and translation invariance, the electric potential of the point charge q_k at y_k is $q_k \phi_0(x - y_k)$, where ϕ_0 is the electric potential of the (positive) unit point charges at the origin. Therefore,
$$\phi = \sum_{k=1}^{K} q_k \phi_0(x - y_k).$$
- Next, consider a general a smooth charge distribution ρ. Let us view it as the "continuous sum of point charges $\rho(y)$ at $y \in \mathbb{R}^3$". By linearity, ϕ should be "continuous sum of the electric potentials $\rho(y)\phi_0(x-y)$ of these point charges". In other words,
$$\phi(x) = \int \rho(y)\phi_0(x - y) \, \mathrm{d}y.$$
- Now it only remains to determine the electric potential ϕ_0 of a (positive) unit point charge at 0. By the rotational symmetry of the problem, ϕ_0 must be radial, i.e., $\phi = \phi(r)$ in the polar coordinates on $\mathbb{R}^3$ ($r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$). For any $r > 0$, the total amount of charge inside the ball $B(0, r)$ is 1, since there is only the unit point charge at 0. By the Gauss law and the divergence theorem,
$$1 = \int_{B(0,r)} \nabla \cdot \mathbf{E} = \int_{\partial B(0,r)} \nu \cdot \mathbf{E}.$$

Since $\nu(x) = \frac{x}{|x|}$ for $x \in \partial B(0,r)$, $\nu \cdot \nabla\phi = \partial_r \phi$ in the polar coordinates. Moreover, since ϕ is radial, $\partial_r\phi$ is constant on $\partial B(0,r)$. Therefore,

$$1 = -\int_{\partial B(0,r)} \nu \cdot \nabla\phi = -4\pi r^2 \partial_r \phi(r),$$

or in other words,

$$\partial_r\phi(r) = -\frac{1}{4\pi r^2}.$$

(Actually, this is Coulomb's inverse square law!) For physical reasons, it is reasonable to normalize ϕ so that $\phi(r) \to 0$ as $r \to \infty$. Then by integration,

$$\phi(r) = \frac{1}{4\pi}\frac{1}{r}, \quad \text{or equivalently, } \phi(x) = \frac{1}{4\pi}\frac{1}{|x|}.$$

– In sum, the electric potential on ϕ in $\mathbb{R}^3$ corresponding to a given charge distribution ρ in $\mathbb{R}^3$ is

$$\phi(x) = \frac{1}{4\pi}\int \frac{1}{|x-y|}\rho(y)\,\mathrm{d}y.$$

This clever procedure has a few jumps that are difficult to justify with usual calculus, such as the notion of a point charge, representation of $\rho(x)$ as the "continuous sum of point charges" and the electrostatic potential of a point charge. However, we do not want to give up on these nice ideas! As a natural setting in which the above argument can be made rigorous (and also generalized), we will introduce the concept of a *distribution*.

3.2. **Definition of a distribution, first take.** The basic idea of a distribution is as follows. Consider a continuous function $u : U \to \mathbb{R}$. The most obvious way to characterize u is by its pointwise values $u(x)$, i.e., two continuous functions u and v on U are the same if and only if $u(x) = v(x)$ for all $x \in U$. Equivalently, we may also characterize f in terms of the weighted averages

$$\langle u, \phi\rangle = \int u\phi\,\mathrm{d}x,$$

where the weight ϕ varies over a vector space of functions on U that is

(1) "nice enough" so that $\langle u, \phi\rangle$ is well-defined for each ϕ; and
(2) "rich enough" so that, for instance, $\langle u, \phi\rangle = \langle v, \phi\rangle$ for all ϕ in this space if and only if $u = v$.

Observe that $\phi \mapsto \langle u, \phi\rangle$ is a linear functional on this vector space. The weights ϕ are also called *test functions*.

The power of this viewpoint lies in the simple observation that, in fact, if the test functions are "nice enough", then the local weighted average $\langle u, \phi\rangle$ makes sense for a much larger class of objects than just the continuous functions. For instance, if ϕ's are assumed to be continuous, then $\langle u, \phi\rangle$ naturally makes sense for any *measure* u that vanishes outside a compact subset K of U.

We arrive at the notion of a distribution u by following the above idea to an extreme: Throw away the pointwise values of u and keep only the linear functionals $\phi \mapsto u(\phi)$ defined on a suitable vector space of test functions. It is a generalization of the notion of a continuous function in the sense that, when u is a continuous function, $u(\phi)$ is given by the weighted average $u(\phi) = \langle u, \phi\rangle = \int u\phi\,\mathrm{d}x$.

What is then the suitable vector space of ϕ's? First, it would be nice if each ϕ is infinitely differentiable, or smooth. Second, it would also be nice if each ϕ vanishes outside a compact set. For later purposes, it is useful to formalize this point with the following definition:

Definition 3.1 (Support of a continuous function)**.** Let $f \in C(U)$. The *support of* f, denoted by $\operatorname{supp} f$ is the closure of the subset of U where f is non-zero, i.e.,

$$\operatorname{supp} f = \overline{\{x \in U : f(x) \neq 0\}}.$$

We say that f is *compactly supported* if $\operatorname{supp} f$ is compact.

The space of all smooth (i.e., infinitely differentiable) functions $\phi : U \to \mathbb{R}$ that are compactly supported (i.e., $\operatorname{supp} \phi$ is compact) is denoted by $C_c^\infty(U)$. Using $C_c^\infty(U)$ as the space of test functions, we (almost) arrive at the standard definition of a distribution:

Definition 3.2 (Distribution; rough version)**.** A distribution u in U is a linear functional $u : C_c^\infty(U) \to \mathbb{R}$ that is "continuous."

The continuity condition is a natural regularity condition to impose on u to avoid crazy counterexamples. In order to make precise the notion of continuity in Definition 3.2, we need to discuss the notion of convergence (i.e., topology) of the space $C_c^\infty(U)$; this is one of the topics of the next subsection.

3.3. **The space of test functions $C_c^\infty(U)$.** To properly formulate the notion of continuity in Definition 3.2, we now turn to the description of the space $C_c^\infty(U)$ of smooth and compactly supported functions in U. We have chosen this space so that its elements are "nice". According to the discussion in the previous subsection, in order for Definition 3.2 to be reasonable, we would like to demonstrate that the space $C_c^\infty(U)$ is "rich enough" as well; for instance, if u and v are continuous functions on U such that $\langle u, \phi \rangle = \langle v, \phi \rangle$ for all $\phi \in C_c^\infty(U)$, then we better have $u = v$.

We start with the simplest question, namely, *can we construct a single example of a smooth function on $\mathbb{R}^d$ whose support belongs to $\overline{B(0,1)}$?* Here is one way to construct such an example. Consider the function $\varphi : \mathbb{R} \to \mathbb{R}$ defined by

$$\varphi(x) = \begin{cases} e^{-\frac{1}{x}} & x > 0 \\ 0 & x \leq 0. \end{cases}$$

It is not difficult to check that φ is infinitely differentiable and $\operatorname{supp} \varphi = [0, \infty)$. Next, consider

$$\phi(x) = \varphi\left(1 - \left((x^1)^2 + \cdots + (x^d)^2\right)\right).$$

Since ϕ is the composition of two smooth functions, $\phi \in C^\infty(\mathbb{R}^d)$. Moreover, since $1 - \left((x^1)^2 + \cdots + (x^d)^2\right) \geq 0$ if and only if $x \in \overline{B(0,1)}$, it follows that with $\operatorname{supp} \phi \subset \overline{B(0,1)}$.

To generate more examples, let us introduce the idea of *convolution*:

Definition 3.3 (Convolution)**.** Let f be a continuous function in U and $\phi \in C_c^\infty(U)$. We define the convolution of f and ϕ by

$$f * \phi(x) = \int f(y)\phi(x-y)\,\mathrm{d}y.$$

As we will see below, the convolution operation can be generalized to more general f and ϕ.

Let us quickly go over a few important properties of convolution. First, note that $*$ is commutative:

$$f * \phi(x) = \int f(x-y)\phi(y)\,\mathrm{d}y = \int f(y)\phi(x-y)\,\mathrm{d}y = g * f(x).$$

Next, even if f is merely continuous (so in particular, non-differentiable), note that its convolution $f*\phi$ with $\phi \in C_c^\infty(\mathbb{R}^d)$ is smooth. Indeed, by the formal chain of identities

$$\partial_{x^j}(f*\phi)(x) = \partial_{x^j}\int f(y)\phi(x-y)\,\mathrm{d}y = \int \partial_{x^j}(f(y)\phi(x-y))\,\mathrm{d}y = \int f(y)\partial_{x^j}\phi(x-y)\,\mathrm{d}y,$$

we see that, provided that the order of the differentiation and the integration can be interchanged (the second equality), we may arrange so that each x-derivative of $f*\phi(x)$ falls only on $\phi(x-\cdot)$, which still belongs to $C_c^\infty(U)$. The hypothesis that f is continuous and $\phi \in C_c^\infty(\mathbb{R}^d)$ is sufficient to justify this interchange.

Another important property of convolution is:

$$\operatorname{supp} f*\phi \subseteq \operatorname{supp} f + \operatorname{supp}\phi,$$

where by $A+B$ for two subsets $A, B \subseteq \mathbb{R}^d$, we mean

$$A + B = \{a+b \in \mathbb{R}^d : a \in A,\ b \in B\}.$$

In particular, if $\operatorname{supp} f$ is compact, then $f*\phi$ is also compactly supported. To see how this inclusion is proved, take a point $x \in \mathbb{R}^d$ such that $f*\phi(x) \neq 0$. Then the integral

$$f*\phi(x) = \int f(y)\phi(x-y)\,\mathrm{d}y$$

must be non-zero, which means that there exists some $y \in \{y : f(y) \neq 0\} \subseteq \operatorname{supp} f$ and $x - y \in \{z : \phi(z) \neq 0\} \subseteq \operatorname{supp}\phi$. Thus, $\{x : f*\phi(x) \neq 0\} \subseteq \operatorname{supp} f + \operatorname{supp}\phi$. Taking the closure of both sides, we arrive at the desired conclusion.

Lemma 3.4. *Let $f \in C^k(\mathbb{R}^d)$, $0 \leq k < \infty$. Let ϕ be a smooth function with support contained in $\overline{B(0,1)}$ and $\int \phi = 1$. Set $\phi_\delta(x) = \delta^{-d}\phi(\delta^{-1}x)$ and let*

$$f_\delta(x) = \phi_\delta * f(x) = \int \delta^{-d}\phi\left(\frac{x-y}{\delta}\right) f(y)\,\mathrm{d}y = \int f(x-\delta z)\phi(z)\,\mathrm{d}z.$$

Then

- *The functions f_δ are C^∞ and $\operatorname{supp} f_\delta \subseteq \operatorname{supp} f + B(0,\delta)$.*
- *For $|\alpha| \leq k$, we have $\partial^\alpha f_\delta \to \partial^\alpha f$ uniformly on each compact set as $\delta \to 0$.*

Proof. By definition, the regularity and support properties of f_δ in the statement of Lemma 3.4 are clear. The key is to prove the uniform convergence assertion.

Let us first consider the case $k = 0$. Let L be a compact subset of U. We write

$$\begin{aligned}\phi_\delta * f(x) - f(x) &= \int \phi_\delta(y) f(x-y)\,\mathrm{d}y - \int \phi_\delta(y) f(x)\,\mathrm{d}y \\ &= \int \phi(z)\left(f(x-\delta z) - f(x)\right)\mathrm{d}z,\end{aligned}$$

where in the first equality, we used the property that $\int \phi_\delta = 1$, and in the second equality, we used the change of variables $y = \delta z$. For each fixed z, $f(x-\delta z) - f(x) \to 0$ as $\delta \to 0$ by the continuity of f. Moreover, for $z \in \operatorname{supp}\phi$, which is compact, and x in the compact set L, $f(x-\delta z)-f(x)\to 0$ as $\delta\to 0$ uniformly. Thus we can exchange the order of the limit $\lim_{\delta\to 0}$ and the integration, and conclude that $\phi_\delta * f(x)\to f(x)$ uniformly on L, as desired.

In the case $0<|\alpha|\leq k$, we begin by writing

$$\begin{aligned} D^\alpha(\phi_\delta * f)(x) - D^\alpha f(x) &= \int \phi_\delta(y) D^\alpha f(x-y)\,\mathrm{d}y - \int \phi_\delta(y) D^\alpha f(x)\,\mathrm{d}y \\ &= \int \phi(z)\left(D^\alpha f(x-\delta z) - D^\alpha f(x)\right)\mathrm{d}z, \end{aligned}$$

where the point is that we let up to k derivatives fall on f. At this point, we may adapt the argument in the case $k=0$ and prove $D^\alpha(\phi_\delta * f)(x) - D^\alpha f(x)\to 0$ uniformly on any compact subset L of U; we omit the straightforward details. □

As a consequence, we have a huge space of test functions, which closely follows the behavior of continuous compactly supported functions. The space of continuous compactly supported functions is very rich, as we learned in point-set topology (recall Urysohn's lemma, etc.). From Lemma 3.4 and "richness" of $C(U)$, it is not difficult to show $C^\infty_c(U)$ is "rich enough" so that, in particular, if u and v are continuous functions on U such that $\langle u,\phi\rangle = \langle v,\phi\rangle$ for all $\phi\in C^\infty_c(U)$, then $u=v$.

We are finally ready to specify the topology of $C^\infty_c(U)$. Instead of an abstract description, let us describe how *sequential convergence* is defined:

Definition 3.5. A sequence $\phi_j\in C^\infty_c(U)$ converges to $\phi\in C^\infty_c(U)$ if there exists a compact set $K\subset U$ such that $\operatorname{supp}\phi_j, \operatorname{supp}\phi\subseteq K$ and

$$\lim_{j\to\infty}\sup_{x\in K}|D^\alpha\phi_j(x) - D^\alpha\phi(x)| = 0$$

for every multi-index α.

Remark 3.6 (For those who are familiar with functional analysis). The topology on $C^\infty_c(U)$ is the strongest (i.e., smallest) topology such that for each compact subset $K\subset U$, the space

$$C^\infty_c(K) = \{\phi : \phi\in C^\infty, \operatorname{supp}\phi\subseteq K\}$$

equipped with the complete invariant (i.e., $d(\phi-\varphi,\psi-\varphi) = d(\phi,\psi)$) metric

$$d(\phi,\psi) = \sum_{n=0}^{\infty} 2^{-n}\frac{p_n(\phi-\psi)}{1+p_n(\phi-\psi)},\quad p_n(\phi) = \sup_{\alpha:|\alpha|=n}\sup_{x\in K}|D^\alpha\phi(x)|,$$

(in fact, $(C^\infty_c(K), d)$ is a Fréchet space) embeds continuously into $C^\infty_c(U)$.

3.4. Definition of a distribution, second take, and some basic concepts.

We are now ready to give the precise definition of a distribution, following L. Schwartz.

Definition 3.7. A *distribution* u on U is a linear functional $u : C^\infty_c(U)\to\mathbb{R}$ that is *continuous* in the following sense: For any sequence $\phi_j\in C^\infty_c(U)$ such that $\phi_j\to\phi\in C^\infty_c(U)$ in the sense of Definition 3.5, then $u(\phi_j)\to u(\phi)$.

It is customary to write $\mathcal{D}'(U)$ for the space of distributions on U. Also, the pairing $u(\phi)$ of the linear functional u and a test function ϕ is usually written in the form

$$u(\phi) = \langle u,\phi\rangle,$$

motivated by the notation in the case u is continuous.

A useful reformulation of Definition 3.7 that does not directly involve Definition 3.5 (but of course, they are deeply related!) is as follows:

Lemma 3.8. *A linear functional $u : C_c^\infty(U) \to \mathbb{R}$ is a* distribution *if and only if it is* bounded *in the following sense: For any compact set $K \subset U$, there exists an integer N and a constant $C = C_{K,N}$ such that for all $\phi \in C_c^\infty(U)$ with* $\operatorname{supp}\phi \subseteq K$, *we have*

$$|\langle u, \phi\rangle| \leq C \sum_{|\alpha|\leq N} \sup_{x\in K} |D^\alpha \phi(x)|. \tag{3.1}$$

Proof. That boundedness implies (sequential) continuity is obvious. To show the converse, we argue by contradiction.

Suppose that $u : C_c^\infty(U) \to \mathbb{R}$ is a distribution, but not bounded. Then there exists a compact set $K \subset U$ such that for all integers N and $C > 0$, there exists $\phi_{N,C} \in C_c^\infty(U)$ such that

$$|\langle u, \phi_{N,C}\rangle| \geq C \sum_{|\alpha|\leq N} \sup_{x\in K} |D^\alpha \phi_{N,C}(x)|.$$

Choosing $C = N$, we arrive at a sequence $\phi_N = \phi_{N,N}$ obeying

$$|\langle u, \phi_N\rangle| \geq N \sum_{|\alpha|\leq N} \sup_{x\in K} |D^\alpha \phi_N(x)|. \tag{3.2}$$

Now put

$$\psi_N(x) := \frac{1}{N \sum_{|\alpha|\leq N} \sup_{y\in K} |D^\alpha \phi_N(y)|} \phi_N(x).$$

Then $\operatorname{supp}\psi_N \subseteq K$ and $|D^\beta \psi_N(x)|_N \leq \frac{1}{N}$ for $x \in K$ and $|\beta| \leq N$. By Definition 3.5, we see that $\psi_N \to 0$ in $C_c^\infty(U)$. Since u is a distribution, $\langle u, \psi_N\rangle \to 0$, but this property contradicts (3.2) that implies instead $|\langle u, \psi_N\rangle| \geq 1$. □

Using Lemma 3.8, we can introduce the concept of the *order* of a distribution.

Definition 3.9 (Order of a distribution)**.** Let $u \in \mathcal{D}'(U)$. If there is an integer N such that (3.1) holds for all compact set $K \subset U$ and some $C = C_K$, then we say that u *has order* $\leq N$. The smallest such N is called the *order* of the distribution u.

Another useful concept to keep in mind is that of the *support* of a distribution.

Definition 3.10 (Support of a distribution)**.** We say that a distribution $u \in \mathcal{D}'(U)$ *vanishes in an open subset* $V \subseteq U$ if $\langle u, \phi\rangle = 0$ for every test function ϕ such that $\operatorname{supp}\phi \subset V$. Let

$$V_{\max} = \bigcup \{V : V \text{ is an open subset of } U \text{ in which } u \text{ vanishes}\}.$$

The *support of* u, denoted by $\operatorname{supp} u$, is defined as the complement of $V_{\max}$, i.e.,

$$\operatorname{supp} u = U \setminus \left(\bigcup \{V : V \text{ is an open subset of } U \text{ in which } u \text{ vanishes}\}\right).$$

At this point, let us cover some first examples of distributions.

- *Locally integrable functions.* We say that a function $u : U \to \mathbb{R}$ is *locally integrable* if it is measurable and absolutely integrable on every compact subset K of U with respect to the Lebesgue measure (i.e., $\int_K |f| < \infty$); we denote by $L^1_{loc}(U)$ the space of such functions. Any locally integrable function u defines a distribution by

$$\langle u, \phi \rangle := \int u\phi \, \mathrm{d}x,$$

where the RHS is well-defined thanks to local integrability. It is usual to abuse the notation and use the same letter u to refer to the distribution defined by the function u. Clearly, all distributions arising in this fashion has order 0. When u is continuous, the support of u as a continuous function agrees with the support of u as a distribution.

To give more specific examples, any continuous function or any essentially bounded measurable function is a distribution. A singular function on $\mathbb{R}^d$ of the form

$$u(x) = \frac{1}{|x|^\alpha}$$

with $\alpha < d$ is locally integrable, so it is a distribution. However, when $\alpha \geq d$, it does *not*

- *(Signed) Borel measures.* Any signed Borel measure μ on U defined a distribution by

  $$\langle \mu, \phi \rangle = \int \phi(x) \, \mathrm{d}\mu(x).$$

  Again, we use the same letter μ to denote the distribution defined by μ. All distributions arising in this fashion again has order 0[4]. The support of μ as a distribution coincides with the support of μ as a signed Borel measure.

  An important example of this type of a distribution is the *delta distribution* (or more colloquially, the delta function) at y, which is defined by

  $$\langle \delta_y, \phi \rangle = \phi(y).$$

  It corresponds to the atomic measure with total measure 1 at $y \in \mathbb{R}^d$. From this example, we can make sense of the charge distribution ρ of point charges q_i at y_i:

  $$\rho = \sum_i q_i \delta_{y_i}.$$

- *Higher order examples.* The simplest example of a N-th order distribution on $\mathbb{R}^d$ is

  $$\langle u, \phi \rangle = \partial^\alpha \phi(0)$$

  where α is a multi-index of order $|\alpha| = N$. The support of this distribution is $\{0\}$.

More interesting examples of distributions will be given after we discuss the basic operations and limit theorems for distributions.

[4]In fact, although we will not prove it, it turns out that every distribution of order zero on U is a continuous linear functional on $C_0(U)$, which may be identified with a signed Borel measure on U.

3.5. **Basic operations for distributions.** We now discuss how to generalize basic operations for smooth functions to the case of distributions. As we will see, the basic idea is as follows:

Basic principle (the adjoint method): An operation $\mathcal{A}$ on smooth functions are generalized to distributions by computing the adjoint operation $\mathcal{A}'$ defined by

$$\int_U (\mathcal{A}u)\phi \,\mathrm{d}x = \int_U u(\mathcal{A}'\phi)\,\mathrm{d}x \quad \forall \phi \in C_c^\infty(U), \forall u \in C^\infty(U),$$

such that $\mathcal{A}'\phi \in C_c^\infty(U)$, then defining

$$\langle \mathcal{A}u, \phi\rangle := \langle u, \mathcal{A}'\phi\rangle \quad \forall \phi \in C_c^\infty(U), \forall u \in \mathcal{D}'(U).$$

- *Multiplication by smooth function.* Given $u \in \mathcal{D}'(U)$ and $f \in C^\infty(U)$, we define

  $$\langle fu, \phi\rangle := \langle u, f\phi\rangle, \quad \forall \phi \in C_c^\infty(U).$$

  Indeed, this definition is motivated by the fact that for $u \in C^\infty(U)$ and $\phi \in C_c^\infty(U)$,

  $$\langle fu, \phi\rangle = \int_U (fu)\phi = \int_U u(f\phi)$$

  and the observation that $f\phi \in C_c^\infty(U)$. Moreover, it is not difficult to check that $\phi_n \to \phi$ in $C_c^\infty(U)$ implies $f\phi_n \to f\phi$ in $C_c^\infty(U)$, which makes $\phi \mapsto \langle fu, \phi\rangle$ indeed a distribution.
- *Differentiation.* Given $u \in \mathcal{D}'(U)$, we define

  $$\langle \partial_j u, \phi\rangle := -\langle u, \partial_j \phi\rangle, \quad \forall \phi \in C_c^\infty(U).$$

  Indeed, for $u \in C^\infty(U)$ and $\phi \in C_c^\infty(U)$, we have

  $$\langle \partial_j u, \phi\rangle = \int_U \partial_j u\phi = -\int_U u\partial_j \phi$$

  by integration by parts. Note moreover that $\phi_n \to \phi$ in $C_c^\infty(U)$ implies $\partial_j \phi_n \to \partial_j \phi$ in $C_c^\infty(U)$, so that $\phi \mapsto \langle \partial_j u, \phi\rangle$ is indeed a distribution.

  It is worth emphasizing that *every distribution is differentiable* with the above simple definition. So distribution theory is an extension of the usual differential calculus, where *every object (including any continuous functions, which are distributions) is differentiable.* Distribution theory is the minimal among such extensions, in the sense that every distribution is locally a (in general, high order) derivative of a continuous function; this is the *structure theorem for distributions*, which we will cover later.
- *Convolution with a smooth compactly supported function.* Given $u \in \mathcal{D}'(\mathbb{R}^d)$ and $f \in C_c^\infty(\mathbb{R}^d)$, we define

  $$\langle f * u, \phi\rangle := \langle u, f *' \phi\rangle \quad \forall \phi \in C_c^\infty(U),$$

  where $*'$ is the "adjoint" convolution defined by

  $$f *' \phi(x) = \int_{\mathbb{R}^d} f(y-x)\phi(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f(y)\phi(x+y)\,\mathrm{d}y.$$

  Indeed, for $u \in C^\infty$, note that

  $$\begin{aligned}\langle f * u, \phi\rangle &= \int_{\mathbb{R}^d} (f*u)(x)\phi(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x-y)u(y)\phi(x)\,\mathrm{d}y\mathrm{d}x\end{aligned}$$

$$= \int_{\mathbb{R}^d} u(y) \left(\int_{\mathbb{R}^d} f(x-y)\phi(x) \, \mathrm{d}x \right) \mathrm{d}y$$
$$= \langle u, f *' \phi \rangle.$$

It can be checked in a straightforward manner (although it is more complicated than the preceding two operations) that $\phi \mapsto \langle f * u, \phi \rangle$ is continuous.

In fact, by the condition $f \in C_c^\infty(\mathbb{R}^d)$, $f * u$ is more than merely a distribution.

Lemma 3.11. *For $f \in C^\infty(\mathbb{R}^d)$ and $u \in \mathcal{D}'(\mathbb{R}^d)$, $f * u$ is a smooth function (i.e., $f * u \in C^\infty(\mathbb{R}^d)$). In fact,*

$$D^\alpha (f * u)(x) = ((D^\alpha f) * u)(x)$$

for any multi-index α.

Proof. Let us first check that $f * u$ is a continuous function. Note that the expression

$$f * u(x) = \int_{\mathbb{R}^d} f(x-y) u(y) \, \mathrm{d}y = \langle u, f(x - \cdot) \rangle$$

already makes sense for $u \in \mathcal{D}'(\mathbb{R}^d)$, since $f(x - \cdot) \in C_c^\infty(\mathbb{R}^d)$. By the continuity property of u in Definition 3.7, it follows that $f * u(x)$ is continuous. Moreover, it agrees with the distribution $f * u$ as defined above.

To see that $f * u$ is smooth, the idea is to work with difference quotients. Let us demonstrate the idea in detail in the case $|\alpha| = 1$; the rest of the proof is a straightforward extension of this case.

Let e_i be the unit vector in the positive direction along the x^i-axis, and consider the directional difference quotient

$$\frac{1}{h} \left(f * u(x + h e_i) - f * u(x) \right) = \left\langle u, \frac{1}{h} \left(f(x + h e_i - \cdot) - f(x - \cdot) \right) \right\rangle.$$

It is not difficult to check that $\frac{1}{h} \left(f(x + h e_i - \cdot) - f(x - \cdot) \right) \to \partial_i f(x - \cdot)$ in the $C_c^\infty(\mathbb{R}^d)$ topology. Thus, by the continuity property of u in Definition 3.7, it follows that

$$\begin{aligned} \partial_i (f * u)(x) &= \lim_{h \to 0} \frac{1}{h} \left(f * u(x + h e_i) - f * u(x) \right) \\ &= \lim_{h \to 0} \left\langle u, \frac{1}{h} \left(f(x + h e_i - \cdot) - f(x - \cdot) \right) \right\rangle \\ &= \langle u, \partial_i f(x - \cdot) \rangle = ((\partial_i f) * u)(x). \end{aligned}$$

This proves that $f * u$ is differentiable and $D^\alpha (f * u)(x) = ((D^\alpha f) * u)(x)$ when $|\alpha| = 1$, as desired. □

An analogue of Lemma 3.4 holds in this case as well; let us hold off the discussion of this result until we introduce the notion of convergence of distributions.

The property of the convolution regarding the supports remains valid in this case, too.

Lemma 3.12. *For $f \in C^\infty(\mathbb{R}^d)$ and $u \in \mathcal{D}'(\mathbb{R}^d)$, we have*

$$\mathrm{supp}(f * u) \subseteq \mathrm{supp}\, f + \mathrm{supp}\, u.$$

Proof. To prove this, it suffices to prove the contrapositive: If $z \notin \mathrm{supp}\, f + \mathrm{supp}\, u$, then $z \notin \mathrm{supp}(f * u)$. Equivalently, we need to show that if $z - \mathrm{supp}\, f = \{z - x : x \in \mathrm{supp}\, f\} \subseteq \mathbb{R}^d \setminus \mathrm{supp}\, u$, then $f * u(z) = 0$. Since $f * u = \langle u, f(x - \cdot) \rangle$ by Lemma 3.11, the last statement is obvious. □

3.6. **More examples.** Here are more examples of distributions on $\mathbb{R}$.

- *Heaviside function.* Consider

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

Since H is locally integrable, it defines a distribution.

Note the following computation:

$$H' = \delta_0,$$

where δ_0 is the delta distribution at 0.

To see this, we recall the definition of differentiation of a distribution and compute as follows: For $\phi \in C_c^\infty(\mathbb{R})$,

$$\begin{aligned} \langle H', \phi \rangle = -\langle H, \phi' \rangle &= -\int H(x)\phi'(x)\,\mathrm{d}x \\ &= -\int_0^\infty \phi'(x)\,\mathrm{d}x \\ &= \phi(0) - \lim_{x\to\infty} \phi(x) = \phi(0), \end{aligned}$$

where we used the fundamental theorem of calculus on the second line, and the compact support property of ϕ on the last line.

Remark 3.13. This simple computation will be generalized below when we compute the derivative of the characteristic function of a C^1 domain, which in turn will lead to the distribution-theoretic proof of the divergence theorem.

- *Principal value distribution.* We start with the following question: Although $\frac{1}{x}$ is not locally integrable (so it does not directly define a distribution), can we find a distribution $u \in \mathcal{D}'(\mathbb{R})$ that agrees with $\frac{1}{x}$ (i.e., $u - \frac{1}{x}$ vanishes) in the open set $\mathbb{R} \setminus \{0\}$?

There are many ways to go about this question, but a good idea is to notice that the indefinite integral of $\frac{1}{x}$, namely $\log|x|$, is locally integrable so that it defines a distribution. Then we define the desired distribution u by

$$u = \frac{\mathrm{d}}{\mathrm{d}x} \log|x|,$$

where the differentiation is taken in the sense of distributions. Such a distribution is called the *principal value distribution*, and is denoted by pv $\frac{1}{x}$. It can be shown that

$$\langle \mathrm{pv}\left(\frac{1}{x}\right), \phi \rangle = \lim_{\epsilon\to 0+} \left(\int_\epsilon^\infty \frac{1}{x}\phi(x) + \int_{-\infty}^{-\epsilon} \frac{1}{x}\phi(x)\,\mathrm{d}x \right).$$

Indeed, consider a test function $\phi \in C_c^\infty(\mathbb{R})$ and choose L large enough so that $\operatorname{supp}\phi \subset (-L, L)$. We compute

$$\begin{aligned} \langle pv\left(\frac{1}{x}\right), \phi \rangle &= -\int_{-\infty}^\infty \log|x|\phi'(x)\,\mathrm{d}x \\ &= -\int_0^L \log|x|(\phi(x) - \phi(0))'\,\mathrm{d}x - \int_{-L}^0 \log|x|(\phi(x) - \phi(0))'\,\mathrm{d}x \\ &= \int_{-L}^L \frac{1}{x}(\phi(x) - \phi(0))\,\mathrm{d}x. \end{aligned}$$

Splitting $(-L, L)$ into $(-L, -\epsilon) \cup (-\epsilon, \epsilon) \cup (\epsilon, L)$, we see that

$$\begin{aligned}
&\int_{-L}^{L} \frac{1}{x}(\phi(x) - \phi(0)) \, \mathrm{d}x \\
&= \lim_{\epsilon \to 0+} \left(\int_{\epsilon}^{L} \frac{1}{x}(\phi(x) - \phi(0)) \, \mathrm{d}x + \int_{-\epsilon}^{\epsilon} \frac{1}{x}(\phi(x) - \phi(0)) \, \mathrm{d}x + \int_{-L}^{-\epsilon} \frac{1}{x}(\phi(x) - \phi(0)) \, \mathrm{d}x \right) \\
&= \lim_{\epsilon \to 0+} \left(\int_{\epsilon}^{L} \frac{1}{x}(\phi(x) - \phi(0)) \, \mathrm{d}x + \int_{-\epsilon}^{\epsilon} \int_{0}^{1} \phi'(\sigma x) \, \mathrm{d}\sigma \, \mathrm{d}x + \int_{-L}^{-\epsilon} \frac{1}{x}(\phi(x) - \phi(0)) \, \mathrm{d}x \right) \\
&= \lim_{\epsilon \to 0+} \left(\int_{\epsilon}^{L} \frac{1}{x}\phi(x) \, \mathrm{d}x + \int_{-L}^{-\epsilon} \frac{1}{x}\phi(x) \, \mathrm{d}x \right),
\end{aligned}$$

where on the last line, we used

$$\left| \int_{-\epsilon}^{\epsilon} \int_{0}^{1} \phi'(\sigma x) \, \mathrm{d}\sigma \, \mathrm{d}x \right| \leq 2 \sup_{x} |\phi'(x)| \epsilon \to 0 \text{ as } \epsilon \to 0,$$

$$\int_{-L}^{-\epsilon} \frac{1}{x}\phi(0) + \int_{\epsilon}^{L} \frac{1}{x}\phi(0) = 0.$$

If we do not take $\epsilon \to 0$, then we obtain an estimate on $\langle \mathrm{pv}\, \frac{1}{x}, \phi \rangle$ implying that the order of $\mathrm{pv}\, \frac{1}{x}$ is ≤ 1. To see that its order is exactly 1, we use the following result:

Lemma 3.14. *Let u be a distribution on $\mathbb{R}$ that agrees with $\frac{1}{x}$ on $(0, \infty)$. Then its order is at least* 1.

Proof. For the purpose of contradiction, suppose that the order of u is 0. Then there exists $C > 0$ such that

$$|\langle u, \phi \rangle| \leq C \sup_{x \in [-2,2]} |\phi(x)|$$

for all $\phi \in C_c^\infty(\mathbb{R})$ with $\operatorname{supp} \phi \subseteq [-2, 2]$. Now take $\phi(x) = \sum_{j=0}^{J} \varphi(2^j x)$, where J is a parameter that we will choose later and $\varphi \in C_c^\infty(\mathbb{R})$ obeys $|\varphi| \leq 1$, $\operatorname{supp} \varphi \subset (1, 2)$ and $\int \varphi \frac{\mathrm{d}x}{x} > 0$. Then on the one hand, by the assumption,

$$\langle u, \phi \rangle = \sum_{j=0}^{J} \int_{2^{-j}}^{2^{-j+1}} \frac{1}{x} \varphi(2^j x) \, \mathrm{d}x = (J + 1) \int_{1}^{2} \varphi(x) \frac{\mathrm{d}x}{x},$$

which can be made arbitrarily large by taking $J \to \infty$. On the other hand, since $\varphi(2^j x)$'s have disjoint supports, we have $|\phi| \leq 1$ and therefore $|\langle u, \phi \rangle| \leq C$ independent of J. This is a contradiction. $\square$

Finally, we remark that $\mathrm{pv}\, \frac{1}{x}$ is not the unique distribution that agrees with $\frac{1}{x}$ in $\mathbb{R} \setminus \{0\}$; however, as we will see below, all such distributions are of the form $\mathrm{pv}\, \frac{1}{x} + \sum_{k=0}^{K} c_k \delta_0^{(k)}$ for some $K \geq 0$ and coefficients $c_k \in \mathbb{R}$.

3.7. **Sequence of distributions.** We now discuss the notion of convergence of a sequence of distributions.

Definition 3.15. A sequence $u_n \in \mathcal{D}'(U)$ converges to $u \in \mathcal{D}'(U)$ if for all $\phi \in C_c^\infty(U)$, we have

$$\langle u_n, \phi \rangle \to \langle u, \phi \rangle.$$

When such a convergence holds, we say that $u_n \rightharpoonup u$ in the sense of distributions.

The key "sequential completeness" theorem for distributions is as follows.

Theorem 3.16. *Let u_n be a sequence of distributions on U with the following property: For each $\phi \in C_c^\infty(U)$, the sequence $\langle u_n, \phi\rangle$ converges as $n \to \infty$. Then there exists a distribution $u \in \mathcal{D}'(U)$ characterized by the property*

$$\langle u, \phi\rangle = \lim_{n\to\infty} \langle u_n, \phi\rangle, \quad \forall \phi \in C_c^\infty(U). \tag{3.3}$$

For every compact set $K \subset U$, there exist N and C (independent of n) such that (3.1) *holds for all u_n and u. Moreover, if $\phi_n \to \phi$ in $C_c^\infty(U)$, then $\langle u_n, \phi_n\rangle \to u(\phi)$.*

This theorem is very useful in two regards: First, when we check the existence of the limit of $\langle u_j, \phi\rangle$ for each $\phi \in C_c^\infty(U)$, the continuity of the limit u follows immediately, which is more cumbersome to check directly. Second, it implies that we can always "distribute the limits" in $\lim_{n\to\infty}\langle u_n, \phi_n\rangle = \langle \lim_{n\to\infty} u, \lim_{n\to\infty} \phi_n\rangle$.

If we consider a sequence of continuous linear functionals u_n on a Banach space X, then the analogous result is a consequence of the *uniform boundedness theorem*, which ensure that any family $\{u\}_{u\in\mathcal{U}}$ of continuous linear functionals on X that is pointwise bounded (i.e., $\{u(x)\}_{u\in\mathcal{U}}$ is bounded for each $x \in X$) is *always* equicontinuous. Unfortunately, $C_c^\infty(U)$ is not a Banach space, but a similar idea turns out to be true.

Proof. (Optional; for those who are familiar with functional analysis) It suffices to check the continuity of u on $C_c^\infty(U)$ as defined above. For any compact subset K of U, note that (3.3) holds for all $\phi \in C_c^\infty(K)$. Since $C_c^\infty(K)$ can be endowed with a complete invariant metric (see Remark 3.6), we can apply the uniform boundedness principle (or the Banach–Steinhaus theorem; see, for instance, [Rud91, Theorem 2.6]). As a result, there exists N_K and C_K independent of n such that each u_n obeys

$$\langle u_n, \phi\rangle \le C_K \sum_{|\alpha|\le N_K} \sup_{x\in K} |D^\alpha \phi(x)| \tag{3.4}$$

for $\phi \in C_c^\infty(K)$. Taking $n \to \infty$, the limit u obeys the same bound. Moreover, since K is an arbitrary compact subset of U, by Lemma 3.8 it follow that u is continuous on $C_c^\infty(U)$, as desired. Finally, the conclusion that $\langle u_n, \phi_n\rangle \to \langle u, \phi\rangle$ follows from the uniform bounds (3.4). ☐

Observe that "term-wise differentiation" is a triviality in distribution theory.

Lemma 3.17. *If $u_n \rightharpoonup u$, then $D^\alpha u_n \rightharpoonup D^\alpha u$ (both in the sense of distributions).*

Proof. It suffices to check the case $|\alpha| = 1$, i.e., $D^\alpha = \partial_j$. Since $u_n \rightharpoonup u$ in the sense of distributions, for all $\phi \in C_c^\infty(U)$ we have

$$\langle u_n, -\partial_j \phi\rangle \to \langle u, -\partial_j\phi\rangle$$

The left- and the right-hand sides are $\langle \partial_j u_n, \phi\rangle$ and $\langle \partial_j u, \phi\rangle$, respectively, so the claim follows. ☐

Equipped with these theoretical tools, let us discuss some concrete examples of convergence in the sense of distributions.

- The dominated convergence theorem allows us to connect pointwise convergence to that in the sense of distributions.

Lemma 3.18. *Suppose that $u_n(x) \in L^1_{loc}(\mathbb{R})$ satisfies $u_n(x) \to u_\infty(x)$ as $n \to \infty$. If there exists $v \in L^1_{loc}(\mathbb{R})$ such that*

$$|u_n(x)| \leq v(x) \quad \text{for almost every } x \in \mathbb{R},$$

then

$$u_t \rightharpoonup u_\infty \quad \text{as } t \to \infty \quad \text{in the sense of distributions.}$$

Proof. Indeed, for any $\phi \in C^\infty_c(U)$, we have

$$u_n\phi \to u\phi \quad \text{for a.e. } x \in U, \quad |u_n\phi(x)| \leq v(x)|\phi(x)| \quad \text{for a.e. } x \in U$$

where $v|\phi|$ is integrable since $v \in L^1_{loc}(U)$ and ϕ is bounded and has a compact support. Thus, by the dominated convergence theorem, $\langle u_n, \phi\rangle = \int u_n\phi \,\mathrm{d}x \to \int u\phi \,\mathrm{d}x = \langle u, \phi\rangle$ as desired. □

For example, if we take $h(x)$ to be a smooth function such that $h(x) = 1$ on $[1, \infty)$ and $\operatorname{supp} h \subset [0, \infty)$, then

$$h_\delta(x) = h(\delta^{-1}x) \to H(x),$$

both pointwisely and in the sense of distributions.

- When there are rapid oscillations, convergence in the sense of distributions may capture some cancellation that pointwise convergence does not. Take, for example,

$$u_n(x) = e^{inx} \quad \text{for } x > 0, \quad u_n(x) = 0 \text{ for } x \leq 0.$$

Then $u_n(x)$ does not have any pointwise limit for $x \in (0, \infty) \setminus 2\pi\mathbb{Z}$. However,

$$u_n \rightharpoonup 0 \quad \text{as } n \to \infty \quad \text{in the sense of distributions.}$$

Indeed,

$$\begin{aligned}
\langle u_n, \phi\rangle &= \int_0^\infty e^{inx}\phi(x)\,\mathrm{d}x \\
&= \int_0^\infty \frac{1}{in}\partial_x e^{inx}\phi(x)\,\mathrm{d}x \\
&= \frac{i}{n}\phi(0) + \int_0^\infty \frac{i}{n}e^{inx}\partial_x\phi(x)\,\mathrm{d}x \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

- In the preceding example, we see that if we normalize

$$v_n(x) = ne^{inx} \quad \text{for } x > 0, \quad v_n(x) = 0 \text{ for } x \leq 0,$$

then

$$v_n \rightharpoonup i\delta_0 \quad \text{as } n \to \infty \quad \text{in the sense of distributions.}$$

Indeed, for any $\phi \in C^\infty_c(\mathbb{R})$,

$$\begin{aligned}
\langle v_n, \phi\rangle &= \int_0^\infty ne^{inx}\phi(x)\,\mathrm{d}x \\
&= i\phi(0) + i\int_0^\infty e^{inx}\phi'(x)\,\mathrm{d}x \\
&= i\phi(0) - \frac{\phi'(0)}{n} - \int_0^\infty \frac{1}{n}e^{inx}\phi''(x)\,\mathrm{d}x \to i\phi(0) \quad \text{as } n \to \infty.
\end{aligned}$$

3.8. **Approximation of a distribution by C^∞ (or C_c^∞) functions.** Let us start with a simple computation that underlies the mollification procedure using convolution.

Lemma 3.19 (Approximation of the identity)**.** *Let $\phi \in C_c^\infty(\mathbb{R}^d)$ and $\phi_\delta = \delta^{-d}\phi(\delta^{-1}x)$. Then*

$$\phi_\delta \rightharpoonup \left(\int \phi(x)\,\mathrm{d}x\right)\delta_0 \quad \text{as } \delta \to 0 \quad \text{in the sense of distributions.}$$

Proof. To see this, for every $\psi \in C_c^\infty(\mathbb{R}^d)$, we compute

$$\begin{aligned}\langle \phi_\delta, \psi\rangle &= \int \delta^{-d}\phi(\delta^{-1}x)\psi(x)\,\mathrm{d}x \\ &= \int \phi(z)\psi(\delta z)\,\mathrm{d}z,\end{aligned}$$

where we made the change of variables $z = \delta^{-1}x$. By the dominated convergence theorem, it follows that

$$\int \phi(z)\psi(\delta z)\,\mathrm{d}z \to \int \phi(z)\,\mathrm{d}z\psi(0) \quad \text{as } \delta \to 0,$$

which is equivalent to the above claim. □

As a consequence of the last computation, we obtain the following useful result:

Proposition 3.20. *Let $\phi \in C_c^\infty(\mathbb{R}^d)$ satisfy $\int \phi = 1$. For every $\delta > 0$, define $\phi_\delta(x) = \delta^{-d}\phi(\delta^{-1}x)$. Then*

$$\phi_\delta * u \rightharpoonup u \quad \text{as } \delta \to 0 \quad \text{in the sense of distributions.}$$

Proof. We need to show that, for all $\psi \in C_c^\infty(U)$

$$\langle \phi_\delta * u, \psi\rangle \to \langle u, \psi\rangle \quad \text{as } \delta \to 0.$$

Recall that $\langle \phi_\delta * u, \psi\rangle = \langle u, \phi_\delta *' \psi\rangle$ and

$$\phi_\delta *' \psi(y) = \int \phi_\delta(x-y)\psi(x)\,\mathrm{d}x = \langle \phi_\delta(\cdot - y), \psi\rangle.$$

By the preceding computation, it follows that the RHS $\to \psi(y)$ as $\delta \to 0$. Moreover, writing

$$\langle \phi_\delta(\cdot - y), \psi\rangle = \langle \phi_\delta, \psi(\cdot + y)\rangle$$

it is not difficult to verify that $D^\alpha\langle \phi_\delta(\cdot - y), \psi\rangle \to D^\alpha\psi(y)$ for every $y \in \operatorname{supp}\psi$ and α. It follows that $\langle \phi_\delta(\cdot - y), \psi\rangle \to D^\alpha\psi(y)$ in $C_c^\infty(U)$. Then by the continuity of U, the desired conclusion follows. □

Note that $\phi_\delta * u$'s are smooth functions that approximate u in the sense of distributions. By taking an additional cutoff, we may approximate $u \in \mathcal{D}'(U)$ by smooth and compactly supported functions in $C_c^\infty(U)$ in the sense of distributions. More precisely, take a sequence K_n of compact subsets of U such that $K_n \subset \mathrm{int}K_{n+1}$ and $\cup_n K_n = U$. Then define $\chi_n \in C_c^\infty(U)$ such that $\chi_n = 1$ on K_n and $\operatorname{supp}\chi_n \subset \mathrm{int}K_{n+1}$ (the construction of such χ_n is easily achieved by starting with a continuous function with similar properties, then applying mollification). For some sequence $\delta_n \to 0$ to be chosen below, we take as our approximating sequence

$$u_n = \phi_{\delta_n} * (\chi_n u).$$

Note that the convolution is well-defined, since $\chi_n u$ is now a compactly supported distribution in $\mathcal{D}'(U)$. Moreover, choosing $\delta_n < \mathrm{dist}(K_{n+1}, \partial U)$, we may ensure that $\operatorname{supp} u_n \subset U$ using Lemma 3.12. Using Proposition 3.20 and the properties of χ_n, it is not difficult to prove that:

Proposition 3.21. *For any $u \in \mathcal{D}'(U)$, the sequence u_n defined above converges to $u \in \mathcal{D}'(U)$ in the sense of distributions. Hence, $C_c^\infty(U)$ is dense in $\mathcal{D}'(U)$.*

These properties present us another way to generalize operations on smooth functions to distributions:

Basic principle (the approximation method): Let $\mathcal{A}$ be an operation on smooth (resp. and compactly supported) functions. Then $\mathcal{A}$ is generalized to a distribution $u \in \mathcal{D}'(U)$ by considering a sequence u_j in $C^\infty(U)$ (resp. $C_c^\infty(U)$) that approximates u in the sense of distributions, and then "defining" $\mathcal{A}u$ to be $\lim_{n\to\infty} \mathcal{A}u_n$.

For this method to work, $\lim_{n\to\infty} \mathcal{A}u_n$ needs to be independent of the choice of u_n. For many basic operations of interest (including those discussed above), this property holds thanks to the "continuity" property of the operation. In practice, this method is often the more useful and flexible way to define basic operations on distributions.

Let us close this subsection with a simple example where the idea of approximation suggests a straightforward proof of a statement regarding distributions:

Lemma 3.22. *If $u \in \mathcal{D}'(U)$ such that $\partial_j u = 0$, then u is a constant.*

Proof. Let us first take the case $U = \mathbb{R}^d$, which is very simple. By Proposition 3.20, we have $\phi_\delta * u \rightharpoonup u$ as $\delta \to 0$ in the sense of distributions (following the notation of the proposition). On the other hand, $\partial_j(\phi_\delta * u) = \phi_\delta * \partial_j u = 0$, so each $\phi_\delta * u$ is a constant, which we denote by C_δ. Thus, $\langle \phi_\delta * u, \psi \rangle = C_\delta \int \psi \, \mathrm{d}x$ is convergent for every $\psi \in C_c^\infty(\mathbb{R}^d)$, from which it follows that C_δ is convergent and $u = \lim_{\delta\to 0} C_\delta$.

In the case of a general domain U, we take the approximating sequence $u_n = \phi_{\delta_n} * (\chi_n u)$, where χ_n and δ_n are chosen as above. On the one hand, it converges to u in the sense of distributions. On the other hand, for every fixed n_0, since $\chi_n(x) = 1$ for $x \in K_{n_0}$ for $n > n_0$, we have

$$\partial_j u_n(x) = \partial_j(\phi_{\delta_n} * (\chi_n u))(x) = 0 \text{ for all } x \in K_{n_0} \text{ and } n > n_0.$$

Thus, u_n is a constant on K_{n_0} for $n > n_0$. By a similar argument as before, it follows that for any $\psi \in C_c^\infty(K_{n_0})$, $\langle u_n, \psi \rangle \to C_{K_{n_0}} \psi$ for some constant $C_{K_{n_0}}$. Since n_0 is arbitrary, it follows that $C = C_{K_{n_0}}$ is independent of K_{n_0} and we have $\langle u, \psi \rangle = \lim_{n\to\infty} \langle u_n, \psi \rangle = C\psi$ for any $\psi \in C_c^\infty(U)$, as we wished. □

It is instructive to come up with a proof of this lemma that only uses the adjoint method (i.e., the definition $\langle \partial_j u, \phi \rangle = -\langle u, \partial_j \phi \rangle$), and compare it with the preceding straightforward proof.

3.9. **Differentiation of the characteristic function.** As an application of the theory developed so far, let us generalize the computation $H'(x) = \delta_0$ to higher dimensions. Given a set $U \subset \mathbb{R}^d$, we introduce its *characteristic function* $\mathbf{1}_U$ defined by

$$\mathbf{1}_U(x) = \begin{cases} 1 & \text{when } x \in U \\ 0 & \text{when } x \in \mathbb{R}^d \setminus U. \end{cases}$$

Proposition 3.23. *Let U be an open domain in $\mathbb{R}^d$ with a C^1 boundary. Then*

$$\partial_j \mathbf{1}_U = -(\nu_{\partial U})_j \mathrm{d}S_{\partial U},$$

where $\mathrm{d}S_{\partial U}$ is the (Euclidean) surface element on ∂U.

Here, by the notation $\mathrm{d}S_{\partial U}$, we mean the distribution

$$\langle \mathrm{d}S_{\partial U}, \phi \rangle = \int_{\partial U} \phi|_{\partial U} \, \mathrm{d}S_{\partial U}.$$

(In other words, view $\mathrm{d}S_{\partial U}$ as a Borel measure supported on ∂U.)

Before getting to the proof, let us note that Proposition 3.23 furnishes a proof of the divergence theorem for smooth vector fields. Indeed, if $b^j \in C^\infty(\mathbb{R}^d)$, then

$$\int_U \mathrm{div} b \, \mathrm{d}x = \langle \mathbf{1}_U, \sum_j \partial_j b^j \rangle = -\sum_j \langle \partial_j \mathbf{1}_U, b^j \rangle = \int_{\partial U} \nu_j b^j \, \mathrm{d}S.$$

In fact, Proposition 3.23 is *equivalent* to the divergence theorem for smooth vector fields. However, we will present an independent proof. The idea will be to use a suitable approximating sequence of $\mathbf{1}_U$.

Proof. Note that $\partial_j \mathbf{1}_U$ is supported in ∂U. Our strategy is to first find a covering ∂U by finitely many open balls B_α (i.e., $\partial U \subset \cup_\alpha B_\alpha$) such that $\langle \partial_j \mathbf{1}_U, \phi \rangle$ may be easily computed for $\phi \in C^\infty_c(B_\alpha)$. Then we will use a smooth partition of unity to piece together these local computations.

By the definition of a domain with a C^1 boundary, for every $x_0 \in \partial U$ we can find $r_{x_0} > 0$ such that, after suitably rearranging and reorienting coordinates, we have

$$B(x_0, r_{x_0}) \cap \partial U = \{x^d = \gamma(x^1, \dots, x^{d-1})\}, \quad B(x_0, r_{x_0}) \cap U = \{x^d < \gamma(x^1, \dots, x^{d-1})\}$$

for some C^1 function γ. By compactness, we can find finitely many points x_α and balls $B_\alpha = B(x_\alpha, r_{x_\alpha})$ with this property, such that $\partial U \subset \cup_\alpha B_\alpha$.

Let us fix one ball B_α. Let h be a smooth function that equals 1 on $[1, \infty)$ and $\mathrm{supp}\, h \subset (0, \infty)$.

$$\mathbf{1}_U = \lim_{\delta \to 0} h(\delta^{-1}(\gamma(x^1, \dots, x^{d-1}) - x^d)) \quad \text{in } B_\alpha$$

pointwisely, and thus also in the sense of distributions (by the dominated convergence theorem). For any $\phi \in C^\infty_c(U)$ with $\mathrm{supp}\, \phi \subset B_\alpha$,

$$\begin{aligned} \langle \partial_j \mathbf{1}_U, \phi \rangle &= \lim_{\delta \to 0} \langle \partial_j \left(h(\delta^{-1}(\gamma(x^1, \dots, x^{d-1}) - x^d)) \right), \phi \rangle \\ &= -\lim_{\delta \to 0} \langle \tilde{\nu}_j \delta^{-1} h'(\delta^{-1}(\gamma(x^1, \dots, x^{d-1}) - x^d)), \phi \rangle \end{aligned}$$

where

$$\tilde{\nu} = \nabla(x^d - \gamma(x^1, \dots, x^{d-1})) = (-\partial_1 \gamma, \dots, -\partial_{d-1} \gamma, 1)$$

points outwards of U. If we freeze $(x^1, \dots, x^{d-1})$, then $\delta^{-1} h'(\delta^{-1}(\gamma(x^1, \dots, x^{d-1}) - x^d)) \rightharpoonup \delta_{\gamma(x^1, \dots, x^{d-1})}(x^d)$ (as distributions in x^d). Thus,

$$\langle \partial_j \mathbf{1}_U, \phi \rangle = -\int \tilde{\nu}_j(x^1, \dots, x^{d-1}) \phi(x^1, \dots, x^{d-1}, \gamma(x^1, \dots, x^{d-1})) \, \mathrm{d}x^1 \cdots \mathrm{d}x^{d-1}.$$

Now note that

$$\tilde{\nu} = (-\partial_1 \gamma, \dots, -\partial_{d-1} \gamma, 1) = \nu \sqrt{1 + |\nabla \gamma|^2},$$

where ν is the outward unit normal to ∂U, whereas we recall from multivariable calculus that

$$\sqrt{1+|\nabla\gamma|^2}\,\mathrm{d}x^1\dots\mathrm{d}x^{d-1}=\mathrm{d}S \quad \text{on } \partial U.$$

Thus, it follows that

$$\langle \partial_j \mathbf{1}_U, \phi\rangle = -\int_{\partial U} \nu_j\, \phi|_{\partial U}\,\mathrm{d}S,$$

for $\phi \in C^\infty_c(B_\alpha)$.

Finally, to piece together the local computations, we use a smooth partition of unity χ_α adapted to B_α. More precisely, there exists a family $\{\chi_\alpha\} \subset C^\infty_c(\mathbb{R}^d)$ such that $\operatorname{supp}\chi_\alpha \subset B_\alpha$, and $\sum_\alpha \chi_\alpha(x) = 1$ in a neighborhood of $\partial U(\subset \cup_\alpha B_\alpha)$. Such a family of functions can be constructed by starting with a continuous partition of unity $\tilde{\chi}_\alpha$ adapted to B_α, which is easier to construct, and then taking $\chi_\alpha = \varphi_\delta * \chi_\alpha$ for $\varphi \in C^\infty_c(\mathbb{R}^d)$ with $\int \varphi = 1$ and a suitably small $\delta > 0$. To ensure the last property, note that $\sum_\alpha \chi_\alpha(x) = \varphi_\delta * \sum_\alpha \tilde{\chi}_\alpha(x) = 1$ if δ is small enough so that $B(x,\delta) \subset \{x : \sum_\alpha \tilde{\chi}_\alpha(x) = 1\}$.

Given any $\phi \in C^\infty_c(U)$, we write $\phi = \sum_\alpha \chi_\alpha \phi$ and apply the local computation in B_α to each $\chi_\alpha\phi$. Then the desired result follows. □

3.10. **Operations between distributions; simple cases.** We may now ask if we can make sense of operations between distributions. Let us cover some simple (but useful, as we will see) situations when this can be done.

- *Integral of a compactly supported distribution.* For any test function $u \in C^\infty_c(U)$, $\int u(y)\,\mathrm{d}y = \langle u, 1\rangle$. This (unary) operation can be extended to compactly supported distributions in the following simple way. First, observe that $u \in \mathcal{D}'(U)$ has compact support, then u can be tested against any smooth function $\phi \in C^\infty(U)$.

  Lemma 3.24. *Let $u \in \mathcal{D}'(U)$ be a distribution with compact support. Let $\chi \in C^\infty_c(U)$ be such that $\chi = 1$ on $\operatorname{supp} u$. Then*

  $$u = \chi u.$$

  Moreover, for any $f \in C^\infty(U)$, if we define

  $$\langle u, f\rangle := \langle u, \chi f\rangle,$$

  then this definition is independent of χ.

  We omit the straightforward proof.

  Let $u \in \mathcal{D}'(U)$ be a distribution with compact support. By Lemma 3.24, it follows that $\langle u, 1\rangle$ is well-defined. We introduce the notation

  $$\int u = \langle u, 1\rangle.$$

  Note that this notation is consistent with the usual notion of integration for $u \in C^\infty_c(U)$. In fact, we will often abuse the notation and write $\int u(y)\,\mathrm{d}y$ for the LHS as well, when it helps to clarify the dependence of $\int u$ on parameters.

  We note a very simple lemma, whose easy proof we will omit, which allows us to perform "integration by parts" for distributions:

Lemma 3.25. *Let $u \in \mathcal{D}'(U)$ be a distribution with compact support. Then for any j,*

$$\int \partial_j u = 0.$$

- *Multiplication of two distributions with disjoint singular supports.* The product of two distributions $u, v \in \mathcal{D}'(U)$ is, in general, not well-defined. However, if for every $x \in U$ at least one of u or v is a smooth function in a neighborhood V of x (say u), then locally we are in the same situation as before (i.e., defining multiplication of a smooth function and a distribution) so uv should be well-defined.

  To formalize this idea, we introduce the notion of the *singular support*:

Definition 3.26 (Singular support of a distribution)**.** Let $u \in \mathcal{D}'(U)$. The *singular support of* u is defined to be the complement of the union of all open sets on which u coincides with a smooth function, i.e.,

$$\operatorname{sing\,supp} u = U \setminus \left(\bigcup \{V : V \text{ is an open subset of } U \text{ on which } u \in C^\infty(V)\} \right).$$

With this definition, we can formulate and prove the following result:

Proposition 3.27. *Let $u, v \in \mathcal{D}'(U)$ such that* $\operatorname{sing\,supp} u \cap \operatorname{sing\,supp} v = \emptyset$. *Then the product uv is well-defined.*

Proof. Take approximating sequences u_n and v_n in $C_c^\infty(U)$ of u and v, respectively. Our goal is to show that

$$\lim_{n \to \infty} u_n v_n$$

is well-defined in the sense of distributions. For this purpose, by Theorem 3.16, it suffice to show that, for every $\phi \in C_c^\infty(U)$,

$$\lim_{n \to \infty} \int u_n v_n \phi \, \mathrm{d}x$$

is well-defined.

Note that, since $\operatorname{sing\,supp} u$ and $\operatorname{sing\,supp} v$ are disjoint closed sets in $U \subseteq \mathbb{R}^d$, there exists $\chi \in C^\infty(\mathbb{R}^d)$ such that $\chi = 1$ on $\operatorname{sing\,supp} v$ and $\operatorname{supp} \chi \cap \operatorname{sing\,supp} u = \emptyset$. We split $\phi = \chi\phi + (1 - \chi)\phi$, and note that $\operatorname{supp}(\chi\phi) \cap \operatorname{sing\,supp} u = \emptyset$ while $\operatorname{supp}((1 - \chi)\phi) \cap \operatorname{sing\,supp} v = \emptyset$.

By the construction and Lemma 3.4, it follows that, for every α, $D^\alpha u_n \to D^\alpha u$ uniformly on every compact set K such that $K \cap \operatorname{sing\,supp} u = \emptyset$. Thus $u_n \chi\phi \to u\chi\phi$ in $C_c^\infty(U)$. Then by Theorem 3.16, $\int u_n v_n \chi\phi \, \mathrm{d}x \to \langle v, \chi u\phi \rangle$. Similarly, $\int u_n v_n \chi\phi \, \mathrm{d}x \to \langle u, (1 - \chi)v\phi \rangle$. Putting these two statements together, the conclusion follows. $\square$

In fact, this theorem could be proved more quickly using the adjoint method. However, the approximation method is more flexible, in that the same recipe can be used to justify the definition of uv even when Proposition 3.27 does not apply. Here is one example:

– *Product of surface elements on transversal hyperplanes.* On $\mathbb{R}^2$, take

$$\langle u, \phi \rangle = \int \phi(0, y) \, \mathrm{d}y, \quad \langle v, \phi \rangle = \int \phi(x, 0) \, \mathrm{d}x.$$

With the notation as before, we may write $u = \mathrm{d}S_{\{x=0\}}$ and $v = \mathrm{d}S_{\{y=0\}}$. Clearly, $\operatorname{sing\,supp} u = \{x = 0\}$ and $\operatorname{sing\,supp} v = \{y = 0\}$, so that $\operatorname{sing\,supp} u \cap \operatorname{sing\,supp} v \neq \emptyset$. However, since

$$u = \lim_{\delta \to 0} \delta^{-1} h'(\delta^{-1} x), \quad v = \lim_{\delta \to 0} \delta^{-1} h'(\delta^{-1} y)$$

where h is a smooth function with $h = 1$ on $[1, \infty)$ and $\operatorname{supp} h \subseteq [0, \infty)$, we wish to define (approximation method)

$$uv \text{``}=\text{''} \lim_{\delta \to 0} \delta^{-2} h'(\delta^{-1} x) h'(\delta^{-1} y).$$

According to an earlier example, the limit on the RHS equals δ_0 on $\mathbb{R}^2$; thus $uv = \delta_0$ is well-defined.

- *Convolution with a compactly supported distribution.* Next, we turn to the task of defining the convolution of two distributions $u, v \in \mathcal{D}'(\mathbb{R}^d)$. Again, in general, this operation is not well-defined. However, we have the following result:

Proposition 3.28. *Let $u, v \in \mathcal{D}'(\mathbb{R}^d)$ such that at least one of u or v has a compact support. Then $u * v$ is well-defined. Moreover, we have*

$$u * v = v * u$$

and

$$\operatorname{supp}(u * v) \subseteq \operatorname{supp} u + \operatorname{supp} v.$$

Proof. Take approximating sequences $u_n = \varphi_{2^{-n}} * u$ and $v_n = \varphi_{2^{-n}} * v$. In view of Theorem 3.16, we wish to show that, for every $\phi \in C^\infty_c(\mathbb{R}^d)$,

$$\lim_{n \to \infty} \int u_n * v_n(x) \phi(x) \, \mathrm{d}x$$

is well-defined.

Let us assume that v is compactly supported; the other case is similar. As we have seen before,

$$\begin{aligned} \int u_n * v_n(x) \phi(x) \, \mathrm{d}x &= \int u_n(y) \left(\int v_n(x - y) \phi(x) \, \mathrm{d}x \right) \mathrm{d}y \\ &= \int u_n(y) \left(\int v_n(z) \phi(z + y) \, \mathrm{d}z \right) \mathrm{d}y \end{aligned}$$

For each fixed $y \in \mathbb{R}^d$, we have $\int v_n(z)\phi(z + y) \, \mathrm{d}z \to \langle v, \phi(\cdot + y) \rangle$ by the convergence $v_n \to v$. Differentiating in y, we obtain the same conclusion for any derivatives. Finally, by (a variant of) Lemma 3.12, we see that there exists a compact set K that contains the supports of $\int v_n(z)\phi(z+y) \, \mathrm{d}z$ and $\langle v, \phi(\cdot + y) \rangle$ (both viewed as functions of y). It follows that $\int v_n(z)\phi(z+y) \, \mathrm{d}z \to \langle v, \phi(\cdot + y) \rangle = v *' \phi(y)$ in $C^\infty_c(\mathbb{R}^d)$. Finally, by Theorem 3.16, it follows that $\lim_{n \to \infty} \int u_n * v_n(x)\phi(x) \, \mathrm{d}x = \langle u, v *' \phi \rangle$, as desired.

At last, the properties $u * v = v * u$ and $\operatorname{supp}(u * v) \subseteq \operatorname{supp} u + \operatorname{supp} v$ easily follow from the corresponding properties for functions via approximation; we omit the straightforward verification. □

Again, the adjoint method would have led to a quicker proof, but we followed the approximation method since it provides a strategy for defining $u * v$ in more general situations.

As an application of Proposition 3.28, we note that the convolution with δ_0 is well-defined for any distribution $u \in \mathcal{D}'(\mathbb{R}^d)$. In fact,

$$\delta_0 * u = u * \delta_0 = u.$$

3.11. **Fundamental solutions and representation formula (optional).** The purpose of this subsection is to motivate the concept of a fundamental solution, and explain the general strategy for using fundamental solutions to understand a linear PDE problem. The arguments here will mostly be formal (i.e., without proof), but they will motivate how we approach constant coefficient linear scalar PDEs (e.g., the Laplace and the wave equations) later, where we will indeed follow and justify (parts of) the strategies outlined here.

Consider a linear scalar differential operator $\mathcal{P}$ on $\mathbb{R}^d$, which is of the form

$$\mathcal{P}u = \sum_{\alpha:|\alpha|\leq k} a_\alpha(x) D^\alpha u.$$

Let us assume that each a_α is $C^\infty(\mathbb{R}^d)$. Its formal adjoint is given by

$$\mathcal{P}'v = \sum_{\alpha:|\alpha|\leq k} (-1)^{|\alpha|} D^\alpha(a_\alpha v).$$

Indeed, $\langle \mathcal{P}u, v\rangle = \langle u, \mathcal{P}'v\rangle$ if one of u or v is in $\mathcal{D}'(\mathbb{R}^d)$ and the other is in $C_c^\infty(\mathbb{R}^d)$.

We are interested in the question of the *existence* of a solution to the inhomogeneous problem

$$\mathcal{P}u = f, \tag{3.5}$$

and also in the question of its *uniqueness*.

Fundamental solutions and the problem of existence. Let us first consider the problem of *existence*, and proceed by an analogy with (finite dimensional) linear algebra. If u and f belonged to finite dimensional vector spaces $\mathbf{U}$ and $\mathbf{F}$, respectively, and $\mathcal{P} : \mathbf{U} \to \mathbf{F}$ is a linear operator, then we can find a solution to (3.5) by the following procedure in linear algebra:

(1) find a set of vectors $\{d_i\}$ that spans the space where f belongs;
(2) for each i, find a solution e_i to $\mathcal{P}e_i = d_i$;
(3) for $f = \sum c_i d_i$, write $u = \sum c_i e_i$, which solves $\mathcal{P}u = f$ by linearity.

Let us try to follow this strategy for (3.5). For the moment, let us take f to be as nice as it can be, e.g., $f \in C_c^\infty(\mathbb{R}^d)$. To motivate the notion of a fundamental solution, let us start by recalling the identity

$$f(y) = \delta_0 * f(y) = \int \delta_0(y-x) f(x)\,\mathrm{d}x = \int \delta_y(x) f(x)\,\mathrm{d}x,$$

where the last two integrals should be interpreted as $\langle \delta_0(y-\cdot), f\rangle$ and $\langle \delta_y, f\rangle$. (Note that, as we saw in the last subsection, this identity makes sense for any distribution u, too!) This identity suggests that f belongs to the "span" of $\{\delta_y\}$, so by linearity, once we know a solution to $\mathcal{P}u = f$ with f equal to the delta distributions, then we can find a solution u to (3.5).

Motivated by the preceding considerations, we make the following definition:

Definition 3.29. For each $y \in \mathbb{R}^d$, we define a *fundamental solution* E_y *for* $\mathcal{P}$ *at* y to be a distribution $E_y \in \mathcal{D}'(y)$ satisfying

$$\mathcal{P}E_y = \delta_y.$$

It should be emphasized that a fundamental solution is usually not *unique*. Indeed, if E_y is a fundamental solution at y, then $E_y + h$ for any homogeneous solution $\mathcal{P}h = 0$ is also a fundamental solution at y; conversely, any fundamental solution at y would be of the form $E_y + h$ for some h satisfying $\mathcal{P}h = 0$.

For a sufficiently nice (say, smooth and compactly supported) f, we formally define

$$u[f] \text{``}={}\text{''} \int f(y) E_y(x) \, \mathrm{d}y.$$

The expectation is that

$$\mathcal{P}u[f] \text{``}={}\text{''} \int f(y) \mathcal{P} E_y(x) \, \mathrm{d}y = f(x);$$

or succinctly, that $f \mapsto u[f]$ is a left-inverse for $\mathcal{P}$. In practice, each " = " must be justified in a case-by-case basis.

Fundamental solutions and the problem of uniqueness (representation formula). Amusingly, fundamental solutions, which are primarily vehicles for proving existence, are also useful for investigating *uniqueness* of a solution to (3.5). To see this, we again start by reviewing a similar procedure in linear algebra.

Let $\mathbf{U}$ and $\mathbf{F}$ be finite dimensional vector spaces, and let $\mathcal{P} : \mathbf{U} \to \mathbf{F}$ be a linear operator. Given a vector $u \in \mathbf{U}$, let us try to determine u from $\mathcal{P}u$. Let $\mathbf{U}'$ and $\mathbf{F}'$ be the dual vector spaces of $\mathbf{U}$ and $\mathbf{F}$, respectively, and consider the adjoint $\mathcal{P}' : \mathbf{F}' \to \mathbf{U}'$ of $\mathcal{P}$ defined by

$$\langle \mathcal{P}u, g \rangle = \langle u, \mathcal{P}' g \rangle, \quad \text{for all } u \in \mathbf{U}, \, g \in \mathbf{F}'.$$

Let $\{(d')^i\}$ span the space $\mathbf{U}'$. Then in order to determine u, it suffices to determine $\langle u, (d')^i \rangle$ for each i. Suppose for each i, we know a solution $(e')^i$ to

$$\mathcal{P}'(e')^i = (d')^i.$$

Then

$$\langle u, (d')^i \rangle = \langle u, \mathcal{P}'(e')^i \rangle = \langle \mathcal{P}u, (e')^i \rangle.$$

Hence, we see how the existence of solutions $(e')_i$ to the adjoint problem leads to a representation formula for u in terms of $\mathcal{P}u$!

Let us now return to the case of PDEs. Suppose that for each $x \in \mathbb{R}^d$, there exists a fundamental solution $(E')^x$ to the adjoint problem

$$\mathcal{P}'(E')^x = \delta_x.$$

For a distribution u that is sufficiently regular (say, smooth) and compactly supported, we perform the following formal manipulation:

$$\begin{aligned} u(x) &= \langle u, \delta_x \rangle \\ &= \langle u, \mathcal{P}'(E')^x \rangle \\ &\text{``}={}\text{''} \langle \mathcal{P}u, (E')^x \rangle. \end{aligned}$$

Again, " = " needs to be justified on a case-by-case basis. The compact support property of u (in addition to sufficient regularity) is important to not generate any boundary terms. When this identity can be justified, we say that we have a *representation formula* for u in terms of $\mathcal{P}u$. To express the expectation succinctly:

$$f \mapsto (x \mapsto \langle f, (E')^x \rangle) \text{ should be a left-inverse for } \mathcal{P}.$$

Fundamental solutions and representation formula for boundary value problems. A variant of the preceding procedure leads to a strategy for deriving a representation formula for a solution u to a boundary value problem on a domain U. Suppose that U is a C^1 domain and

$$\mathcal{P}u = f \quad \text{in } U.$$

The strategy for finding a representation formula for u in terms of $\mathcal{P}u$ in U and the data on ∂U is to justify the following formal manipulation:

$$\begin{aligned} u(x) &= \langle u, \delta_x \rangle \\ &= \langle \mathbf{1}_U u, \mathcal{P}'(E')^x \rangle \\ \text{``} &= \langle \mathbf{1}_U \mathcal{P}u, (E')^x \rangle + \cdots . \end{aligned}$$

In the remainder $\cdots$, we expect to see only $u|_{\partial U}, \ldots, D^\alpha u|_{\partial U}$ for $|\alpha| \leq k-1$ since at least one derivative must fall on $\mathbf{1}_U$, which we computed in Proposition 3.23.

The constant coefficient case. When the the coefficients a_α in the definition of $\mathcal{P}$ are constant, there is a natural way to generate fundamental solutions at $y \in \mathbb{R}^d$ starting from one of them. The key idea is that in the constant coefficient case, $\mathcal{P}$ is translation invariant, i.e.,

$$\mathcal{P}(u(\cdot - y)) = (\mathcal{P}u)(\cdot - y).$$

Thus, given a fundamental solution E_0 at δ_0, it follows that its translate

$$E_y = E_0(\cdot - y)$$

satisfies $\mathcal{P}E_y = \delta_y$; in other words, E_y is a fundamental solution at $y \in \mathbb{R}^d$.

Moreover, the formal adjoint $\mathcal{P}'$ of $\mathcal{P}$ is given by

$$\mathcal{P}' = \sum_{\alpha : |\alpha| \leq k} a_\alpha (-1)^{|\alpha|} D^\alpha .$$

or equivalently,

$$\mathcal{P}'(u(-\,\cdot)) = (\mathcal{P}u)(-\,\cdot).$$

Therefore, we see that

$$(E')^x = E_0(x - \,\cdot)$$

is a fundamental solution for $\mathcal{P}'$ at x, i.e., $\mathcal{P}'(E')^x = \delta_x$.

Using the fundamental solutions E_y and E'_x generated from E_0 in the above fashion, the formal formulas that we discussed before take the form of convolution.

- *Existence.* The proposed formula for $u[f]$ is

$$u[f](x) \text{``}=\text{''} \int f(y) E_y(x) \, \mathrm{d}y = \int f(y) E_0(x - y) \, \mathrm{d}y = E_0 * f(x).$$

  In particular, by Proposition 3.28, the last expression always makes sense when f is a compactly supported distribution. Thus,

  Proposition 3.30. *Let $\mathcal{P}$ be a constant coefficient partial differential operator on $\mathbb{R}^d$. Let E_0 be a fundamental solution for $\mathcal{P}$ at 0, and let f be a compactly supported distribution. Then*

$$u[f] = E_0 * f,$$

  solves $\mathcal{P}u = f$.

- *Representation formula for a compactly supported u.* Recall that the proposed representation formula for $u(x)$ is

$$u(x) = \langle u, \delta_x \rangle = \langle u, \mathcal{P}'(E')^x \rangle$$
$$\text{``} = \text{''} \langle \mathcal{P}u, (E')^x \rangle.$$

With the above choices of $(E')^x$,

$$\langle u, \mathcal{P}'(E')^x \rangle = u * \mathcal{P}E_0(x), \quad \langle \mathcal{P}u, (E')^x \rangle = (\mathcal{P}u) * E_0(x).$$

So the justification of the representation formula (or more concretely, "$=$") amounts to justifying the passage of the derivatives in $\mathcal{P}$ from E_0 to u in the convolution, i.e.,

$$u(x) = u * \mathcal{P}E_0(x) \text{``} = \text{''} (\mathcal{P}u) * E_0(x).$$

This is possible when u is compactly supported. Thus,

Proposition 3.31. *Let $\mathcal{P}$ be a constant coefficient partial differential operator on $\mathbb{R}^d$. Let E_0 be a fundamental solution for $\mathcal{P}$ at 0, and let u be a compactly supported distribution. Then*

$$u = E_0 * \mathcal{P}u.$$

- *Representation formula for a u solving a boundary value problem.* Let U be a C^1 domain (not necessarily bounded), and let $u \in C^\infty(\overline{U})$ (i.e., u extends to a smooth function to an open set $V \supset \overline{U}$). In this case, the proposed representation formula for $u(x)$ for $x \in U$ is

$$u(x) = \langle \delta_x, u \rangle = \int \mathcal{P}'(E')^x u \mathbf{1}_U$$
$$\text{``} = \int (E')^x \mathcal{P}u \mathbf{1}_U + \cdots ,$$

so as before, the justification of the representation formula amounts to justifying the passage of the derivatives in $\mathcal{P}$ from E_0 to u in the convolution, i.e.,

$$u(x) = \mathcal{P}E_0 * u\mathbf{1}_U(x) \text{``} = \text{''} E_0 * (\mathcal{P}u)\mathbf{1}_U(x) + \cdots .$$

The omitted terms $\cdots$ would involve the values of $D^\alpha u$ on ∂U with $|\alpha| \leq k-1$ since at least one derivative would fall on $\mathbf{1}_U$.

One useful special case to keep in mind is when U is a bounded domain; in that case, by Proposition 3.31 we have

$$u = E_0 * \mathcal{P}(u\mathbf{1}_U) = E_0 * ((\mathcal{P}u)\mathbf{1}_U) + \mathcal{B},$$

where the boundary integrals should be contained in

$$\mathcal{B} = E_0 * \mathcal{P}(u\mathbf{1}_U) - E_0 * ((\mathcal{P}u)\mathbf{1}_U).$$

It remains to use the product rule to compute $\mathcal{B}$, and justify that they indeed give rise to well-defined integrals on ∂U; the latter step involves checking that $E_0(x - \cdot)$ and its derivatives have good properties on ∂U. We will carry out this abstract procedure in concrete cases (the Laplace and the wave equations) below.

Remark 3.32. It is worth noting that every *constant coefficient* scalar linear partial differential operator has a fundamental solution; this is the celebrated theorem of Malgrange–Ehrenpreis [Rud91, Theorem 8.5]. However, this theorem per se does not tell us much about how the solution looks like. Moreover, fundamental solutions need not exist in the general linear case[5].

Examples. Finally, let us discuss a simple example to illustrate the strategies described above.

- *The operator ∂_x on $\mathbb{R}$*: A fundamental solution for ∂_x (which has constant coefficients) on $\mathbb{R}$ is
$$\partial_x H(x) = \delta_0.$$
Moreover, any fundamental solution $E(x)$ has the property that $\partial_x(E - H) = 0$. Thus, a general fundamental solution is given by $E(x) = H(x) + C$.

  Let us carry out the strategies outlined above for this simple example. For concreteness, we use the fundamental solution $H(x)$.
  - *Existence.* For $f \in C_c^\infty(\mathbb{R})$, we define
$$u[f](x) := \int f(y)H(x-y)\,\mathrm{d}y = f * H(x).$$
Then $\partial_x u[f](x) = (f * \partial_x H) = f$, as desired. Alternatively, note that
$$u[f](x) = \int f(y)H(x-y)\,\mathrm{d}y = \int_{-\infty}^{x} f(y)\,\mathrm{d}y,$$
so $\partial_x u[f](x) = f$ by the fundamental theorem of calculus.
  - *Representation formula for a "nice" u in $\mathbb{R}$.* For $u \in C_c^\infty(\mathbb{R})$, we compute
$$\begin{aligned} u(x) &= u * \delta_0(x) \\ &= u * \partial_x H(x) \\ &= \partial_x u * H(x) \\ &= \int_{-\infty}^{x} \partial_x u(y)\,\mathrm{d}y, \end{aligned}$$
which is the desired representation formula.
  - *Representation formula for a u solving a boundary value problem.* Consider an open interval $I = (a, b)$ and suppose $u \in C^\infty(\overline{I})$. We obtain
$$\begin{aligned} u(x) &= u * \delta_0(x) \\ &= \mathbf{1}_I u * \partial_x H(x) \\ &= \partial_x(\mathbf{1}_I u) * H(x) \\ &= ((\delta_a - \delta_b)u) * H(x) + (\mathbf{1}_I \partial_x u) * H(x) \\ &= u(a) + \int_a^x \partial_x u(y)\,\mathrm{d}y. \end{aligned}$$

  The cases of other fundamental solutions $H(x) + C$ are left as an exercise. One interesting case is when $C = -1$, in which case $\operatorname{supp}(H(x) - 1) \subseteq (-\infty, 0]$; in this case, the formulas for $u(x)$ will be integrated only to the right of x.

---

[5] We have already seen a weaker statement along this direction. Given a fundamental solution E_0 satisfying $\mathcal{P}E_0 = \delta_0$, note that $u[f] = f * E_0$ makes sense for every $f \in C_c^\infty$. Thus, we have the existence of a smooth solution $u[f]$ to $\mathcal{P}u = f$. On the other hand, in HW#1, we saw that there exists $f \in C_c^\infty(\mathbb{R}^2)$ such that the scalar linear PDE
$$\left((\partial_t^2 + t^2\partial_x^2)^2 + \partial_y^2\right) u = f$$
does not have any C^4 solutions near 0.

- *The operator ∂_x^k on $\mathbb{R}$:* A fundamental solution for ∂_x^k (which has constant coefficients) on $\mathbb{R}$ is $E_0 := \frac{1}{(k-1)!} x^{k-1} H(x)$:

$$\partial_x^k \left(\frac{1}{(k-1)!} x^{k-1} H(x) \right) = \delta_0.$$

Moreover, any fundamental solution $E(x)$ has the property that $\partial_x^k (E - E_0) = 0$. Thus, a general fundamental solution is given by $E(x) = E_0(x) + \sum_{j=0}^{k-1} c_j x^j$.

Let us carry out the strategies outlined above for this simple example using the fundamental solution $\frac{1}{(k-1)!} x^{k-1} H(x)$; the general case is again left out as an exercise.

– *Existence.* For $f \in C_c^\infty(\mathbb{R})$, we define

$$u(x) := f * E_0(x).$$

Again, $\partial_x^k u = (f * \partial_x^k E_0) = f$, as desired.

– *Representation formula for a "nice" u in $\mathbb{R}$.* For $u \in C_c^\infty(\mathbb{R})$, we compute

$$\begin{aligned} u(x) &= u * \delta_0(x) = u * \partial_x^k E_0 = \partial_x^k u * E_0 \\ &= \frac{1}{(k-1)!} \int_{-\infty}^x \partial_x^k u(y) (x-y)^{k-1} \, \mathrm{d}y, \end{aligned}$$

which is the desired representation formula.

– *Representation formula for a u solving a boundary value problem.* Consider an open interval $I = (a, b)$, and suppose $u \in C^\infty(\overline{I})$. We now start computing as before, but move ∂_x from E_0 to u carefully so that at most one derivative falls on $\mathbf{1}_I$ each time:

$$\begin{aligned} u(x) &= u * \delta_0(x) \\ &= \mathbf{1}_I u * \partial_x^k E_0(x) \\ &= \partial_x (\mathbf{1}_I u) * \partial_x^{k-1} E_0(x) \\ &= ((\delta_a - \delta_b) u) * \partial_x^{k-1} E_0(x) + (\mathbf{1}_I \partial_x u) * \partial_{x^{k-1}} E_0(x) \\ &= ((\delta_a - \delta_b) u) * \partial_x^{k-1} E_0(x) + \partial_x (\mathbf{1}_I \partial_x u) * \partial_{x^{k-2}} E_0(x) \\ &= \cdots = \sum_{j=0}^{k-1} ((\delta_a - \delta_b) \partial_x^j u) * \partial_x^{k-j-1} E_0(x) + (\mathbf{1}_I \partial_x^k u) * E_0(x). \end{aligned}$$

Note also that $\partial_x^{k-j-1} E_0 = \frac{1}{j!} x^j H(x)$. It follows that

$$u(x) = \sum_{j=0}^{k-1} \frac{1}{j!} \partial_x^j u(a) (x-a)^j + \frac{1}{(k-1)!} \int_a^x \partial_x^k u(y) (x-y)^{k-1} \, \mathrm{d}y. \tag{3.6}$$

The representation formula is nothing but the Taylor expansion of u at a to order $k-1$, with the integral form of the remainder!

Soon, we will carry out (some of) the strategies outlined above for the important second order scalar PDEs, namely the Laplace, the heat and the wave equations (the Schrödinger equation will be discussed after we introduce the Fourier transform).

4. The Laplace equation

The subject of this section is the Laplacian on $\mathbb{R}^d$,

$$-\Delta u = -\sum_{j=1}^{d} \partial_j^2 u$$

and the associated Laplace (or Poisson) equation,

$$-\Delta u = 0 \quad (\text{or } -\Delta u = f).$$

In this section, we will focus on finding a fundamental solution E_0 for $-\Delta$, based on the symmetries enjoyed by $-\Delta$, and then try to derive key properties of solutions to the Laplace equation using fundamental solutions by following the strategies in Section 3.11. Other important ways to study the Laplace equation, namely the *Fourier and energy (or variational) methods*, will be discussed later.

A remark on the conventions. In this section, we will refer to the equation $-\Delta u = f$ as the *inhomogeneous Laplace equation* rather than by the special name Poisson equation, in order to be consistent with the discussion of other linear equations.

4.1. Symmetries of $-\Delta$ and an explicit fundamental solution.

Although the existence of a fundamental solution can be proved through abstract means (see, for instance, Remark 3.32), there is no general recipe for actually computing it. In practice, we need to make an educated guess.

In the case of the Laplacian $-\Delta$, our strategy for finding a fundamental solution will be to make use of the great number of symmetries of $-\Delta$ to narrow down the class of candidates. Since $-\Delta$ is a constant coefficient partial differential operator, it is clearly invariant under the translations $x \mapsto x - x_0$, i.e.,

$$-\Delta(u(x - x_0)) = (-\Delta u)(x - x_0).$$

Recall from Section 3.11 that this property implies that it suffices to look for a fundamental solution for $-\Delta$ at 0, i.e.,

$$-\Delta E_0 = \delta_0 \quad \text{in } \mathbb{R}^d. \tag{4.1}$$

Another important class of symmetries of $-\Delta$ is *rotations*: If R is a $d \times d$ orthogonal matrix (i.e., $R^\top = R^{-1}$) with $\det R = 1$ (i.e., a rotation matrix on $\mathbb{R}^d$), then

$$-\Delta(u(Rx)) = -\Delta u(Rx).$$

Note also that δ_0 is invariant under rotations, in the sense of the adjoint method:

$$\langle \delta_0, \phi(R\cdot)\rangle = \langle \delta_0, \phi(\cdot)\rangle \quad \text{for any } \phi \in C_c^\infty(\mathbb{R}^d).$$

Thus it is natural[6] to *look for a fundamental solution E_0 that is also invariant under rotations* (i.e., radial).

Finally, to pin down a fundamental solution E_0, let us make the bold (unjustified at the moment) assumption that E_0 *agrees with a smooth radial function $E_0(r)$ outside* $\{0\}$. Multiplying (4.1) by the characteristic function $\mathbf{1}_{B(0,r)}$ and integrating

[6] In fact, if one is familiar with the theory of Haar measure on compact Lie groups, then one can argue that if there exists any solution $\tilde{E}_0$ to (4.1), we can average $\tilde{E}_0 \circ R$ for $R \in SO(d)$ using the Haar measure to produce E_0 that is rotationally invariant!

(which can be thought of as testing the compactly supproted distribution against 1; see Lemma 3.24),

$$\int (-\Delta E_0) \mathbf{1}_{B(0,r)} = \int \delta_0 \mathbf{1}_{B(0,r)}.$$

The RHS equals $\mathbf{1}_{B(0,r)}(0) = 1$. The LHS can be computed as follows:

$$\begin{aligned} \int (-\Delta E_0) \mathbf{1}_{B(0,r)} &= \sum_j \int \partial_j (-\partial_j E_0 \mathbf{1}_{B(0,r)}) + \partial_j E_0 \partial_j \mathbf{1}_{B(0,r)} \\ &= -\int \nu \cdot D E_0 \mathrm{d}S_{\partial B(0,r)} \\ &= -|\partial B(0,r)| \partial_r E_0(r). \end{aligned}$$

For the second equality, we used Proposition 3.23 (essentially the divergence theorem) and Lemma 3.25; for the third equality, we used our assumption that E_0 agrees with a smooth radial function $E_0(r)$ outside the origin. It follows that

$$\partial_r E_0(r) = -\frac{1}{|\partial B(0,r)|} = -\frac{1}{d\alpha(d)} \frac{1}{r^{d-1}}. \tag{4.2}$$

Integrating this equation in r, we obtain

$$E_0(r) = \begin{cases} -\frac{1}{2\pi} \log r & d = 2 \\ \frac{1}{d(d-2)\alpha(d)} \frac{1}{r^{d-2}} & d \geq 3. \end{cases}$$

At this point, we can check that $E_0(r)$ indeed solves (4.1) and is locally integrable near 0 (so that it is a distribution).

Remark 4.1. Note that our derivation is nothing but a distribution-theoretic re-do of the discussion in Section 3.1; E_0 is the electrostatic potential associated to a point unit charge at 0.

Remark 4.2. Although it is not a symmetry of $-\Delta$, another important property of $-\Delta$ is the its *homogeneity*: For any $\lambda > 0$,

$$-\Delta(u(\lambda x)) = \lambda^2 (-\Delta u)(\lambda x).$$

Here we took the shortcut as in Section 3.1, but a more systematic way to derive E_0 would have been to make use of homogeneity to narrow down the candidate for E_0. This strategy will be carried out for the wave equation in Section 7.

4.2. **Uses of the fundamental solution E_0.** We now discuss various applications of the fundamental solution E_0 that we just found. Note that most of these applications require very soft properties of the fundamental solution, which means that we can often choose a different fundamental solution adapted to a problem. Indeed, we will use this freedom to prove the mean value property of harmonic functions (see Theorem 4.9).

- **Existence for the problem $-\Delta u = f$ in $\mathbb{R}^d$.** For a compactly supported distribution f, the formula
$$u[f] = E_0 * f$$
defines a solution to the Laplace equation.
- **Uniqueness (or a representation formula) for compactly supported u.** For a compactly supported distribution u on $\mathbb{R}^d$,
$$u = E_0 * (-\Delta u).$$

Indeed,

$$u = \delta_0 * u = (-\Delta E_0) * u = E_0 * (-\Delta u),$$

where the last equality is justified since u is compactly supported (recall our discussion on the convolution of two distributions).

- **Smoothness.** From the representation formula. If $-\Delta u = 0$, then

$$-\Delta(\chi u) = (-\Delta\chi)u + 2D\chi \cdot Du.$$

$$u(x) = \chi(x)u(x) = E_0 * (-\Delta)(\chi u) = \int E_0(x-y)(-\Delta(\chi u))(y)\,\mathrm{d}y.$$

Now, note that $E_0(x - \cdot)$ is smooth away from $\{x\}$, and $-\Delta(\chi u)$ is supported away from x.

Theorem 4.3 (Smoothness of harmonic functions)**.** *If $u \in \mathcal{D}'(U)$ is a solution to $-\Delta u = 0$ in U in the sense of distributions, then u is smooth in U.*

Proof. Let $x_0 \in U$, and consider a smooth function χ such that $\chi = 1$ in a ball $B(x_0, \delta)$, where $\delta > 0$ is small enough so that $\overline{B(x_0, \delta)} \subset U$, $\operatorname{supp}\chi$ is compact and $\operatorname{supp}\chi \subset U$. We will show that u is smooth in a smaller ball $B(x_0, \frac{1}{4}\delta)$.

Even though u is only defined in U, after multiplying by χ, χu is a compactly supported distribution on $\mathbb{R}^d$. We have the representation formula

$$\chi u = E_0 * (-\Delta)(\chi u).$$

Indeed, since χu is a compactly supported distribution on $\mathbb{R}^d$, the convolution $E_0 * (-\Delta)(\chi u)$ is well-defined and

$$E_0 * (-\Delta)(\chi u) = ((-\Delta)E_0) * (\chi u) = \delta_0 * (\chi u) = \chi u.$$

Next, note that

$$(-\Delta)(\chi u) = ((-\Delta)\chi)u - 2D\chi \cdot Du + \chi(-\Delta)u = ((-\Delta)\chi)u - 2D\chi \cdot Du.$$

Observe that each term on the RHS has at least one derivative falling on χ. Therefore, it vanishes on $B(x_0, \delta)$ since χ is constant there. It follows that $\operatorname{supp}(-\Delta)(\chi u) \subseteq \mathbb{R}^d \setminus B(x_0, \delta)$.

Let $\tilde{\chi}$ be a smooth function that equals 1 on $B(0, \frac{1}{4}\delta)$ and $\operatorname{supp}\tilde{\chi} \subset B(0, \frac{1}{2})$. By Lemma 3.12 and elementary geometry, we see that

$$\begin{aligned}
\operatorname{supp}(\tilde{\chi}E_0 * (-\Delta)(\chi u)) &\subseteq \operatorname{supp}\tilde{\chi}E_0 + \operatorname{supp}(-\Delta)(\chi u) \\
&\subseteq \operatorname{supp} B(0, \frac{1}{2}\delta) + (\mathbb{R}^d \setminus B(x_0, \delta)) \\
&\subseteq \mathbb{R}^d \setminus B(x_0, \frac{1}{2}\delta),
\end{aligned}$$

so $\tilde{\chi}(\cdot - x_0)(\tilde{\chi}E_0 * (-\Delta)(\chi u)) = 0$. Thus,

$$\begin{aligned}
\tilde{\chi}(\cdot - x_0)u &= \tilde{\chi}(\cdot - x_0)(E_0 * (-\Delta)(\chi u)) \\
&= \tilde{\chi}(\cdot - x_0)((1 - \tilde{\chi})E_0 * (-\Delta)(\chi u)).
\end{aligned}$$

Observe, finally, that $(1-\tilde{\chi})E_0$ is smooth since $1-\tilde{\chi}$ vanishes near $\operatorname{sing\,supp} E_0 = \{0\}$. It follows that the RHS is smooth, from which it follows that u is smooth in $B(x_0, \frac{1}{4}\delta)$, in which $u(x) = \tilde{\chi}(x - x_0)u(x)$. Since x_0 and $\delta > 0$ were arbitrary, smoothness of u in U follows. ☐

In what follows, we will call a solution u to $-\Delta u$ a *harmonic function* (that u is always a function follows from Theorem 4.3).

- **Derivative estimates.** The last proof can be made quantitative (i.e., in the form of an inequality for u) in the following way.

Theorem 4.4 (Derivative estimates)**.** *Let u be a harmonic function on U. Then for any ball $B(x,r)$ such that $\overline{B(x,r)} \subset U$, we have*

$$|D^{\alpha} u(x)| \leq \frac{C_k}{r^{d+|\alpha|}} \|u\|_{L^1(B(x_0,r))}.$$

Proof. Let $x_0 \in U$ and define $\chi_r = \chi((x-x_0)/r)$, where χ is a smooth function that equals 1 on $B(0,\frac{1}{2})$ and $\operatorname{supp}\chi \subset B(0,1)$. Starting from the representation formula for $\chi_r u$, we compute

$$\begin{aligned}
\chi_r u(x) &= E_0 * (-\Delta)(\chi_r u)(x) \\
&= E_0 * \left((-\Delta\chi_r) u - 2(D\chi_r)\cdot Du\right)(x) \\
&= E_0 * (\Delta\chi_r) u(x) - E_0 * D\cdot((D\chi_r)u)(x) \\
&= E_0 * (\Delta\chi_r) u(x) - \sum_{j=1}^{d} \partial_j E_0 * (\partial_j \chi_r) u(x).
\end{aligned}$$

Note that we moved all derivatives away from u on the RHS. Taking D^{α} and evaluating at $x = x_0$, we arrive at

$$\begin{aligned}
D^{\alpha} u(x_0) &= D^{\alpha}(\chi_r u)(x_0) \\
&= D^{\alpha} E_0 * (\Delta\chi_r) u(x_0) - \sum_{j=1}^{d} D^{\alpha}\partial_j E_0 * (\partial_j\chi_r) u(x_0) \\
&= \int D^{\alpha} E_0(y)(\Delta\chi_r) u(x_0 - y)\,\mathrm{d}y \\
&= -\sum_{j=1}^{d} \int D^{\alpha}\partial_j E_0(y)(\partial_j\chi_r) u(x_0-y)\,\mathrm{d}y.
\end{aligned}$$

Observe that $\operatorname{supp}\Delta\chi_r$ and $\operatorname{supp}\partial_j\chi_r$ are contained in the annulus $A_r := \{y \in \mathbb{R}^d : \frac{1}{2} < |y - x_0| < 1\}$. Then by the estimate

$$\sup_{y\in A_r} |D^{\beta} E_0(y)| \leq C_{\beta} r^{-|\beta|-d+2},$$

as well as the relation

$$\sup_{y\in A_r} |D^{\beta}\chi_r(x_0 - y)| = r^{-|\beta|} \sup_{y:\frac{1}{2}<|y|<1} |D^{\alpha}\chi(y)|,$$

we obtain

$$|D^{\alpha} u(x_0)| \leq C_{\alpha} r^{-d+|\alpha|} \int_{A_r} |u|(x_0 - y)\,\mathrm{d}y \leq C_{\alpha} r^{-d+|\alpha|} \int_{B(x_0,r)} |u|,$$

which implies the desired estimate. □

Remark 4.5. A similar strategy yields real-analyticity of harmonic functions; here, the key property of E_0 is that it is analytic near every point $x \neq 0$.

- **Liouville's theorem.** From the derivative estimate, we obtain the celebrated Liouville theorem for harmonic functions:

Theorem 4.6 (Liouville theorem)**.** *Suppose that u is a harmonic function on the whole space $\mathbb{R}^d$ that is bounded. Then u is constant.*

Proof. Let $M = \sup_{y\in\mathbb{R}^d} |u|$, which is finite by hypothesis. For any $x \in \mathbb{R}^d$ and $r > 0$, let us apply the derivative estimate on $B(x, r)$ with $|\alpha| = 1$. Then

$$|Du(x)| \leq Cr^{-d-1} \int_{B(x,r)} |u| \leq CMr^{-1}$$

Since u is harmonic on $\mathbb{R}^d$, we can take $r \to \infty$, which implies that $Du(x) = 0$. Since x may be chosen arbitrarily, it follows that Du vanishes and thus u is constant. □

As a consequence, we can classify solutions u to $-\Delta u = f$ in $\mathbb{R}^d$ that "behave nicely" at infinity:

Corollary 4.7 (Representation formula on $\mathbb{R}^d$)**.** *Let $f \in C(\mathbb{R}^d)$ be compactly supported.*

(1) Let $d \geq 3$. Then any bounded solution of $-\Delta u = f$ has the form

$$u = E_0 * f + c$$

for some constant c.

(2) Let $d = 2$. Then any locally integrable solution u of $-\Delta u = f$ satisfying the condition

$$\sup_{x\in\mathbb{R}^d} |Du(x)| < \infty$$

has the form

$$u = E_0 * f + \sum_j b_j x^j + c$$

for some constants $b_1, \ldots, b_d$ and c.

Note that in the case $d = 2$, the condition $\sup_{x\in\mathbb{R}^d} |Du(x)| < \infty$ implies, by the fundamental theorem of calculus, that u obeys the growth condition $|u(x)| \leq C(1 + |x|)$ for some constant $C > 0$.

Proof. When $d \geq 3$, $u[f] = E_0 * f$ is bounded; thus the desired theorem follows by applying Theorem 4.6 to $u - u[f]$.

When $d = 2$, $Du[f] = DE_0 * f$ is bounded. Therefore, $v := u - u[f]$ is a harmonic function on $\mathbb{R}^d$ such that Dv is bounded. Since each component of Dv is also harmonic, we can apply Theorem 4.6 to conclude that Dv is constant. Then the desired conclusion follows. □

Finally, we turn to the celebrated *mean value property* (see Theorem 4.9 for the statement) for harmonic functions. For us, it will be a consequence of the representation formula for boundary value problems as in Section 3.11, which is derived in the following lemma:

Lemma 4.8. *Let U be a bounded C^1 domain and $u \in C^\infty(\overline{U})$. Let $\tilde{E}_0$ be a fundamental solution for $-\Delta$ at 0. Then for $x \in U$*

$$\begin{aligned} u(x) = &\int_U \tilde{E}_0(x-y)(-\Delta u)(y)\,\mathrm{d}y - \int_{\partial U} \nu(y) \cdot D_y \tilde{E}_0(x-y) u(y)\,\mathrm{d}S(y) \\ &+ \int_{\partial U} \tilde{E}_0(x-y)\nu(y) \cdot Du(y)\,\mathrm{d}S(y), \end{aligned}$$

where $\nu(y)$ is the unit outer normal vector to ∂U at $y \in \partial U$.

By approximation, this formula can be generalized to u that only satisfies $C^2(U) \cap C^1(\overline{U})$, but we will not carry out the details here.

Proof. We begin with the observation that for any fundamental solution $\tilde{E}_0$ for $-\Delta$ at 0, $\operatorname{sing\,supp} \tilde{E}_0 = \{0\}$, just like E_0; indeed, $\tilde{E}_0 - E_0$ is a harmonic function on $\mathbb{R}^d$, which is smooth by Theorem 4.3.

Now we compute

$$\begin{aligned} u\mathbf{1}_U &= (u\mathbf{1}_U) * ((-\Delta)\tilde{E}_0) \\ &= -\sum_{j=1}^{d} (\partial_j u \mathbf{1}_U) * (\partial_j \tilde{E}_0) - \sum_{j=1}^{d} (u \partial_j \mathbf{1}_U) * (\partial_j \tilde{E}_0) \\ &= (-\Delta u \mathbf{1}_U) * \tilde{E}_0 - \sum_{j=1}^{d} (\partial_j u \partial_j \mathbf{1}_U) * \tilde{E}_0 - \sum_{j=1}^{d} (u \partial_j \mathbf{1}_U) * (\partial_j \tilde{E}_0), \end{aligned}$$

which are all justified since $\mathbf{1}_U$ is compactly supported (see Proposition 3.28). Recall Proposition 3.23, which says $\partial_j \mathbf{1}_U = -\nu_j \mathrm{d}S_{\partial U}$. Since $\operatorname{sing\,supp} \tilde{E}_0$ and $\operatorname{supp} \partial_j \mathbf{1}_U = \partial U$ does not contain $x \in U$, it follows that the last two terms are smooth near x and

$$\begin{aligned} &- \sum_{j=1}^{d} (\partial_j u \partial_j \mathbf{1}_U) * \tilde{E}_0 - \sum_{j=1}^{d} (u \partial_j \mathbf{1}_U) * (\partial_j \tilde{E}_0) \\ &= \sum_{j=1}^{d} \int_{\partial U} \nu_j(y) \partial_j u(y) \tilde{E}_0(x-y) \, \mathrm{d}S(y) + \sum_{j=1}^{d} \int_{\partial U} \nu_j(y) u(y) (\partial_j \tilde{E}_0)(x-y) \, \mathrm{d}S(y) \\ &= \sum_{j=1}^{d} \int_{\partial U} \nu_j(y) \partial_j u(y) \tilde{E}_0(x-y) \, \mathrm{d}S(y) - \sum_{j=1}^{d} \int_{\partial U} \nu_j(y) u(y) \partial_{y^j} \tilde{E}_0(x-y) \, \mathrm{d}S(y), \end{aligned}$$

as desired. □

We are now ready to prove the celebrated mean-value property of harmonic functions:

Theorem 4.9. *Let u be a harmonic function on U. Then for any ball $B(x,r)$ such that $\overline{B(x,r)} \subset U$, we have*

$$u(x) = \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} u \, \mathrm{d}S \tag{4.3}$$

$$= \frac{1}{|B(x,r)|} \int_{B(x,r)} u \, \mathrm{d}y. \tag{4.4}$$

As we will see in the proof, the key point is that $\partial B(0,r)$ is a level hypersurface of the fundamental solution $E_0(y)$.

Proof. First, we note that the identity $u(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} u \, \mathrm{d}y$ is a straightforward consequence of (4.3). Indeed,

$$\int_{B(x,r)} u \, \mathrm{d}y = \int_0^r \int_{\partial B(x,r')} u \, \mathrm{d}S(y) \, \mathrm{d}r' = u(x) \int_0^r \int_{\partial B(x,r')} \mathrm{d}S(y) \, \mathrm{d}r' = u(x) |B(x,r)|.$$

Let us focus on proving (4.3). Let $\tilde{E}_0$ be a fundamental solution for $-\Delta$ at 0. We begin by applying Lemma 4.8 with $U = B(x,r)$, which gives

$$u(x) = \int_{B(x,r)} \tilde{E}(x-y)(-\Delta u)(y)\,\mathrm{d}y - \int_{\partial B(x,r)} \nu(y)\cdot D_y \tilde{E}_0(x-y) u(y)\,\mathrm{d}S(y)$$
$$+ \int_{\partial B(x,r)} \tilde{E}_0(x-y)\nu(y)\cdot Du(y)\,\mathrm{d}S(y).$$

The first term vanishes since u is harmonic. To kill the last term, we choose $\tilde{E}_0(y) = E_0(|y|) - E_0(r)$ so that $\tilde{E}_0(x-y)$ vanishes on the sphere $\partial B(x,r)$. To compute out the second term, we note that $\nu(y) = \frac{y-x}{r}$ and

$$-\nu(y)\cdot D_y \tilde{E}_0(x-y) = -\frac{y-x}{r}\cdot\left(-\frac{x-y}{r}E_0'(r)\right) = -E_0'(r)$$

on $\partial B(x,r)$. Recalling that $E_0'(r) = -\frac{1}{d\alpha(d)r^{d-1}} = -\frac{1}{|\partial B(0,r)|}$, the mean value property follows. □

Remark 4.10. If we directly apply Lemma 4.8 with the above choice of $\tilde{E}_0$ for an arbitrary smooth function u, then

$$u(x) = \frac{1}{|\partial B(x,r)|}\int_{\partial B(x,r)} u(y)\,\mathrm{d}S(y)$$
$$+ \int_{B(x,r)} (-\Delta u)(y)\left(E_0(|x-y|) - E_0(r)\right)\mathrm{d}y.$$

This formula can be justified provided that $-\Delta u$ is continuous in a neighborhood of x, so that the last term makes sense. It is useful for showing the converse of the mean value property, i.e., a smooth function u is harmonic in U if and only if it obeys the mean value property. See also Remark 4.31 for a further application.

4.3. **Maximum principles and Harnack's inequality.** From the mean-value property, we obtain the so-called *maximum principles* for harmonic functions:

Theorem 4.11 (Maximum principles). *Suppose $u \in C^2(U)\cap C(\overline{U})$ is harmonic in U.*

(1) **Weak maximum principle.** *We have*

$$\max_{\overline{U}} u = \max_{\partial U} u.$$

(2) **Strong maximum principle.** *Moreover, if U is connected and there exists $x_0 \in U$ such that*

$$u(x_0) = \max_{\overline{U}} u,$$

then u is constant in U.

Proof. Suppose that u attains a maximum at a point $x_0 \in U$, i.e., $u(x_0) = \max_{\overline{U}} u$. Then the set

$$V = \{x \in U : u(x) = \max_{\overline{U}} u\}$$

is nonempty. Clearly, V is a closed subset of V. We claim that V is open as well. Then by connectedness, $U = V$, which proves (2). Moreover, (1) is a quick consequence of (2).

To prove that V is open, take any $x_0 \in V$. By the mean value property, for sufficiently small $r > 0$ such that $\overline{B(x_0, r)} \subset U$, we have

$$0 = u(x_0) - \frac{1}{|B(x_0,r)|} \int_{B(x_0,r)} u \, \mathrm{d}y = \frac{1}{|B(x_0,r)|} \int_{B(x_0,r)} (\max_{\overline{U}} u - u) \, \mathrm{d}y.$$

Since $\max_{\overline{U}} u - u \geq 0$ on $B(x_0, r)$, it follows that $\max_{\overline{U}} u = u$ on $B(x_0, r)$, i.e., $B(x_0, r) \subseteq V$. Hence V is open, as desired. □

From the maximum principle, we obtain:

Theorem 4.12 (Uniqueness of the Dirichlet problem)**.** *Let U be a bounded domain, $g \in C(\partial U)$ and $f \in C(U)$. There exists at most one solution $u \in C^2(U) \cap C(\overline{U})$ to the boundary value problem*

$$\begin{cases} -\Delta u = f & \text{in } U, \\ \quad u = g & \text{on } \partial U. \end{cases}$$

Remark 4.13 (For those of you familiar with the Cauch–Kovalevskaya theorem). It is important (and enlightening) to compare this result with the Cauchy–Kovalevskaya theorem [Eva10, Section 4.6.3]. Apart from the requirement of analyticity, the Cauchy–Kovalevskaya theorem requires both the boundary data $u|_\Gamma$ and the normal derivative $\nu \cdot Du|_\Gamma$ for existence and uniqueness[7], whereas Theorem 4.12 only requires the boundary data $u|_{\partial U}$. How are these facts consistent?

The key difference, which is responsible for this phenomenon, is that the Cauchy–Kovalevskaya theorem is *local* (i.e., it gives a unique solution to the boundary value problem only near the boundary portion Γ), whereas the boundary value problem in Theorem 4.12 is global (i.e., uniqueness only holds among solutions defined in the *whole* domain U). One may try to prescribe both $\tilde{u}|_{\partial U} = g$ and $\nu \cdot D\tilde{u}|_{\partial U} = h$ and appeal to Cauchy–Kovalevskaya to find a solution $\tilde{u}$ to $-\Delta \tilde{u} = f$ in U. What will happen is that unless $\nu \cdot D\tilde{u}|_{\partial U}$ matches with the unique values given by the unique solution u defined on the whole U (uniqueness given by Theorem 4.12), $\tilde{u}$ will *not* be well-defined on the whole U.

The simplest instance of this phenomenon can be seen in the context of the second order ODE $\ddot{x} = 0$, when one compares between the initial value problem (analogous to Cauchy–Kovalevskaya) $x(a) = x_0, \dot{x}(a) = y_0$ and the boundary value problem (analogous to Theorem 4.12) $x(a) = x_0$, $x(b) = x_1$.

From the mean-value property, we can derive *Harnack's inequality*:

Theorem 4.14. *Let u be a nonnegative harmonic function on a domain U. For each connected open set V such that $\overline{V}$ is compact and $\overline{V} \subset U$, there exists a positive constant $C = C(d, V, U)$ such that*

$$\max_{\overline{V}} u \leq C \min_{\overline{V}} u.$$

Harnack's inequality should be thought of as the quantitative version of the strong maximum principle. Indeed, if u is a nonnegative harmonic function, then the strong maximum principle applied to $-u$ tells you the qualitative fact that $u(x) > 0$ for all $x \in U$; in particular, $\min_{\overline{V}} u > 0$. Harnack's inequality gives us a quantitative lower bound, in terms of $\overline{V}$ and $\sup_{\overline{V}} u$, for $\min_{\overline{V}} u$.

[7]Note that for the Laplace equation, any boundary data is noncharacteristic.

Proof. Let $r = \frac{1}{4}\operatorname{dist}(V, \partial U)$. Consider $x, y \in V$ such that $|x - y| \leq r$. Then by the triangle inequality, $B(y, r) \subseteq B(x, 2r)$. Therefore, we have

$$u(x) = \frac{1}{|B(x,2r)|}\int_{B(x,2r)} u \, \mathrm{d}z \geq \frac{1}{2^d |B(x,r)|}\int_{B(y,r)} u \, \mathrm{d}z = \frac{1}{2^d} u(y).$$

Now since V is connected and $\overline{V}$ is compact, we can cover $\overline{V}$ by finitely many (say, N-many) open balls $\{B_i\}_{i=1}^N$ of radius $\frac{r}{2}$. As we have seen, for each i, we have $u(x) \geq 2^{-d} u(y)$ for any $x, y \in B_i$. For any pair $(x, y) \in V$, we may find distinct balls $B_{i_1}, \dots, B_{i_M}$ and points $x_{i_j} \in B_{i_j}$ such that

$$x = x_{i_0}, \quad x_{i_{j-1}}, x_{i_j} \in B_{i_j} \quad (j = 1, \dots, M), \quad x_{i_M} = y.$$

Then interweaving the above bound in each ball,

$$u(x) \geq 2^{-dM} u(y) \geq 2^{-dN} u(y),$$

where we used the trivial bound $M \leq N$ (the number of balls involved $\leq$ the total number of balls). Taking the infimum in x and the supremum in y, we obtain the theorem with $C = 2^{dN}$. □

4.4. **Green's function for the Dirichlet problem.** We now turn to the discussion of Green's functions, which are fundamental solutions for the Dirichlet problem for $-\Delta$ (recall that in Theorem 4.12, we saw that the solution is unique). As we will see, they allow us to derive a representation formula for the solution to the Dirichlet problem. Moreover, under suitable assumptions, we can turn the table around and use the representation formula to write down the solution to the Dirichlet problem (Poisson's integral formula).

Let us start with the definition of a Green's function.

Definition 4.15. Let U be domain. We say that $G(x, y)$ is a *Green's function on* U if $G(\cdot, y) \in \mathcal{D}'(U) \cap C^1(\overline{U} \setminus \{y\})$ and[8]

$$\begin{cases} -\Delta G(\cdot, y) = \delta_y & \text{in } U, \\ \quad G(\cdot, y) = 0 & \text{on } \partial U. \end{cases}$$

Note that $G(x, y) - E_0(x - y)$ is harmonic in U, so it is unique (Theorem 4.12) and smooth for $x \in U \setminus \{y\}$ for each $y \in U$ (Theorem 4.3). Since E_0 is smooth outside $\{0\}$, it follows that $G(\cdot, y)$ is smooth in $U \setminus \{y\}$.

Remark 4.16 (Existence of Green's function). If we know, by some means, the existence of a solution $u \in C^\infty(U) \cap C^1(\overline{U})$ to the homogeneous Dirichlet problem

$$\begin{cases} -\Delta u = 0 \text{ in } U \\ \quad u = g \text{ on } \partial U, \end{cases} \tag{4.5}$$

for $g \in C^\infty(\partial U)$, then there exists a Green's function. Indeed, for every $y \in U$, we can solve (4.5) with $g_y(x) = -\Gamma(x - y)$ to obtain a solution $u_y(x)$ and write $G(x, y) = E_0(x - y) + u_y(x)$. Soon, we will be able to complete the circle and conclude (modulo technicalities on regularity assumptions) that the existence of a Green's function implies the existence of a solution to (4.5); so the two statements can be thought of as being equivalent.

[8]Here, we are deviating from the notation in Evans's book, but we will quickly show that the definitions here and in Evans's book are equivalent.

A sufficient condition for the existence of Green's function as in Definition 4.15 is that U is a bounded $C^{1,\alpha}$ domain[9] [10]. Existence theory for (4.5) is known for much rougher domains (e.g., C^1 or even Lipschitz), but the regularity of the solution u near boundary is much worse. So in such rough domains, Green's function $G(x,y)$ can still be constructed according to the above procedure, but its behavior for $x \in \partial U$ be more delicate (in particular, it may not be in $C^1(\overline{U} \setminus y)$).

Uniqueness and symmetry of Green's function. Interestingly, the existence of a Green's function gives another proof of uniqueness. Along the way, we also obtain the useful result that G is symmetric in x, y (i.e., $G(x,y) = G(y,x)$). Both results are ultimately due to the fact that $-\Delta$ is symmetric (i.e., $\langle -\Delta u, v\rangle = \langle u, -\Delta v\rangle$ for $u, v \in C^\infty_c(U)$).

Lemma 4.17. *Suppose that there exists a Green's function $G(x,y)$ on a C^1 domain U. Then*

$$G(x,y) = G(y,x) \quad \text{for any } x, y \in U.$$

As a corollary, we see that

$$\begin{cases} -\Delta G(x,\cdot) = \delta_x & \text{in } U, \\ G(x,\cdot) = 0 & \text{on } \partial U. \end{cases} \tag{4.6}$$

Note that this is the adjoint of the problem in Definition 4.15.

Lemma 4.18. *Suppose that there exists a Green's function $G(x,y)$ on a C^1 domain U. If $G'(x,y)$ is also a Green's function on U, then*

$$G(x,y) = G'(x,y).$$

Proof of Lemmas 4.17 and 4.18. We follow the ideas used in the derivation of a representation formula in a boundary value problem. Formally, the manipulation we wish to perform is as follows: For any two Green's functions G', G on U,

$$\begin{aligned} G'(x,y) &= \delta_x[G'(\cdot,y)] \\ &= \int_U (-\Delta_z G(z,x)) G'(z,y) \, \mathrm{d}z \\ &= \int_U G(z,x)(-\Delta_z G'(z,y)) \, \mathrm{d}z \\ &\quad - \sum_{j=1}^d \int G(z,x) \partial_{z^j} G'(z,y) \partial_{z^j} \mathbf{1}_U \, \mathrm{d}z + \sum_{j=1}^d \int \partial_{z^j} G(z,x) G'(z,y) \partial_{z^j} \mathbf{1}_U \, \mathrm{d}z \\ &= \delta_y[G(\cdot,x)] \\ &= G(y,x). \end{aligned}$$

If we apply the proof to the same Green's function, then we obtain the symmetry property of any Green's function. Then for two different Green's functions, we have $G'(x,y) = G(y,x) = G(x,y)$, which is the desired uniqueness statement.

[9]For $k \in \mathbb{N}$ and $0 < \alpha < 1$, we say that f is $C^{k,\alpha}(U)$ if f is continuously differentiable and $\sup_{x,y \in U : |x-y| \leq 1} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^\alpha} < \infty$. We say ∂U is $C^{k,\alpha}$ regular if, after suitably relabeling and reorienting the coordinate axes, ∂U locally coincides with the graph of a $C^{1,\alpha}$ function.

[10]For instance, [GT01, Problem 2.12] for the existence of a solution $u \in C^2(U) \cap C(\overline{U})$ to (4.5) with $g \in C(\partial U)$. Then by [GT01, Chapter 6, Notes], $u \in C^{1,\alpha}(\overline{U})$ provided that ∂U is $C^{1,\alpha}$ regular and $g \in C^{1,\alpha}(\partial U)$.

A precise justification of the preceding formal manipulation is as follows (alternatively, one can also proceed by approximating the two Green's functions by smooth objects via mollification). Let χ be a smooth cutoff that equals 1 on $B(0,1)$ and 0 outside $B(0,2)$. Introduce two smooth cutoffs

$$\chi_x(z) = \chi(\epsilon^{-1}(z-x)), \quad \chi_y(z) = \chi(\epsilon^{-1}(z-y)),$$

where $\epsilon > 0$ is chosen so that $\operatorname{supp}\chi_x \cap \operatorname{supp}\chi_y = \emptyset$, while $\operatorname{supp}\chi_x, \operatorname{supp}\chi_y \subset U$.

To begin with, observe that for any fixed y, $G(x,y)$ and $G'(x',y)$ are harmonic and thus smooth on $U \setminus \{y\}$. Therefore

$$\begin{aligned} G'(x,y) &= \delta_x[\chi_x G'(\cdot,y)] \\ &= \int_U (-\Delta_z)\left(\chi_x(z)G'(z,y)\right) G(z,x)\,\mathrm{d}z \\ &= \int_U (-\Delta_z)\left(\chi_x(z)G'(z,y)\right) G(z,x)\,\mathrm{d}z + \int_U (1-\chi_x)(z)G'(z,y)(-\Delta_z G(z,x))\,\mathrm{d}z \end{aligned}$$

where we used $(1-\chi_x)(-\Delta_z)G(z,x) = 0$ for the last equality. Splitting $G(z,x) = (1-\chi_y(z))G(z,x) + \chi_y(z)G(z,x)$, and using the properties $\operatorname{supp}\chi_x \cap \operatorname{supp}\chi_y = \emptyset$ and $(1-\chi_y)(-\Delta_z)G'(z,y) = 0$, the last line equals

$$\begin{aligned} &- \int_U (-\Delta_z)\left((1-\chi_x)(z)G'(z,y)\right)(1-\chi_y)(z)G(z,x)\,\mathrm{d}z \\ &+ \int_U (1-\chi_x)(z)G'(z,y)(-\Delta_z)\left((1-\chi_y)(z)G(z,x)\right)\,\mathrm{d}z \\ &+ \int_U G'(z,y)(-\Delta_z)\left(\chi_y(z)G(z,x)\right)\,\mathrm{d}z. \end{aligned}$$

The third term equals $\delta_y(\chi_y(z)G(z,x)) = G(y,x)$. For the first two terms, since the integrand is smooth thanks to the cutoffs, and since the support of the cutoffs are disjoint from ∂U, we may apply integration by parts (or equivalently Proposition 3.23) to conclude

$$\begin{aligned} &- \int_U (-\Delta_z)\left((1-\chi_x)(z)G'(z,y)\right)(1-\chi_y)(z)G(z,x)\,\mathrm{d}z \\ &+ \int_U (1-\chi_x)(z)G'(z,y)(-\Delta_z)\left((1-\chi_y)(z)G(z,x)\right)\,\mathrm{d}z \\ &= \sum_{j=1}^{d} \int G(z,x)\partial_{z^j}G'(z,y)\partial_{z^j}\mathbf{1}_U\,\mathrm{d}z + \sum_{j=1}^{d} \int \partial_{z^j}G(z,x)G'(z,y)\partial_{z^j}\mathbf{1}_U\,\mathrm{d}z \\ &= 0, \end{aligned}$$

where we used $G(z,x) = G'(z,y) = 0$ for $z \in \partial U$ on the last line. □

Green's function and existence for the Dirichlet problem (Optional). Suppose that there exists a Green's function $G(x,y)$ on U. Then following the first strategy in Section 3.11, given a function f on U, the formula

$$u[f](x) = \int_U G(x,y)f(y)\,\mathrm{d}y$$

should give us a solution to the inhomogeneous Dirichlet problem

$$\begin{cases} -\Delta u = f & \text{in } U, \\ \quad u = 0 & \text{on } \partial U. \end{cases} \tag{4.7}$$

This procedure is not difficult to justify when f is a continuous function such that $\operatorname{supp} f \subset U$. The key step is to understand the regularity properties of $G(x,y)$. If we write

$$h(x,y) = G(x,y) - \Gamma(x-y)$$

then by the symmetry of Green's function (Lemma 4.17), we see that $h(x,y)$ is harmonic in $x \in U$ for each fixed $y \in U$ and vice versa. Going through a similar argument as in Theorem 4.3, it can be shown that $h(x,y)$ is smooth in $U \times U$. Moreover, by the regularity of $G(x,y)$ near ∂U, it follows that $h(\cdot,y) \in C^1(\overline{U})$ for each fixed $y \in U$ and vice versa (by Lemma 4.17). These properties are sufficient to justify the definition and the desired properties of $u[f]$ when $f \in C(U)$ and $\operatorname{supp} f \subset U$.

The formula still works for f that is nontrivial on ∂U, but in order to justify the desired properties (especially, to prove $u[f](x) \to 0$ as $x \to x_0 \in \partial U$) we need to know more about the behavior of $G(x,y)$ as both x, y approach a same boundary point $x_0 \in \partial\Omega$. Let us not go deeper into this issue here.

Remark 4.19 (From homogeneous to inhomogeneous Dirichlet problem). Let us point out that, by a fairly general trick, solving the homogeneous Dirichlet problem (4.5) can be reduced to solving the inhomogeneous Dirichlet problem with zero boundary data (4.7), at least when the boundary values $g(x)$ are smooth enough. The idea is to first find an extension $\tilde{g}$ of g to U, and then consider $v = u - \tilde{g}$. Then (4.5) transforms to $-\Delta v = -\Delta \tilde{g}$ in U and $v = 0$ on ∂U, which is in the same form as (4.7).

This shows that the existence of Green's function is *essentially* equivalent to solvability of (4.5) or (4.7), modulo specific regularity assumptions and properties of $G(x,y)$ as $x, y \to \partial\Omega$.

Representation formula for the Dirichlet problem (Poisson integral formula). We now derive a representation formula for the solution to the Dirichlet problem using Green's function.

Theorem 4.20 (Poisson integral formula). *Let U be a C^1 domain and suppose that there exists a Green's function $G(x,y)$ on U. Then for any $u \in C^2(U) \cap C(\overline{U})$, we have*

$$u(x) = -\int_{\partial U} u(y)\nu(y) \cdot D_y G(x,y) \, \mathrm{d}S(y) + \int_U (-\Delta u)(y) G(x,y) \, \mathrm{d}y.$$

In case $-\Delta u = 0$, this representation formula is often called the *Poisson integral formula* for harmonic functions, and the function $K(x,y) := \nu(y) \cdot D_y G(x,y)$ on ∂U is called the *Poisson kernel*.

Proof. In the following computation, all derivatives are taken with respect to y. First, we assume that $u \in C^\infty(\overline{U})$ and repeat the derivation of the representation formula for a boundary value problem. Formally, we manipulate as follows:

$$\begin{aligned} u(x) &= \int \delta_x(y) u(y) \mathbf{1}_U(y) \, \mathrm{d}y \\ &= \int (-\Delta G(x,y)) u(y) \mathbf{1}_U(y) \, \mathrm{d}y \\ &= \sum_{j=1}^d \int \partial_j G(x,y) \partial_j u(y) \mathbf{1}_U(y) \, \mathrm{d}y \end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=1}^{d}\int \partial_j G(x,y)u(y)\partial_j \mathbf{1}_U(y)\,\mathrm{d}y\\
=&\int G(x,y)(-\Delta u)(y)\mathbf{1}_U(y)\,\mathrm{d}y\\
&-\sum_{j=1}^{d}\int G(x,y)\partial_j u(y)\partial_j \mathbf{1}_U(y)\,\mathrm{d}y\\
&+\sum_{j=1}^{d}\int \partial_j G(x,y)u(y)\partial_j \mathbf{1}_U(y)\,\mathrm{d}y.
\end{aligned}
$$

Here, unlike what we had before, $G(x,y)=0$ for $y\in\partial U$. Thus the second to last term, which involves $\partial_j u$ on ∂U, vanishes. We are left with

$$u(x) = -\int_{\partial U} u(y)\nu(y)\cdot DG(x,y)\,\mathrm{d}S(y) + \int_U (-\Delta u)(y)G(x,y)\,\mathrm{d}y.$$

We leave the rigorous justification, which may proceed like the proof of Lemmas 4.17–4.18, as an exercise. The case of a more general solution u follows from approximation. □

When Green's function $G(x,y)$ is known, the Poisson integral formula suggests us a way to find a solution to the homogeneous Dirichlet problem (4.5), namely, to simply write down the Poisson integral formula

$$u(x) = -\int_{\partial U} \nu(y)\cdot D_y G(x,y)g(y)\,\mathrm{d}S(y)$$

and check that it is a solution. This procedure works for a wide class domains and g [Dah79], but its justification requires more information about the behavior of $\nu(y)\cdot D_y G(x,y)$ as x approaches ∂U than we have right now. Instead, we will concentrate on simple examples of U, for which $G(x,y)$ can be written down explicitly, and then verify this assertion on a case-by-case basis.

Remark 4.21. We remark that the existence of a representation formula does *not* guarantee the existence of a solution to a boundary value problem. Recall, for instance, that in Complex Analysis, the Cauchy integral formula expresses any solution f to the Cauchy–Riemann equation in U in terms of the data $f|_{\partial U}$, but not every continuous function g on ∂U can be the boundary values of a holomorphic function (e.g., take $U=B(0,1)$ and $g=e^{-i\theta}$ on $\partial B(0,1)$).

In the case of the Laplace equation, it is ultimately because of the symmetry of $-\Delta$ that uniqueness (which follows from having a representation formula) is equivalent to existence!

Computation of Green's function for some domains: Method of image charges. For some domains U, Green's function can be constructed by the *method of image charges*. Using the analogies in electrostatics, this method can be summarized as follows[11]:

[11] Note that, as in Evans's book, we are solving the adjoint problem (4.6) to find Green's function, which is equivalent to Definition 4.15 thanks to Lemma 4.17. This choice is more convenient here, because we will be using it in the context of a representation formula (Poisson integral formula).

- To construct $G(x,y)$, start with the potential $E_0(y-x)$ corresponding to a unit point charge at $x \in U$.
- Place other point charges outside U (image charges), with charges q_j and locations $\{\bar{x}_j\}$, so that the corresponding potential $\sum_j q_j E_0(y-\bar{x}_j)$ exactly cancels $E_0(y-x)$ for $y \in \partial U$. Then $G(x,y) = E_0(y-x) + \sum_j q_j E_0(y-\bar{x}_j)$ is Green's function that we looked for.

We discuss two cases, namely when U is a half-space or a ball, which involve putting one image charge.

- *Half-space:* $U = \mathbb{R}^d_+ = \{x \in \mathbb{R}^d : x^d > 0\}$. In this case, we put an image charge with charge -1 at $\bar{x}$, where $\bar{x}$ is the reflection of x across $\partial\mathbb{R}^d_+$, i.e.,

$$\bar{x} = (x^1, \dots, x^{d-1}, -x^d).$$

Since $\partial\mathbb{R}^d_+ = \{y^d = 0\}$ is exactly the set of points that are equidistant to x and $\bar{x}$, clearly $E_0(y-x) = E_0(y-\bar{x})$ for $y \in \partial\mathbb{R}^d_+$. Therefore,

$$G(x,y) = E_0(y-x) - E_0(y-\bar{x}).$$

From the above expression, let us compute the Poisson kernel on $\partial\mathbb{R}^d_+$. Since $-\nu(y)\cdot D_y = \partial_{y^d}$ on $\partial\mathbb{R}^d_+$, we first compute $\partial_{y^d} G(x,y)$ for $x, y \in \mathbb{R}^d_+$:

$$\begin{aligned}
\partial_{y^d} G(x,y) &= \partial_{y^d}\left(E_0(y-x) - E_0(y-\bar{x})\right) \\
&= \partial_{y^d}|y-x| E_0'(|y-x|) - \partial_{y^d}|y-\bar{x}| E_0'(|y-\bar{x}|) \\
&= -\frac{1}{d\alpha(d)} \frac{y^d - x^d}{|y-x|^d} + \frac{1}{d\alpha(d)} \frac{y^d + x^d}{|y-\bar{x}|^d}.
\end{aligned}$$

Now we put $y = (y', 0)$ and write $x = (x', x^d)$, which makes $|y-x| = |y-\bar{x}| = \sqrt{|y'-x'|^2 + (x^d)^2}$. Thus,

$$-\nu(y)\cdot D_y G(x,y) = \frac{2x^d}{d\alpha(d)} \frac{1}{(|y'-x'|^2 + (x^d)^2)^{\frac{d}{2}}}.$$

The following theorem then can be directly verified:

Theorem 4.22. *Assume that $g \in C(\mathbb{R}^{d-1}) \cap L^\infty(\mathbb{R}^{d-1})$ and for $x \in \mathbb{R}^d_+$, define*

$$u(x) = \frac{2x^d}{d\alpha(d)} \int_{\mathbb{R}^{d-1}} \frac{g(y')}{(|y'-x'|^2 + (x^d)^2)^{\frac{d}{2}}} \, \mathrm{d}y'.$$

Then
(1) $u \in C^\infty(\mathbb{R}^d_+) \cap L^\infty(\mathbb{R}^d_+)$;
(2) $-\Delta u = 0$ in $\mathbb{R}^d_+$;
(3) for each point $x_0 \in \partial\mathbb{R}^d_+$, $\lim_{x\to x_0} u(x) = g(x_0)$.

The most delicate part is the proof of (3); we need to observe that as $x^d \to 0$, the Poisson kernel is an approximation to the identity. We refer to Evans's book for the details of the proof.

- *Unit ball:* $U = B(0,1)$. To construct a Green's function in this case, we use the following elementary (but very amusing!) geometric fact: Given two points $x, \bar{x} \in \mathbb{R}^d$, the set of points y such that the ratio between $|y-x|$ and $|y-\bar{x}|$ is constant is the sphere. More precisely,

Lemma 4.23. *Let x be a point in the unit ball $B(0,1)$, and define $\bar{x} = \frac{x}{|x|^2}$. Then*

$$\partial B(0,1) = \{y \in \mathbb{R}^d : |y-x| = |x||y-\bar{x}|\}.$$

Proof. Unraveling the definition $\bar{x} = \frac{x}{|x|^2}$, we compute

$$\begin{aligned} |y-x|^2 &= |y|^2 - 2x\cdot y + |x|^2, \\ |x|^2\left|y - \frac{x}{|x|^2}\right|^2 &= |x|^2\left(|y|^2 - 2\frac{x\cdot y}{|x|^2} + \frac{|x|^2}{|x|^4}\right) \\ &= |x|^2|y|^2 - 2x\cdot y + 1 \end{aligned}$$

Equating both sides,

$$\begin{aligned} |y|^2 - 2x\cdot y + |x|^2 &= |x|^2|y|^2 - 2x\cdot y + 1 \\ \Leftrightarrow (1-|x|^2)|y|^2 &= 1 - |x|^2. \end{aligned}$$

Since $|x| < 1$, it follows that the last line is equivalent to $|y| = 1$, as desired. □

So Green's function is

$$G(x,y) = E_0(y-x) - E_0\left(|x|(y-\bar{x})\right),$$

where

$$\bar{x} = \frac{x}{|x|^2}.$$

In other words, we placed an image charge at $\bar{x}$ with charge $-|x|^{2-d}$.

Let us compute the Poisson kernel on $\partial B(0,1)$. In this case, $-\nu(y)\cdot D_y = -\sum_{j=1}^d y^j \partial_{y^j}$, so we begin by computing, for $x, y \in B(0,1)$,

$$-\sum_{j=1}^d y^j \partial_{y^j} G(x,y) = -\sum_{j=1}^d y^j \partial_{y^j} E_0(y-x) + \sum_{j=1}^d y^j \partial_{y^j} E_0\left(|x|(y-\bar{x})\right).$$

Note that

$$\begin{aligned} \partial_{y^j} E_0(y-x) &= \partial_{y^j}|y-x| E_0'(y-x) \\ &= -\frac{(y^j - x^j)}{d\alpha(d)}\frac{1}{|y-x|^d}, \\ \partial_{y^j} E_0\left(|x|(y-\bar{x})\right) &= \partial_{y^j}\left(|x||y-\bar{x}|\right) E_0'\left(|x||y-\bar{x}|\right) \\ &= -\frac{(|x|^2 y^j - x^j)}{d\alpha(d)}\frac{1}{|x|^d|y-\bar{x}|^d} \end{aligned}$$

so

$$-\sum_{j=1}^d y^j \partial_{y^j} G(x,y) = \frac{|y|^2 - x\cdot y}{d\alpha(d)}\frac{1}{|y-x|^d} - \frac{|x|^2 - x\cdot y}{d\alpha(d)}\frac{1}{|x|^d|y-\bar{x}|^d}.$$

Now we restrict $y \in \partial B(0,1)$, on which $|y|^2 = 1$ and $|y-x| = |x||y-\bar{x}|$. Thus, for $x \in B(0,1)$ and $y \in \partial B(0,1)$,

$$\nu(y)\cdot D_y G(x,y) = \frac{1-|x|^2}{d\alpha(d)}\frac{1}{|y-x|^d}.$$

As before, the following theorem can be proved by computation:

Theorem 4.24. *Assume that $g \in C(\partial B(0,1))$ and for $x \in B(0,1)$, define*

$$u(x) = \frac{1-|x|^2}{d\alpha(d)} \int_{\partial B(0,1)} \frac{g(y')}{|y-x|^d} \, \mathrm{d}S(y).$$

Then
(1) $u \in C^\infty(B(0,1))$;
(2) $-\Delta u = 0$ in $B(0,1)$;
(3) for each point $x_0 \in \partial B(0,1)$, $\lim_{x \to x_0} u(x) = g(x_0)$.

We omit the proof. We remark that this result can be extended to balls of arbitrary radii by scaling.

4.5. **The Cauchy–Riemann equation and holomorphic functions.** Let us now consider an application of our strategies to the Cauchy–Riemann equation

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

where f is a complex-valued function on a domain U in $\mathbb{C} = \mathbb{R}^2$ (i.e., $f = u + iv$ where u, v are real-valued functions on U). We will also use the notation $z = x + iy$. As we will see, the very basic pillars of complex analysis (Morera's theorem, Cauchy integral formula, equivalence of complex-differentiability with complex-analyticity) follow from the strategies outlined in Section 3.11.

In this section, we work with *complex-valued distributions* on U, which are simply pairs of real-valued distributions $u, v \in \mathcal{D}'(U)$, combined in the form $f = u + iv$. Given a complex-valued test function $\phi = \operatorname{Re}\phi + i \operatorname{Im}\phi \in C^\infty_c(U; \mathbb{C})$, the pairing is defined as

$$(f, \phi) = \langle u, \operatorname{Re}\phi \rangle + \langle v, \operatorname{Im}\phi \rangle + i(\langle v, \operatorname{Re}\phi \rangle - \langle u, \operatorname{Im}\phi \rangle),$$

so that when f is a function, $(f, \phi) = \int f \overline{\phi} \, \mathrm{d}x \mathrm{d}y$. We will discuss further properties of complex-valued distributions when we discuss the Fourier transform.

Let us start by deriving the Cauchy–Riemann equation from *complex differentiability*: We say that f is *complex-differentiable* at $z \in \mathbb{C}$ if the limit

$$\lim_{w \to 0} \frac{f(z+w) - f(z)}{w}$$

exists, where w is a complex number. As usual, f is complex-differentiable on a domain $U \subseteq \mathbb{C}$ if it is complex-differentiable at every point $z \in U$.

If f is complex-differentiable on U, the limit must agree whether w approaches zero along the real axis ($w = h$ as $h \to 0$ with $h \in \mathbb{R}$) or ($w = ih$ as $h \to 0$). Thus,

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} = \lim_{h \to 0} \frac{f(z+ih) - f(z)}{ih}.$$

If we write $z = x + iy$, then the above identity becomes

$$\partial_x f(z) = \frac{1}{i} \partial_y f,$$

so rearranging terms, we arrive at the *Cauchy–Riemann equation*:

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

We will call $(\partial_x + i\partial_y)$ the Cauchy–Riemann operator.

Note the following algebraic identities:

$$(\partial_x - i\partial_y)(\partial_x + i\partial_y) f = (\partial_x + i\partial_y)(\partial_x - i\partial_y) f = \Delta f.$$

The identity $(\partial_x - i\partial_y)(\partial_x + i\partial_y)f = \Delta f$ tells us that the components u, v in $f = u + iv$ are harmonic if f solves the Cauchy–Riemann equation. The other identity, $(\partial_x + i\partial_y)(\partial_x - i\partial_y)f = \Delta f$ tells us how to construct a fundamental solution for the Cauchy–Riemann operator, from a fundamental solution for $-\Delta$. Recall from Section 4.1 that $-\frac{1}{2\pi}\log r$ is a fundamental solution for $-\Delta$, or equivalently,

$$\Delta\left(\frac{1}{2\pi}\log r\right) = \delta_0.$$

By the identity $(\partial_x + i\partial_y)(\partial_x - i\partial_y) = \Delta$, we see that $E_0 := (\partial_x - i\partial_y)\left(\frac{1}{2\pi}\log r\right)$ is a fundamental solution for $(\partial_x + i\partial_y)$. Note that

$$\begin{aligned} E_0 &= (\partial_x - i\partial_y)\left(\frac{1}{2\pi}\log r\right) \\ &= \frac{1}{2\pi}\left(\frac{x}{r^2} - i\frac{y}{r^2}\right) \\ &= \frac{1}{2\pi}\frac{1}{x+iy} = \frac{1}{2\pi}\frac{1}{z}. \end{aligned}$$

Hence, we have derived

$$(\partial_x + i\partial_y)\left(\frac{1}{2\pi}\frac{1}{z}\right) = \delta_0.$$

With the fundamental solution $E_0 = \frac{1}{2\pi}\frac{1}{z}$ in our hands, let us carry out the strategies outlined in Section 3.11. In particular, as in the case of the Laplace equation, representation formula for a "nice" u (more precisely, compactly supported), combined with the observation that E_0 is smooth outside $\{0\}$, leads to the following regularity result:

Theorem 4.25. *If $f \in \mathcal{D}'(U)$ is a solution to $(\partial_x + i\partial_y)f = 0$ in U (in the sense of distributions), then f is smooth in U.*

We omit the proof, which is very similar to Theorem 4.3. We will call a smooth solution to the Cauchy–Riemann equation *holomorphic*. It turns out that Theorem 4.25 is the main thrust behind Morera's theorem:

Corollary 4.26 (Morera's theorem)**.** *If f is a continuous function on U such that for every bounded domain Ω such that $\overline{\Omega} \subset U$ and $\partial\Omega$ is a triangle, then we have*

$$\int_{\partial\Omega} f \,\mathrm{d}z = 0, \tag{4.8}$$

then f is holomorphic in U.

In order to prove this corollary, we need to carry out the computation of $(\partial_x + i\partial_y)\mathbf{1}_\Omega$. Let us record the result as a lemma, since it will be useful again later:

Lemma 4.27. *Let U be a domain in $\mathbb{C}$ and consider a bounded piecewise C^1 domain $\Omega \subset \overline{\Omega} \subset U$. For any $\phi \in C_c^\infty(U)$, we have*

$$\int ((\partial_x + i\partial_y)\mathbf{1}_\Omega)\,\phi \,\mathrm{d}x\mathrm{d}y = \int_{\partial\Omega} i\phi(z)\mathrm{d}z.$$

On the LHS, we are using the convention of writing $\int u \,\mathrm{d}x\mathrm{d}y$ for $\langle u, 1\rangle$ when u is a distribution with a compact support. The RHS is the integral of the 1-form $i\phi(z)\,\mathrm{d}z = i\phi(z)\,(\mathrm{d}x + i\mathrm{d}y)$ on the curve $\partial\Omega$ with the induced orientation. More concretely, if $(x + iy)(t)$ $(t \in I)$ is a positively oriented (i.e., Ω is always left to the tangent vector $\dot{x} + i\dot{y}(t)$ at $(x+iy)(t)$) parametrization of $\partial\Omega$, which can be seen to be a piecewise C^1 curve by the assumption, then

$$\int_{\partial\Omega} i\phi(z)\mathrm{d}z = \int_I i\phi(x(t)+iy(t))(\dot{x}(t)+i\dot{y}(t))\,\mathrm{d}t.$$

Proof. We will only carry out the key computation when Ω is a bounded C^1 domain; the piecewise C^1 case then follows from a straightforward approximation argument. By Proposition 3.23,

$$(\partial_x + i\partial_y)\mathbf{1}_\Omega = -(\nu_x + i\nu_y)\mathrm{d}S_{\partial\Omega},$$

where $\nu_x + i\nu_y$ is the outer unit normal vector field on $\partial\Omega$ and $\mathrm{d}S_{\partial\Omega}$ is the induced measure on $\partial\Omega$. If $(x+iy)(t)$ is a positively oriented parametrization of $\partial\Omega$, we have

$$\int \varphi \mathrm{d}S_{\partial\Omega} = \int_{\partial\Omega} \varphi(x(t)+iy(t))\sqrt{\dot{x}^2+\dot{y}^2}\,\mathrm{d}t \qquad \text{for } \varphi \in C^\infty_c(U).$$

Moreover, the unit tangent vector is $\tau_x + i\tau_y = (\sqrt{\dot{x}^2+\dot{y}^2})^{-1}(\dot{x}+i\dot{y})$, so the outward unit normal vector is

$$\nu_x + i\nu_y = -i(\tau_x + i\tau_y) = -i\frac{1}{\sqrt{\dot{x}^2+\dot{y}^2}}(\dot{x}+i\dot{y}).$$

Putting all these together, the lemma follows. □

Proof of Morera's theorem. First, let us prove the equivalence of (4.8) with the Cauchy–Riemann equation when f is smooth. For every bounded domain Ω with $\overline{\Omega} \subset U$ such that $\partial\Omega$ is a triangle, we have

$$0 = \int_{\partial\Omega} f\,\mathrm{d}z = -i\int ((\partial_x + i\partial_y)\mathbf{1}_\Omega)\, f\,\mathrm{d}x\mathrm{d}y = i\int (\partial_x + i\partial_y) f \mathbf{1}_\Omega\,\mathrm{d}x\mathrm{d}y,$$

where the last equality follows from the definition of the distributional derivative (to be pedantic, we need $f \in C^\infty_c(U)$ to apply Lemma 4.27 and the definition of the distributional derivative, but here it is okay since $\mathbf{1}_\Omega$ is compactly supported). Varying the domain Ω, it is not difficult to show that $(\partial_x + i\partial_y)f = 0$, i.e., f is holomorphic.

Next, let us consider the case when f is merely continuous. Here, the strategy is to use the approximation method; we make auxiliary preparations to avoid issues near the boundary ∂U. Fix an open set V such that $\overline{V}$ is compact and $\overline{V} \subset U$. Then there exists $\delta_0 > 0$ such that $\cup_{z\in V} B(z;\delta_0) \subseteq U$. Consider the convolution $f_\delta = f * \varphi_\delta$, where $\varphi_\delta(z) = \delta^{-2}\varphi(\delta^{-1}z)$ and $\varphi \in C^\infty_c(U)$ obeys $\int \varphi = 1$ and $\operatorname{supp}\varphi \subset B(0,1)$. For any bounded domain Ω such that $\overline{\Omega} \subset V$ and $\partial\Omega$ is a triangle and $\delta \in (0,\delta_0)$, we have

$$\begin{aligned}\int_{\partial\Omega} f_\delta(z)\,\mathrm{d}z &= \int_{\partial\Omega}\left(\int f(z-z')\varphi_\delta(z')\,\mathrm{d}x'\mathrm{d}y'\right)\mathrm{d}z \\ &= \int\left(\int_{-z'+\partial\Omega} f(w)\,\mathrm{d}w\right)\varphi_\delta(z')\,\mathrm{d}x'\mathrm{d}y',\end{aligned}$$

where on the last line, we used Fubini's theorem and the change of variables $(z,z') \mapsto (w = z - z', z')$; $-z' + \partial\Omega$ is the set $\{-z' + z \in \mathbb{C} : z \in \partial\Omega\}$. Note that $\operatorname{supp}\varphi_\delta \subset B(0,\delta)$ and $-z' + \Omega \subset U$ for each $z' \in B(0,\delta) \subseteq B(0,\delta_0)$. Therefore, by (4.8) applied to each $-z'+\Omega$, the last line vanishes. It follows that for each $\delta \in (0,\delta_0)$, f_δ is a smooth function that satisfies the hypothesis of Corollary 4.26

on V; hence $(\partial_x + i\partial_y) f_\delta = 0$ on V by the first part of the proof. Then the distributional limit f also satisfies $(\partial_x + i\partial_y) f = 0$ on V (i.e., when tested against $\phi \in C_c^\infty(V)$) by Lemma 3.17. Since V is an arbitrary bounded domain such that $\overline{V} \subset U$, it follows that f is a solution to the Cauchy–Riemann equation in the sense of distributions. Finally, by Theorem 4.25, Morera's theorem follows. $\square$

The representation formula for boundary value problems in Section 3.11 leads to the Cauchy integral formula:

Theorem 4.28 (Cauchy integral formula)**.** *Let f be a holomorphic function on U. Then for every bounded piecewise C^1 domain Ω and $z_0 \in \Omega$,*

$$f(z_0) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(z)}{z - z_0} \, \mathrm{d}z.$$

Proof. As in the proof of Theorem 4.9, we begin by computing

$$f \mathbf{1}_\Omega = f \mathbf{1}_\Omega * (\partial_x + i\partial_y) E_0 \quad = ((\partial_x + i\partial_y) f \mathbf{1}_\Omega) * E_0 + (f (\partial_x + i\partial_y) \mathbf{1}_\Omega) * E_0.$$

The first term vanishes by the Cauchy–Riemann equation. Since $\operatorname{sing\,supp} E_0 = \{0\}$ and $\operatorname{supp}(\partial_x + i\partial_y)\mathbf{1}_\Omega = \partial\Omega$, it follows that the second term is smooth near z_0 and

$$\begin{aligned} f(z_0) &= f \mathbf{1}_\Omega(z_0) \\ &= \int (f(\partial_x + i\partial_y)\mathbf{1}_\Omega)(z) E_0(z_0 - z) \, \mathrm{d}x\mathrm{d}y \\ &= \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(z)}{z - z_0} \, \mathrm{d}z, \end{aligned}$$

where on the last line, we used Lemma 4.27 and $\frac{i}{z_0 - z} = \frac{1}{i(z - z_0)}$. $\square$

Remark 4.29. If we carry out the computation without using the Cauchy–Riemann equation, then we obtain the more general formula

$$f(z_0) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(z)}{z - z_0} \, \mathrm{d}z - \frac{1}{2\pi} \int_{\Omega} \frac{(\partial_x + i\partial_y) f(z)}{z - z_0} \, \mathrm{d}x\mathrm{d}y,$$

which may be justified as long as $(\partial_x + i\partial_y) f$ is continuous near z_0 (the important point is that the last term should make sense).

The Cauchy integral formula, of course, is where the magic of complex analysis begins. Here, let us end by just closing the loop that we started at the beginning:

Corollary 4.30. *Let f be a continuous function on a domain $U \subseteq \mathbb{C}$. The following statements are equivalent:*

(1) f is complex-differentiable;
(2) f is a solution to the Cauchy–Riemann equation (i.e., f is holomorphic);
(3) f is complex-analytic, i.e., at every point $z_0 \in U$, there exists $r > 0$ and coefficients $c_j \in \mathbb{C}$ such that

$$f(z) = \sum_{j=0}^{\infty} c_j (z - z_0)^j \qquad \text{for } |z - z_0| < r.$$

Proof. The following is a standard proof in complex analysis. Note that (1)⇒(2) was shown at the beginning of this subsection and (3)⇒(1) is obvious; it only remains to verify (2)⇒(3). Applying the Cauchy integral formula for $z \in B(z_0, r)$, where $r > 0$ is chosen so that $\overline{B(z_0, r)} \subset U$, we have

$$f(z) = \frac{1}{2\pi i} \int_{\partial B(z_0,r)} \frac{f(w)}{w - z} \, \mathrm{d}w.$$

Now the point is that $\frac{1}{w-z}$ is complex-analytic near z_0, from which complex-analyticity of f should follow. More precisely, we have

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \int_{\partial B(z_0,r)} \frac{f(w)}{w - z_0} \frac{1}{1 - \frac{z-z_0}{w-z_0}} \, \mathrm{d}w \\ &= \frac{1}{2\pi i} \int_{\partial B(z_0,r)} \frac{f(w)}{w - z_0} \sum_{j=0}^{\infty} \left(\frac{z - z_0}{w - z_0} \right)^j \mathrm{d}w \\ &= \sum_{j=0}^{\infty} \left(\frac{1}{2\pi i} \int_{\partial B(z_0,r)} \frac{f(w)}{(w - z_0)^{j+1}} \, \mathrm{d}w \right) (z - z_0)^j, \end{aligned}$$

where the two identities make sense as long as

$$\left| \frac{z - z_0}{w - z_0} \right| < 1 \text{ for } w \in \partial B(z_0, r),$$

or equivalently, $|z - z_0| < r$. □

Remark 4.31 (Jensen's formula). Another nice application of the results so far is a quick proof of *Jensen's formula*, which is a basic tool for relating the growth of a holomorphic function on $\mathbb{C}$ with the distribution of its zeroes. For this application, we will assume more familiarity with complex analysis.

We start by observing that if g is a holomorphic function on U with no zeroes, then

$$\Delta \left(\frac{1}{2\pi} \log |g| \right) = 0.$$

Indeed, $\frac{1}{2\pi} \log |z|$ is harmonic on $\mathbb{C} \setminus \{0\}$ and g is a holomorphic function whose image is contained in $\mathbb{C} \setminus \{0\}$; it follows that their composition $\frac{1}{2\pi} \log |g|$ is harmonic.

Next, when f is a general non-zero holomorphic function on U, then we can write

$$f(z) = g(z) \prod_k (z - \rho_k),$$

where ρ_k's are the zeroes of f in U counted with multiplicity (note that there can be only finitely many of them in each compact set, since f, being complex-analytic, cannot have accumulated zeroes) and g is holomorphic with no zeroes in U (this statement can be proved by the Cauchy integral formula and Morera's theorem). Thus

$$\Delta \left(\frac{1}{2\pi} \log |f| \right) = \Delta \left(\frac{1}{2\pi} \log |g| \right) + \sum_k \Delta \left(\frac{1}{2\pi} \log |z - \rho_k| \right) = \sum_k \delta_{\rho_k}, \tag{4.9}$$

where we used the preceding observation for g and the fact that $\frac{1}{2\pi} \log |z - \rho_k|$ is a fundamental solution for Δ at ρ_k.

Finally, we apply the general form of the mean value theorem in Remark 4.10 to (4.9). Then

$$\begin{aligned}\log|f(0)| &= \frac{1}{2\pi R}\int_0^{2\pi}\log|f(Re^{i\theta})|\,\mathrm{d}\theta \\ &\quad + \int_{B(0,R)}(-2\pi)\sum_k \delta_{\rho_k}\left(-\frac{1}{2\pi}\log|y| + \frac{1}{2\pi}\log R\right)\,\mathrm{d}y \\ &= \frac{1}{2\pi R}\int_0^{2\pi}\log|f(Re^{i\theta})|\,\mathrm{d}\theta - \sum_{k:|\rho_k|<R}\log\frac{R}{|\rho_k|}.\end{aligned}$$

Rearranging terms, we obtain

$$\frac{1}{2\pi R}\int_0^{2\pi}\log|f(Re^{i\theta})|\,\mathrm{d}\theta = \log|f(0)| + \sum_{k:|\rho_k|<R}\log\frac{R}{|\rho_k|},$$

which is the usual form of Jensen's formula.

5. Heat equation

In this section we study the heat equation

$$(\partial_t - \Delta) u = f, \tag{5.1}$$

using fundamental solutions by following the strategies in Section 3.11. One key difference between (5.1) and the Laplace/Poisson equations (as well as the Cauchy–Riemann equation) studied in Section 4 is that (5.1) is *evolutionary*. Therefore, in the case of the heat equation, we are now interested in the *initial value problem*,

$$\begin{cases} (\partial_t - \Delta) u = f & \text{in } (0, \infty)_t \times \mathbb{R}^d_x, \\ u = g & \text{on } \{t = 0\} \times \mathbb{R}^d_x, \end{cases} \tag{5.2}$$

or the *initial-boundary value problem*,

$$\begin{cases} (\partial_t - \Delta) u = f & \text{in } (0, \infty)_t \times U, \\ u = g & \text{on } \{t = 0\} \times U, \\ u = h & \text{on } (0, \infty)_t \times \partial U. \end{cases} \tag{5.3}$$

To deal with the evolutionary aspect of the heat equation, it is natural to think about what is called a *forward fundamental solution*, which may be thought of as Green's function for the initial value problem. To introduce this concept, we will first consider the simplest class of evolutionary differential equations, namely, ordinary differential equations (ODEs).

5.1. The idea of forward fundamental solution: a case study for ODEs.

For $A \in C((-\infty, \infty)_t; \mathbb{R}^{N \times N})$, consider the first-order linear ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{x}(t) - A(t) \mathbf{x}(t) = \mathbf{f}(t).$$

We say that $\Pi_+(t, s)$ is a *forward fundamental solution* (or *fundamental matrix*) if

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} \Pi_+(t, s) - A(t) \Pi_+(t, s) = \delta_0(t - s) I & \text{for } t \in \mathbb{R}, \\ \Pi_+(t, s) = 0 & \text{for } t \in (-\infty, s). \end{cases} \tag{5.4}$$

The condition $\Pi_+(t, s) = 0$ for $t \in (-\infty, s)$ is called the *forward (support) condition.* Intuitively, we are looking for a fundamental solution that stays 0 until we reach the time s, which is where the forcing term is supported.

To make sense of (5.4), we need to extend the idea of distributions to the case of N-vector (resp. $N \times N$-matrix-valued) functions. At the rudimentary level, these may be simply thought of as collections of N (resp. $N \times N$) many real-valued distributions. We may also view them as continuous linear functionals on the space of N-vector-valued test functions $C^\infty_c(I; \mathbb{R}^N)$ (resp. $N \times N$-matrix-valued test functions $C^\infty_c(I; \mathbb{R}^{N \times N})$), with the duality pairing (in the case of locally integrable functions) given by

$$\langle v, \phi \rangle = \sum_{j=1}^N \int_I v^j \phi^j \, \mathrm{d}t, \quad \left(\text{resp. } \langle A, \phi \rangle = \sum_{j,k=1}^N \int_I A^{jk} \phi^{jk} \, \mathrm{d}t \right).$$

Of course, vector- or matrix-valued distributions on $U \subseteq \mathbb{R}^d$ are defined in the same way.

Now that we have precisely formulated the meaning of (5.4), we turn to the natural question: how do we construct a solution to (5.4)? The idea is to *introduce a unit jump at $t = s$, so that we would see $\delta_0(t-s)$ after differentiation.* More precisely, we define $\Pi_+(t,s)$ in the following way:

(1) for $t < s$, we define $\Pi_+(t,s) = 0$ (forward property);
(2) for $t > s$, we define $\Pi_+(t,s) = \Pi(t,s)$, where $\Pi(t,s)$ is the solution to the homogeneous ODE

$$\begin{cases} \partial_t \Pi(t,s) + A(t)\Pi(t,s) = 0 & \text{for } t \in \mathbb{R}, \\ \Pi(s,s) = I & \text{at } t = s. \end{cases} \tag{5.5}$$

By the standard ODE theory, $\Pi(t,s) \in C^1((-\infty,\infty)_t)$. Therefore, for every fixed s, $\Pi_+(t,s)$ is a locally integrable function on $(-\infty,\infty)_t$ and hence a distribution. Moreover, we have

$$\begin{aligned} \langle (\partial_t - A(t))\Pi_+, \varphi \rangle &= \langle \Pi_+, (-\partial_t + A^*(t))\varphi(t) \rangle \\ &= \int_s^\infty \Pi_+(t,s)(-\partial_t + A^*(t))\varphi(t)\,\mathrm{d}t \\ &= \int_s^\infty (\partial_t - A(t))\Pi_+(t,s)\varphi(t)\,\mathrm{d}t + \varphi(0) \\ &= \varphi(0), \end{aligned}$$

which is exactly the meaning of the expression $(\partial_t - A(t))\Pi_+ = \delta_0 I$.

As we have seen in Section 4 (see also Section 7.3), the following is an immediate consequence of the existence of Π_+:

(1) **Solution formula, forward solution.** For $\mathbf{f} \in L^1(\mathbb{R};\mathbb{R}^N)$ with $\operatorname{supp}\mathbf{f} \subseteq \{t > a\}$ for some $a \in \mathbb{R}$,

$$\mathbf{x}(t) = \int_{-\infty}^\infty \Pi_+(t,s)\mathbf{f}(s)\,\mathrm{d}s = \int_a^t \Pi(t,s)\mathbf{f}(s)\,\mathrm{d}s.$$

solves $(\partial_t + A)\mathbf{x} = \mathbf{f}$ with $\mathbf{x}(t) = 0$ for $t \in (-\infty, a]$.

By the forward support property of $\Pi_+(t,s)$, note that the above formula defines $\mathbf{x}(t)$ for any $t > a$ as long as $\mathbf{f} \in L^1_{loc}((-\infty,\infty);\mathbb{R}^N)$ with $\operatorname{supp}\mathbf{f} \subseteq \{t > a\}$. Moreover, also observe that

$$\mathbf{x}(t) = \Pi(t,a)\mathbf{g}$$

solves the homogeneous problem $(\partial_t + A)\mathbf{x} = 0$ for $t \in (a,\infty)$ with $\mathbf{x}(a) = \mathbf{g}$. By linearity, we obtain the following:

(2) **Solution formula, for inhomogeneous initial value problem.** For $\mathbf{f} \in L^1_{loc}(\mathbb{R};\mathbb{R}^N)$ and $\mathbf{g} \in \mathbb{R}^N$,

$$\mathbf{x}(t) = \int_a^t \Pi(t,s)\mathbf{f}(s)\ \mathrm{d}s + \Pi(t,a)\mathbf{g} \tag{5.6}$$

is a solution to $(\partial_t + A)\mathbf{x} = \mathbf{f}$ on (a,∞) with $\mathbf{x}(a) = \mathbf{g}$. We know that such a solution is unique from the standard theory of ODEs.

We may summarize the above discussion in two ways:

- The forward fundamental solution $\Pi_+(t,s)$ – which is used find a forward solution to the inhomogeneous problem – is constructed from the solution $\Pi(t,s)$ to the initial value problem (5.5) for the homogeneous problem.

- Conversely, finding the solution $\Pi(t,s)$ to the initial value problem (5.5) for the homogeneous problem for $t > s$ amounts to finding the forward solution $\Pi_+(t,s)$ to the inhomogeneous problem with $\mathbf{f} = \delta_0(t-a)$.

Informally speaking, by thinking about the forward fundamental solution, we have discovered that the homogeneous problem can be reduced to the (forward) inhomogeneous problem, and vice versa. This is usually referred to as *Duhamel's principle* or *variation of constants.*

Next, we derive representation formulas.

(3) **Representation formula, forward solution.** For $\mathbf{x} \in C^1(\mathbb{R})$ with $\operatorname{supp} \mathbf{x} \subseteq \{t > a\}$ for some $a \in \mathbb{R}$,

$$\mathbf{x}(t) = \int_{-\infty}^{\infty} \Pi_+(t,s)(\partial_s + A(s))\mathbf{x}(s) \, \mathrm{d}s = \int_{-\infty}^{t} \Pi_+(t,s)(\partial_s + A(s))\mathbf{x}(s) \, \mathrm{d}s$$

To prove this, we observe that

$$\Pi(s,t)\Pi(t,s) = I,$$

hence $\Pi(s,t) = \Pi(t,s)^{-1}$. Using σ_1 and σ_2 (in this order) to denote the two input variables of $\Pi = \Pi(\sigma_1, \sigma_2)$, it follows that

$$\partial_{\sigma_2}\Pi(s,t) = -\Pi(s,t)\partial_{\sigma_1}\Pi(t,s)\Pi(s,t) = \Pi(s,t)A(t)\Pi(t,s)\Pi(s,t) = \Pi(s,t)A(t),$$

or more succinctly,

$$\partial_s \Pi(t,s) = \Pi(t,s)A(s).$$

Writing $\Pi_+(t,s) = \mathbf{1}_{(s,\infty)}(t)\Pi(t,s) = \mathbf{1}_{(-\infty,t)}(s)\Pi(t,s)$ and using the fact that $\Pi(s,s) = I$, we obtain

$$\begin{cases} -\partial_s \Pi_+(t,s) + \Pi_+(t,s)A(s) = \delta_0(s-t)I & \text{for } t \in \mathbb{R}_s, \\ \Pi_+(t,s) = 0 & \text{for } s \in (t,\infty). \end{cases} \tag{5.7}$$

For $\mathbf{x} \in C^1(\mathbb{R})$ with $\operatorname{supp} \mathbf{x} \subseteq \{t > a\}$, the following formal computation may be justified:

$$\begin{aligned} \mathbf{x}^j(t) &= \sum_k \langle \delta_0(s-t)\delta^j_k, \mathbf{x}^k(s) \rangle \\ &= \sum_k \int_{-\infty}^{\infty} (-\partial_s \Pi_+(t,s) + \Pi_+(t,s)A(s))^j_k \mathbf{x}^k(s) \, \mathrm{d}s \\ &= \int_{-\infty}^{\infty} (\Pi_+(t,s)(\partial_s + A(s))\mathbf{x}(s))^j \, \mathrm{d}s. \end{aligned}$$

Remark 5.1. Another way to understand (5.7) is observe that, after taking the adjoint (as matrices) and exchanging s and t, it is equivalent to the following:

$$\begin{cases} (-\partial_t + A^*(t))\Pi^*_+(s,t) = \delta_0(t-s)I & \text{for } s \in \mathbb{R}_t, \\ \Pi^*_+(s,t) = 0 & \text{for } t \in (s,\infty). \end{cases}$$

In other words, for each fixed s, $\Pi^*_+(s,t)$ is a backward fundamental solution for $-\partial_t + A^*(t)$, which is the adjoint of the original operator $\partial_t + A(t)$ (as an operator on the space of $\mathbb{R}^N$-valued functions, with respect to the pairing $\langle u, v \rangle = \sum_{j=1}^N \int_I u^j v^j \, \mathrm{d}t$).

Finally, we also obtain the following statement.

(4) **Representation formula, for inhomogeneous initial value problem.** For $\mathbf{x} \in C^1(I) \cap C(\bar{I})$,

$$\begin{aligned}\mathbf{x}(t) &= \int_{-\infty}^{\infty} \Pi_+(t,s)(\partial_s + A(s))(\mathbf{x}\mathbf{1}_{(a,\infty)})(s)\,\mathrm{d}s \\ &= \int_a^t \Pi_+(t,s)(\partial_s + A(s))\mathbf{x}(s)\,\mathrm{d}s + \Pi_+(t,a)\mathbf{x}(a).\end{aligned}$$

Remark 5.2 (Uniqueness of Π_+). It turns out that a forward fundamental solution is *unique* in a suitable class of objects, or more precisely, among order zero distributions. To see why this class is natural, observe that when $A \in C(\mathbb{R}; \mathbb{R}^{N\times N})$, this is the natural class for which the product $A\Pi_+$ makes sense.

The basic observation is that given any two solutions Π_+ and Π_+' to (5.4), $M := \Pi_+ - \Pi_+'$ solves the homogeneous equation $(\partial_t + A(t))M(t,s) = 0$ with $M = 0$ on $(-\infty, s)$. Then appealing to the uniqueness of the solution $\mathbf{x}$ to the ODE $(\partial_t + A(t))\mathbf{x}(t) = 0$ with $\mathbf{x} = 0$ on $(-\infty, a)$ in the class of order zero distributions (**Exercise:** prove this!), It follows that $M = 0$, i.e., $\Pi_+ = \Pi_+'$.

5.2. **The forward fundamental solution for $\partial_t - \Delta$: the Gaussian.** We now look for the forward fundamental solution for $\partial_t - \Delta$, i.e., a solution $E_+ \in \mathcal{D}'(\mathbb{R}^{1+d})$ to

$$(\partial_t - \Delta)E_+(t,x) = \delta_0(t,x)$$

with the forward property, i.e.,

$$E_+ = 0 \quad \text{in } \{t < 0\}.$$

As in Section 5.1, it will be sufficient look for the solution to the following homogeneous initial value problem:

$$\begin{cases} (\partial_t - \Delta)E_+ = 0 & \text{in } (0,\infty)_t \times \mathbb{R}^d, \\ \quad E_+(0,x) = \delta_0(x) & \text{on } \{t=0\} \times \mathbb{R}^d. \end{cases}$$

As in the case of the Laplace operator, we shall make an educated guess based on the properties of $(\partial_t - \Delta)$. We begin with its symmetries:

- Rotational symmetry: For every orthogonal matrix O, $(\partial_t - \Delta)(u(t,Ox)) = ((\partial_t - \Delta)u)(t,Ox)$.
- Scaling (self-similarity): For every $\mu \in \mathbb{R}$ and $\lambda > 0$, $(\partial_t - \Delta)(\mu u(\lambda^{-2}t, \lambda^{-1}x)) = \mu\lambda^{-2}((\partial_t - \Delta)u)(\lambda^{-2}t, \lambda^{-1}x)$.

In view of these symmetries, we look for $E_+(t,x)$ such that

$$E_+(t,x) = E_+(t,|x|) = \mu(t)E_+(1, \tfrac{|x|}{t^{\frac{1}{2}}}).$$

How should we choose $\mu(t)$? Another property we observe is that

$$(\partial_t - \Delta)u = 0 \Rightarrow \partial_t \int u\,\mathrm{d}x = \int \partial_t u\,\mathrm{d}x = \int \Delta u\,\mathrm{d}x,$$

which would vanish provided that u decays sufficiently fast at infinity. Since $\int E_+(0,x) = \int \delta_0(x) = 1$, it is natural to set $\mu(t) = t^{-\frac{d}{2}}$ and ask for

$$E_+(t,x) = t^{-\frac{d}{2}} E_+(1, \tfrac{|x|}{t^{\frac{1}{2}}}), \qquad \int E_+(1,x)\,\mathrm{d}x = 1.$$

In order to proceed, let us write $w(x) := E_+(1, |x|)$. The heat equation on $E_+(t,x)$ becomes the following equation for w:

$$
\begin{aligned}
&(\partial_t - \Delta)\left(\frac{1}{t^{\frac{d}{2}}} w\left(\frac{|x|^2}{t}\right)\right) \\
&= \left[-\frac{d}{2}\frac{1}{t}\frac{1}{t^{\frac{d}{2}}} w - \frac{|x|^2}{t^2}\frac{1}{t^{\frac{d}{2}}} w' - \sum_j \partial_j \left(\frac{2x^j}{t}\frac{1}{t^{\frac{d}{2}}} w'\right)\right]\left(\frac{|x|^2}{t}\right) \\
&= \left[-\frac{d}{2}\frac{1}{t}\frac{1}{t^{\frac{d}{2}}} w - \frac{|x|^2}{t^2}\frac{1}{t^{\frac{d}{2}}} w' - \frac{2d}{t}\frac{1}{t^{\frac{d}{2}}} w' - \frac{4|x|^2}{t^2}\frac{1}{t^{\frac{d}{2}}} w''\right]\left(\frac{|x|^2}{t}\right) \\
&= -\frac{1}{t^{1+\frac{d}{2}}}\left[\frac{4|x|^2}{t} w''\left(\frac{|x|^2}{t}\right) + (2d + \frac{|x|^2}{t}) w' + \frac{d}{2} w\right]\left(\frac{|x|^2}{t}\right).
\end{aligned}
$$

Introducing $\rho = \frac{|x|^2}{t}$, we look for a solution to

$$4\rho w''(\rho) + (2d + \rho) w'(\rho) + \frac{d}{2} w(\rho) = 0.$$

We may rewrite the above as

$$4\rho w''(\rho) + 2d w'(\rho) + \rho w'(\rho) + \frac{d}{2} w(\rho) = 0,$$

or equivalently,

$$4(\rho^d w')' + (\rho^d w)' = 0.$$

It follows that, for some constant a,

$$4\rho^d w' + \rho^d w = a.$$

For simplicity, we set $a = 0$ (otherwise, w needs to grow very fast as $\rho \to \infty$). Then we arrive at the ODE

$$w' = -\frac{1}{4} w,$$

so

$$w(\rho) = b e^{-\frac{1}{4}\rho}.$$

Returning to $E_+(t,x)$, we arrive at

$$E_+(t,x) = \frac{b}{t^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} \quad \text{for } t > 0.$$

To keep $\int E_+(t,x)\,\mathrm{d}x = 1$, we select $b = \frac{1}{(4\pi)^{\frac{d}{2}}}$.

Let us now verify that $E_+(t,x)$ is indeed a forward fundamental solution for $(\partial_t - \Delta)$.

Proposition 5.3. *The function*

$$E_+(t,x) = \mathbf{1}_{(0,\infty)}(t) \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}}$$

is a forward fundamental solution for $\partial_t - \Delta$ in $\mathbb{R}^{1+d}$.

Proof. The forward support property is clear, so we simply need to check that $(\partial_t - \Delta) E_+ = \delta_0$. Given a test function $\varphi \in C^\infty_c(\mathbb{R}^{1+d})$, we compute

$$\iint E_+(t,x)(-\partial_t - \Delta)\varphi(t,x)\,\mathrm{d}t\mathrm{d}x$$

$$
\begin{aligned}
&= \int \int_0^\infty \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}}(t,x)(-\partial_t - \Delta)\varphi(t,x)\,\mathrm{d}t\mathrm{d}x \\
&= \lim_{\epsilon\to 0+} \int \int_\epsilon^\infty \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}}(-\partial_t - \Delta)\varphi(t,x)\,\mathrm{d}t\mathrm{d}x \\
&= \lim_{\epsilon\to 0+} \int \int_\epsilon^\infty (\partial_t - \Delta)\frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}}\varphi(t,x)\,\mathrm{d}t\mathrm{d}x \\
&\quad + \lim_{\epsilon\to 0+} \int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}}\varphi(\epsilon,x)\,\mathrm{d}x.
\end{aligned}
$$

In the last expression, note that the spacetime integral is zero since $\frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}}$ solves the homogeneous heat equation in $\{t > 0\}$. We rewrite the last expression as

$$
\begin{aligned}
\int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}}\varphi(\epsilon,x)\,\mathrm{d}x = &\int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}}\varphi(0,0)\,\mathrm{d}x \\
&+ \int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}}(\varphi(\epsilon,x) - \varphi(0,0))\,\mathrm{d}x.
\end{aligned}
$$

The first term is equal to $\varphi(0,0)$ by our normalization. For the second term, we may write $\varphi(\epsilon,x) - \varphi(0,0) = O(\epsilon + |x|)$ by the fundamental theorem of calculus (where the implicit constant depends on $\|\partial\varphi\|_{L^\infty}$). We wish to argue that $e^{-\frac{|x|^2}{4\epsilon}}$ essentially localizes the integral to $\{|x| \lesssim \epsilon^{\frac{1}{2}}\}$, so this contribution should vanish as $\epsilon \to 0$. Indeed,

$$
\begin{aligned}
\int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}}(\epsilon + |x|)\,\mathrm{d}x &\leq \epsilon \int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}}\,\mathrm{d}x + \epsilon^{\frac{1}{2}} \int \frac{1}{(4\pi \epsilon)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4\epsilon}} \frac{|x|}{\epsilon^{\frac{1}{2}}}\,\mathrm{d}x \\
&= \epsilon \int \frac{1}{(4\pi)^{\frac{d}{2}}} e^{-\frac{|y|^2}{4}}\,\mathrm{d}y + \epsilon^{\frac{1}{2}} \int \frac{1}{(4\pi)^{\frac{d}{2}}} e^{-\frac{|y|^2}{4}}|y|\,\mathrm{d}y,
\end{aligned}
$$

where we used a change of variables $y = \epsilon^{-\frac{1}{2}}x$ on the last line. The last line clearly vanishes as $\epsilon \to 0$, which finishes the proof.

□

5.3. **Uses of the forward fundamental solution.** We now derive some consequences of our discovery of E_+.

Solution formula. For $f \in \mathcal{D}'_c(\mathbb{R}^{1+d})$,

$$
u(t,x) = \int f(s,y)E_+(t-s, x-y)\,\mathrm{d}s\mathrm{d}y
$$

is a forward solution to $(\partial_t - \Delta)u = f$. Moreover, for $g \in \mathcal{D}'_c(\mathbb{R}^d)$,

$$
u(t,x) = \int f(s,y)E_+(t-s,x-y)\,\mathrm{d}s\mathrm{d}y + \int g(y)E_+(t,x-y)\,\mathrm{d}y
$$

solves the initial value problem (9.1).

Representation formula. For $u \in \mathcal{D}'_c(\mathbb{R}^{1+d})$, we have

$$
u(t,x) = \int (\partial_t - \Delta)u(s,y)E_+(t-s,x-y)\,\mathrm{d}y.
$$

Remark 5.4 (**Optional:** Extension to f, g without compact support). These solution (resp. representation) formulae can be extended to a more general class of distributions f, g (resp. u). We will say that $g \in \mathcal{D}'(\mathbb{R}^d)$ satisfies the *Gaussian growth condition* if, for every $A > 0$, we have

$$\langle g, (\chi_{>1}(\tfrac{x}{R'}) - \chi_{>1}(\tfrac{x}{R}))e^{-A|x|^2}\rangle \to 0 \text{ as } R, R' \to \infty.$$

For instance, of $g \in L^1_{loc}(\mathbb{R}^d)$ and $|g| \leq Ce^{|x|^{2-\epsilon}}$ for some $C > 0$ and $\epsilon > 0$, then g satisfies the Gaussian growth condition.

Moreover, for $f \in \mathcal{D}'(\mathbb{R} \times \mathbb{R}^d)$, we say that f satisfies the Gaussian growth condition in space if for every $\eta \in C^\infty_c(\mathbb{R}_t)$ and $A > 0$, we have

$$\langle f, \eta(t)(\chi_{>1}(\tfrac{x}{R'}) - \chi_{>1}(\tfrac{x}{R}))e^{-A|x|^2}\rangle \to 0 \text{ as } R, R' \to \infty.$$

The solution formulae can be extended to $f \in \mathcal{D}'(\mathbb{R}^{1+d})$ with $\operatorname{supp} f \subseteq \{t > a\}$ for some $a \in \mathbb{R}$ and satisfying the Gaussian growth condition in space, and $g \in \mathcal{D}'(\mathbb{R}^d)$ with the Gaussian growth condition. The representation formula can be extended to $u \in \mathcal{D}'(\mathbb{R}^{1+d})$ with $\operatorname{supp} f \subseteq \{t > a\}$ for some $a \in \mathbb{R}$ and satisfying the Gaussian growth condition in space.

Regularity. From the fact that $E_+(t, x)$ is smooth outside of $(0, 0)$, we obtain the following regularity property of solutions to the heat equation:

Proposition 5.5. *If u solves $(\partial_t - \Delta)u = 0$ in $\mathcal{U} \subseteq \mathbb{R}^{1+d}$, then u is smooth.*

We will skip its proof, since we will discuss the more quantitative version of this result below.

Derivative bounds and Liouville theorem for ancient solutions. From the same proof as before, we can derive the following derivative bounds:

Proposition 5.6. *Let u be a solution to $(\partial_t - \Delta)u = 0$ in $\mathcal{U} \subseteq \mathbb{R}^{1+d}$. Then for any (t, x) and $R > 0$ such that $\overline{(t - R^2, t) \times B_R(x)} \subseteq \mathcal{U}$, we have*

$$|\partial_t^k \partial_x^\alpha u(t, x)| \leq C_\alpha R^{-2k-|\alpha|} \sup_{(s,y)\in(t-R^2,t)\times B_R(x)} |u(s, y)|.$$

Here, note that we only need the information on u to the past of (t, x). As we will see, this feature is due to the forward support property of E_+.

Proof. Without loss of generality, assume that $(t, x) = (0, 0)$. Note also that it suffices to consider the case $k = 0$, since we can always trade ∂_t by Δ using the equation $(\partial_t - \Delta)u = 0$.

We introduce a spacetime cutoff $\chi(s, y) \in C^\infty(\mathbb{R}^{1+d})$ that equals 1 in $(-\frac{1}{2}, \frac{1}{2}) \times B_{\frac{1}{2}}(0)$ and $\operatorname{supp} \chi \subseteq (-1, 1) \times B_1(0)$, and a space cutoff $\underline{\chi}(y) \in C^\infty(\mathbb{R}^d)$ that equals 1 in $B_{\frac{1}{2}}(0)$ and $\operatorname{supp} \underline{\chi} \subseteq B_1(0)$. Define

$$\chi_R(t, x) = \chi(R^{-2}t, R^{-1}x), \quad \underline{\chi}_R(x) = \chi(R^{-1}x).$$

Consider

$$v(s, y) := u(s, y)\chi_R(s, y).$$

Note that v agrees with u on $(-\frac{R}{2}, \frac{R}{2}) \times B_{\frac{R}{2}}(0)$. By the representation formula (which applies since v is compactly supported),

$$\underline{\chi}_{\frac{R}{4}}(x)\partial_x^\alpha u(0, x) = \underline{\chi}_{\frac{R}{4}}(x)\partial_x^\alpha v(0, x)$$

$$= \iint \underline{\chi}_{\frac{R}{4}}(x)(\partial_s - \Delta)v(s,y)\partial_x^\alpha E_+(-s,x-y)\,\mathrm{d}s\mathrm{d}y$$
$$= \int_{-R^2}^0 \int_{B_R(0)} \underline{\chi}_{\frac{R}{4}}(x)(\partial_s - \Delta)v(s,y)\partial_x^\alpha E_+(-s,x-y)\,\mathrm{d}s\mathrm{d}y,$$

where we used the support properties of E_+ and v. Observe furthermore that

$$(\partial_s - \Delta)v(s,y) = R^{-2}u(s,y)((\partial_s - \Delta)\chi)(R^{-2}s, R^{-1}y)$$
$$-2\sum_j R^{-1}\partial_j u(s,y)\cdot(\partial_j\chi)(R^{-2}s, R^{-1}y)$$

is, in fact, supported in $\operatorname{supp}\partial\chi_R \subseteq (-R,R)\times B_R(0)\setminus(-\frac{R}{2},\frac{R}{2})\times B_{\frac{R}{2}}(0)$. It follows that

$$|s| + |x-y|^2 \geq \frac{R^2}{8} \text{ for } x \in \operatorname{supp}\underline{\chi}_R,\ (s,y)\in\operatorname{supp}\partial\chi_R.$$

In particular,

$$\frac{1}{|s|} \leq \frac{8}{R^2}\left(1 + \frac{|x-y|^2}{|s|}\right).$$

In this region,

$$|\partial_x^\alpha E_+(-s,x-y)| = \frac{1}{(4\pi|s|)^{\frac{d}{2}}}\left|\partial_x^\alpha e^{-\frac{|x-y|^2}{4|s|}}\right|$$
$$\leq C_\alpha |s|^{-\frac{d}{2}-|\alpha|}|x-y|^{|\alpha|}e^{-\frac{|x-y|^2}{4|s|}}$$
$$\leq C_\alpha |s|^{-\frac{d+|\alpha|}{2}}\left(1+\frac{|x-y|^2}{|s|}\right)^{\frac{|\alpha|}{2}} e^{-\frac{1}{4}(1+\frac{|x-y|^2}{|s|})}$$
$$\leq C_\alpha R^{-d-|\alpha|}\left(1+\frac{|x-y|^2}{|s|}\right)^{d+|\alpha|} e^{-\frac{1}{4}(1+\frac{|x-y|^2}{|s|})}$$
$$\leq C_\alpha R^{-d-|\alpha|},$$

where, for the last inequality, we used that $s^{d+|\alpha|}e^{-\frac{1}{4}s}$ is uniformly bounded for all $s > 0$. Thus, for the contribution of $R^{-2}u(s,y)((\partial_s - \Delta)\chi)(R^{-2}s, R^{-1}y)$, we estimate

$$\left|\int_{-R^2}^0 \int_{B_R(0)} \underline{\chi}_{\frac{R}{4}}(x)\left(R^{-2}u(s,y)((\partial_s-\Delta)\chi)(R^{-2}s,R^{-1}y)\right)\partial_x^\alpha E_+(-s,x-y)\,\mathrm{d}s\mathrm{d}y\right|$$
$$\leq C_\alpha R^{-2-d-|\alpha|}\sup_{(-R^2,0)\times B_R(0)}|u|\int_{-R^2}^0\int_{B_R(0)}\mathrm{d}y\mathrm{d}s$$
$$\leq C_\alpha R^{-|\alpha|}\sup_{(-R^2,0)\times B_R(0)}|u|.$$

For the contribution of $-2\sum_j R^{-1}\partial_j u(s,y)\cdot(\partial_j\chi)(R^{-2}s,R^{-1}y)$, we first integrate by parts to write

$$\int_{-R^2}^0 \int_{B_R(0)} \underline{\chi}_{\frac{R}{4}}(x)\left(-2\sum_j R^{-1}\partial_j u(s,y)\cdot(\partial_j\chi)(R^{-2}s,R^{-1}y)\right)\partial_x^\alpha E_+(-s,x-y)\,\mathrm{d}s\mathrm{d}y$$
$$= \sum_j 2\int_{-R^2}^0\int_{B_R(0)}\underline{\chi}_{\frac{R}{4}}(x)R^{-1}u(s,y)\partial_j\left((\partial_j\chi)(R^{-2}s,R^{-1}y)\partial_x^\alpha E_+(-s,x-y)\right)\mathrm{d}s\mathrm{d}y,$$

which can then be handled as before. □

From the derivative bounds, we obtain Liouville's theorem for solutions to the heat equation that exists on an interval that is unbounded to the past (i.e., ancient solutions).

Theorem 5.7 (Liouville theorem for ancient solutions)**.** *If $u(t,x) \in C(\mathbb{R}^{1+d})$ solves $(\partial_t - \Delta)u = 0$ in $(-\infty, a) \times \mathbb{R}^d$ and is uniformly bounded, then $u = const$.*

We skip the simple proof.

Mean value property. Amusingly, we can also derive a *mean value property* of solutions to the heat equation, which is not at all obvious.

Theorem 5.8 (Mean value property)**.** *For $(t,x) \in \mathbb{R}^{1+d}$, define*

$$\mathcal{E}_r(t,x) = \{(s,y) \in \mathbb{R}^{1+d} : s \leq t,\ E_+(t-s, x-y) \geq r^{-d}\}.$$

Let $u \in C^2(\mathcal{U})$ satisfy $(\partial_t - \Delta)u = 0$ in $\mathcal{U}$. For any (t,x) such that $\overline{\mathcal{E}_r(t,x)} \subseteq \mathcal{U}$, we have

$$u(t,x) = \frac{1}{4r^d} \iint_{\mathcal{E}_r(t,x)} u(s,y) \frac{|x-y|^2}{(t-s)^2} \mathrm{d}s\mathrm{d}y.$$

We will first discuss the key idea behind the proof of this theorem, in particular how to derive the particular formula. As in the case of the Laplace equation, our starting point is the following representation formula on a bounded spacetime domain:

Lemma 5.9. *Let $u \in C^2(\mathcal{U})$ and consider a connected C^1 spacetime domain $\mathcal{V} \subseteq \overline{\mathcal{V}} \subseteq \mathcal{U}$. Let $\widetilde{E} = E_+ + h$, where h solves $(\partial_t - \Delta)h = 0$ in $\mathbb{R}^{1+d}$. Then for any $(t,x) \in \mathcal{V}$,*

$$\begin{aligned} u(t,x) = & \iint_{\mathcal{V}} \widetilde{E}(t-s, x-y)(\partial_s - \Delta_y)u(s,y)\,\mathrm{d}s\mathrm{d}y \\ & - \iint_{\partial\mathcal{V}} \nu_0(s,y)\widetilde{E}(t-s,x-y)u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}}(s,y) \\ & - \sum_j \iint_{\partial\mathcal{V}} \nu_j(s,y)\partial_{y^j}(\widetilde{E}(t-s,x-y))u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}}(s,y) \\ & + \sum_j \iint_{\partial\mathcal{V}} \nu_j(s,y)\widetilde{E}(t-s,x-y)\partial_j u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}}(s,y). \end{aligned} \tag{5.8}$$

Proof. In the following computation, we shall assume that $u \in C^\infty(\mathcal{U})$; the case $u \in C^2(\mathcal{U})$ follows by approximation. Note that $\delta_0(t-s, x-y) = (\partial_t - \Delta_x)\widetilde{E}(t-s, x-y) = (-\partial_s - \Delta_y)\widetilde{E}(t-s,x-y)$, while $\delta_0(t-s,x-y) = \delta_0(s-t,y-x)$. We have

$$\begin{aligned} u(t,x) &= \langle \delta_0(s-t, y-x), u(s,y)\rangle \\ &= \langle (-\partial_s - \Delta_y)(\widetilde{E}(t-s,x-y)), u(s,y)\rangle \\ &= \langle \mathbf{1}_{\mathcal{V}}(s,y)(-\partial_s - \Delta_y)(\widetilde{E}(t-s,x-y)), u(s,y)\rangle \\ &= \langle \partial_s \mathbf{1}_{\mathcal{V}}(s,y)(\widetilde{E}(t-s,x-y)), u(s,y)\rangle \\ &\quad + \langle \mathbf{1}_{\mathcal{V}}(s,y)(\widetilde{E}(t-s,x-y)), \partial_s u(s,y)\rangle \end{aligned}$$

$$
\begin{aligned}
&+\sum_j \langle \partial_{y^j} \mathbf{1}_{\mathcal{V}}(s,y)\partial_{y^j}(\widetilde{E}(t-s,x-y)), u(s,y)\rangle \\
&-\sum_j \langle \partial_{y^j} \mathbf{1}_{\mathcal{V}}(s,y)(\widetilde{E}(t-s,x-y)), \partial_{y^j} u(s,y)\rangle \\
&+\langle \mathbf{1}_{\mathcal{V}}(s,y)(\widetilde{E}(t-s,x-y)), -\Delta_y u(s,y)\rangle \\
=&\iint_{\mathcal{V}} \widetilde{E}(t-s,x-y)(\partial_s-\Delta_y)u(s,y)\,\mathrm{d}s\mathrm{d}y \\
&-\int \nu_0(s,y)\widetilde{E}(t-s,x-y)u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}}(s,y) \\
&-\sum_j \int \nu_j(s,y)\partial_{y^j}(\widetilde{E}(t-s,x-y))u(s,y)\mathrm{d}\sigma_{\partial\mathcal{V}}(s,y) \\
&+\sum_j \int \nu_j(s,y)\widetilde{E}(t-s,x-y)\partial_{y^j}u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}}(s,y),
\end{aligned}
$$

which is the desired conclusion. □

We will also need the following simple version of the *coarea formula.*

Theorem 5.10 (Corea formula [EG15, Ch. 3])**.** *Let $f:\mathbb{R}^d\to\mathbb{R}$ be Lipschitz continuous and assume that for a.e. $\rho\in\mathbb{R}$ the level set*

$$\{x\in\mathbb{R}^d : f(x)=\rho\}$$

is a smooth, $(n-1)$-dimensional hypersurface in $\mathbb{R}^d$. Assume also $u:\mathbb{R}^d\to\mathbb{R}$ is continuous and integrable. Then

$$\int_{\mathbb{R}^d} u|Df|\,\mathrm{d}x = \int_{-\infty}^{\infty}\left(\int_{\{f=\rho\}} f\,\mathrm{d}\sigma\right)\mathrm{d}\rho.$$

We are now ready to prove Theorem 5.8.

Proof of Theorem 5.8. Let $u\in C^2(\mathcal{U})$ solve the heat equation, $(\partial_t-\Delta)u=0$. Without loss of generality, take $(t,x)=(0,0)$.
Step 1: Derivation of the formula. To derive the mean value formula for $u(0,0)$, we apply Lemma 5.9 with $\mathcal{V}_\rho := \{E_+(-s,-y)\geq\rho\}$ and $\widetilde{E}_\rho := E_+-\rho$, so that $\widetilde{E}_\rho=0$ on $\partial\mathcal{V}_\rho$. Note that the LHS of (5.8) is $u(0,0)$, while

$$
\begin{aligned}
&(\text{RHS of (5.8)}) \\
&=-\sum_j\int_{\partial\mathcal{V}_\rho}\nu_j(s,y)\partial_{y^j}(\widetilde{E}_\rho(-s,-y))u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}_\rho}(s,y) \\
&=-\sum_j\int_{\partial\mathcal{V}_\rho}\nu_j(s,y)\partial_{y^j}(E_+)(-s,-y))u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}_\rho}(s,y) \\
&=\sum_j\int_{\partial\mathcal{V}_\rho}\frac{(\partial_{y^j}(E_+)(-s,-y)))^2}{|\nabla_{s,y}E_+(-s,-y)|}u(s,y)\,\mathrm{d}\sigma_{\partial\mathcal{V}_\rho}(s,y)
\end{aligned}
$$

where

$$|\nabla_{s,y}E_+(-s,-y)|$$

$$= \left[(\partial_s E_+(-s,-y))^2 + (\partial_{y^1} E_+(-s,-y))^2 + \cdots (\partial_{y^d} E_+(-s,-y))^2 \right]^{\frac{1}{2}}.$$

On the last line, we used the fact that $\partial \mathcal{V}_\rho$ is a level hypersurface of $E_+(-s,-y)$ so

$$\nu_j(s,y) = \frac{-\partial_{y^j} E_+(-s,-y)}{|\nabla_{s,y} E_+(-s,-y)|}.$$

Observe also that on $\partial \mathcal{V}_\rho$, we have

$$\partial_{y^j}(E_+)(-s,-y)) = \frac{2y^j}{-4s} \left[\frac{1}{(4\pi(-s))^{\frac{d}{2}}} e^{-\frac{|y|^2}{4(-s)}} \right] = -\frac{y^j}{2s}\rho.$$

Thus,

$$\sum_j (\partial_{y^j}(E_+)(-s,-y)))^2 = \frac{|y|^2}{4s^2}\rho^2,$$

and

$$\begin{aligned} &(\text{RHS of } (5.8)) \\ &= \int_{\partial \mathcal{V}_\rho} u(s,y) \frac{|y|^2}{4s^2}\rho^2 \, \frac{\mathrm{d}\sigma_{\partial \mathcal{V}_\rho}(s,y)}{|\nabla_{s,y} E_+(-s,-y)|}. \end{aligned}$$

To obtain a spacetime integral, we average this expression in ρ^{-1} for $0 < \rho^{-1} \leq R^{-1}$. Since $\rho^2 \mathrm{d}\rho^{-1} = -\mathrm{d}\rho$, we obtain

$$\begin{aligned} &R \int_0^{R^{-1}} (\text{RHS of } (5.8)) \, \mathrm{d}(\rho^{-1}) \\ &= R \int_R^\infty \int_{\partial \mathcal{V}_\rho} u(s,y) \frac{|y|^2}{4s^2} \, \frac{\mathrm{d}\sigma_{\partial \mathcal{V}_\rho}(s,y) \mathrm{d}\rho}{|\nabla_{s,y} E_+(-s,-y)|} \\ &= R \iint_{\{E_+(-s,-y) \geq R\}} u(s,y) \frac{|y|^2}{4s^2} \, \mathrm{d}s \mathrm{d}y. \end{aligned}$$

Choosing $R = r^{-d}$, this coincides with the RHS of the mean value formula in Theorem 5.8!

Step 2. Justification of (5.8). However, we are not done yet since Lemma 5.9 does *not* directly apply. Indeed, note that $(0,0)$ lies only on the boundary of $\mathcal{V}$, not in its interior as is required by Lemma 5.9. To finish the proof, we need to justify (5.8) holds for each $\mathcal{V}_\rho$ and $\widetilde{E}_\rho$ as above.

Fix $\rho > 0$. Given $\epsilon > 0$, we consider the following deformation of $\mathcal{V}_\rho$:

$$\mathcal{V}_{\rho,\epsilon} = \{s < -\epsilon^2\} \cap \mathcal{V}_\rho \cup [-\epsilon^2, \epsilon^2)_s \times B_{r(\epsilon)},$$

where $r(\epsilon)$ is defined so that $E_+(-\epsilon^2, r(\epsilon)) = \frac{1}{(4\pi\epsilon^2)^{\frac{d}{2}}} e^{-\frac{r(\epsilon)^2}{\epsilon^2}} = \rho$. Then $\mathcal{V}_{\rho,\epsilon}$ contains $(0,0)$, has a piecewise C^1 boundary, and

$$\partial V_{\rho,\epsilon} = \left(\partial V_\rho \cap \{s < -\epsilon^2\}\right) \cup \left([-\epsilon^2, \epsilon^2) \times \partial B_{r(\epsilon)}\right) \cup \left(\{\epsilon^2\} \times \overline{B_{r(\epsilon)}}\ \right).$$

Let us introduce the abbreviation $\partial_{top} V_{\rho,\epsilon} := \left([-\epsilon^2, \epsilon^2) \times \partial B_{r(\epsilon)}\right) \cup \left(\{\epsilon^2\} \times \overline{B_{r(\epsilon)}}\ \right)$. Applying Lemma 5.9 (which is now possible since $(0,0) \in \mathcal{V}_{\rho,\epsilon}$), we have

$$u(0,0) = -\sum_j \int_{\partial \mathcal{V}_\rho \cap \{s < -\epsilon^2\}} \nu_j(s,y) \partial_{y^j}(\widetilde{E}_\rho(-s,-y)) u(s,y) \, \mathrm{d}\sigma_{\partial \mathcal{V}}(s,y)$$

$$
\begin{aligned}
&- \int_{\partial_{top} V_{\rho,\epsilon}} \nu_0(s,y) \widetilde{E}_\rho(-s,-y) u(s,y) \, \mathrm{d}\sigma_{\partial \mathcal{V}}(s,y) \\
&- \sum_j \int_{\partial_{top} \mathcal{V}_{\rho,\epsilon}} \nu_j(s,y) \partial_{y^j} (\widetilde{E}_\rho(-s,-y)) u(s,y) \, \mathrm{d}\sigma_{\partial \mathcal{V}}(s,y) \\
&+ \sum_j \int_{\partial_{top} \mathcal{V}_{\rho,\epsilon}} \nu_j(s,y) \widetilde{E}_\rho(-s,-y) \partial_j u(s,y) \, \mathrm{d}\sigma_{\partial \mathcal{V}}(s,y).
\end{aligned}
$$

By the dominated convergence theorem, we have

$$
\begin{aligned}
&- \sum_j \int_{\partial \mathcal{V}_\rho \cap \{s < -\epsilon^2\}} \nu_j(s,y) \partial_{y^j} (\widetilde{E}_\rho(-s,-y)) u(s,y) \, \mathrm{d}\sigma_{\partial \mathcal{V}}(s,y) \\
&\to - \sum_j \int_{\partial \mathcal{V}_\rho} \nu_j(s,y) \partial_{y^j} (\widetilde{E}_\rho(-s,-y)) u(s,y) \, \mathrm{d}\sigma_{\partial \mathcal{V}}(s,y),
\end{aligned}
$$

as $\epsilon \to 0$, which is already the desired expression. It remains to show that the integrals on $\partial_{top} \mathcal{V}_{\rho,\epsilon}$ go to zero. First observe that, from $\frac{1}{(4\pi\epsilon^2)^{\frac{d}{2}}} e^{-\frac{r(\epsilon)^2}{4\epsilon^2}} = \rho$,

$$
\frac{r(\epsilon)^2}{4\epsilon^2} = -\log \rho (4\pi)^{\frac{d}{2}} - d \log \epsilon
$$

so $\frac{r(\epsilon)^2}{\epsilon^2(-\log \epsilon)} \to d$ as $\epsilon \to 0$. Since

$$
\begin{aligned}
\partial_s E_+(-s,-y) &= \partial_s \left(\frac{1}{(4\pi(-s))^{\frac{d}{2}}} e^{-\frac{|y|^2}{4(-s)}} \right) \\
&= -\frac{1}{(-s)} \left(\frac{|y|^2}{4(-s)} - \frac{d}{2} \right) \left(\frac{1}{(4\pi(-s))^{\frac{d}{2}}} e^{-\frac{|y|^2}{4(-s)}} \right)
\end{aligned}
$$

we see that $E_+(-s, r(\epsilon))$ is decreasing on $(-\epsilon^2, 0)$ if ϵ is sufficiently small (namely, so that $\frac{r(\epsilon)^2}{4\epsilon^2} - \frac{d}{2} > 0$). Note also that $E_+(-s, r(\epsilon)) = 0$ for $s \geq 0$. It follows that $|\widetilde{E}_\rho(-s,-y)| = E_+(-s,-y) - \rho$ is uniformly bounded (independent of ϵ) on $\partial_{top} \mathcal{V}_{\rho,\epsilon}$. A similar consideration shows that $\partial_{y^j} E_+(-s,-y)$ is uniformly bounded by $|\partial_{y^j} E_+(-\epsilon^2, r(\epsilon))| \leq \frac{r(\epsilon)}{2\epsilon^2} \rho$. Both lead to the desired vanishing statement, in view of the fact that $\int_{\partial_{top} \mathcal{V}_\rho \cap \{s<0\}} \mathrm{d}\sigma_{\partial \mathcal{V}} = O(\epsilon^2 r(\epsilon)^{d-1}) = O(\epsilon^{d+1} |\log \epsilon|^{d-1})$, which always vanishes faster than $\frac{r(\epsilon)}{\epsilon^2} = O(\frac{|\log \epsilon|}{\epsilon})$. □

Remark 5.11. The reader is encouraged to also look at the proof of Theorem 5.8 in [Eva10, Section 2.3], which is elementary and does not require distribution theory. Meanwhile, the proof presented above, while technical and long when carried out in full detail, clarifies the *derivation* of the mean value formula from the fundamental solution.

Maximum principles in bounded domain. Given $\mathcal{U} = (a,b) \times U$, define the *heat boundary* of $\mathcal{U}$ to be $\partial_h \mathcal{U} = ((a,b) \times \partial U) \cup (\{a\} \times \overline{U})$. Also, given $(t,x) \in \mathbb{R}^{1+d}$ and $r > 0$, introduce the $\mathcal{B}_r(t,x) := \{(s,y) \in \mathbb{R}^{1+d} : t - r < s \leq t, \, |x-y| < r\}$.

Theorem 5.12 (Maximum principles)**.** *Let U be a bounded domain in $\mathbb{R}^d$. Suppose $u \in C^2(I \times U) \cap C(\overline{I \times U})$ solves $(\partial_t - \Delta) u = 0$.*

(1) **Weak maximum principle.** *We have*

$$\max_{\overline{I\times U}} u = \max_{\partial_h(I\times U)} u.$$

(2) **Strong maximum principle.** *Moreover, if U is connected and there exists $(t_0, x_0) \in \overline{I\times U} \setminus \partial_h(I\times U)$ such that*

$$u(t_0, x_0) = \max_{\overline{I\times U}} u,$$

then u is constant in $I \times U \cap \{t \leq t_0\}$ (i.e., to the past of $t = t_0$).

Proof. We use the mean value formula, Theorem 5.8, to argue as in the case of the Laplace equation. We focus on the strong maximum principle, since the weak maximum principle would be a consequence.

Let $(t_0, x_0) \in \overline{I\times U} \setminus \partial_h(I\times U)$ with $u(t_0, x_0) = M$, where $M = \max_{\overline{I\times U}} u$. Then for all sufficiently small $r > 0$ such that $\mathcal{E}_r(t_0, x_0) \in I \times U$, the mean value property implies that

$$M = u(t_0, x_0) = \frac{1}{4r^n} \int_{\mathcal{E}_r(t_0,x_0)} u \frac{|x_0 - y|^2}{(t_0 - s)^2} \,\mathrm{d}y\mathrm{d}s \leq M.$$

Since $\frac{1}{4r^n} \int_{\mathcal{E}_r(t_0,x_0)} \frac{|x_0-y|^2}{(t_0-s)^2} = 1$ (which can be easily checked by applying the mean value property to $u = 1$), it follows that equality holds if and only if u is identically equal to M in $\mathcal{E}_r(t_0, x_0)$.

In order to extend this property, we follow the argument in [Eva10, Section 2.3]. Fix any $(s, y) \in I \times U$ with $s < t$. Note that there exists a polygonal line $[x_0, x_1] \cup \cdots \cup [x_{n-1} \cup x_n = y]$ in U that connects x_0 and y (here, $[x, y]$ denotes the line segment between x and y). Select times $t_0 > t_1 > \cdots > t_n = s$ and notice that the (spacetime) polygonal line $[(t_0, x_0), (t_1, x_1)] \cup \cdots \cup [(t_{n-1}, x_{n-1}), (t_n, x_n)]$ lies inside $I \times U \cap \{t \leq t_0\}$ and connects (t_0, x_0) with (s, y).

To conclude the proof, it suffices to show that if $u(t_i, x_i) = M$, then $u(t, x) = M$ on the whole line segment $[(t_i, x_i), (t_{i+1}, x_{i+1})]$. To show this, we set a continuity argument. Denote by J the subset of $[(t_i, x_i), (t_{i+1}, x_{i+1})]$ on which $u(t, x) = M$. Let $t_* := \min\{t : (t, x) \in J \text{ for some } x\}$. Since J is nonempty, $t_* \leq t_i$; since J is clearly closed, there exists x_* such that $(t_*, x_*) \in [(t_i, x_i), (t_{i+1}, x_{i+1})]$ and $u(t_*, x_*) = M$. The proof will be complete if we show that $t_* = t_{i+1}$. Indeed, if $t_* > t_{i+1}$, then $u = M$ on $\mathcal{E}_r(t_*, x_*)$ for some $r > 0$. However, observe that $\mathcal{E}_r(t_*, x_*)$ intersects $[(t_*, x_*), (t_{i+1}, x_{i+1})]$ on a segment of nonzero length. Since $u = M$ on this intersection, we arrive at a contradiction. □

As a quick corollary of the maximum principle, we obtain the following uniqueness result.

Corollary 5.13 (Uniqueness of the initial boundary value problem)**.** *Let $I = (a, b)$, U a bounded domain, $f \in C(I \times U)$, $g \in C(\overline{U})$ and $h \in C(I \times \partial U)$. There exists at most one solution $u \in C^2(I \times U) \cap C(\overline{I\times U})$ to the initial boundary value problem*

$$\begin{cases} (\partial_t - \Delta)u = f \text{ in } U, \\ \qquad\quad u = g \text{ on } \{a\} \times \overline{U}, \\ \qquad\quad u = h \text{ on } I \times \partial U. \end{cases}$$

6. More distribution theory

We have already seen the usefulness of fundamental solutions and distribution theory in the study of various equations, including the Laplace/Poisson equations, the Cauchy–Riemann equation and the heat equation. In this section, we collect some more distribution-theoretic tools that is needed to study the wave equation[12].

6.1. **Change of variables and pullback of distributions.**

When solving problems, we often need to change coordinates to better suit our needs. The following proposition justifies the procedure of change of coordinates for distributions.

Proposition 6.1. *Let $\Phi : X_1 \to X_2$ be a diffeomorphism, where X_1, X_2 are open subsets of $\mathbb{R}^d$. To every distribution $u \in \mathcal{D}'(X_2)$ on X_2, there exists a way to associate a unique distribution $u \circ \Phi \in \mathcal{D}'(X_1)$ on X_1 so that $u \circ \Phi$ agrees with the usual composition for $u \in C_c^\infty(X_2) \subseteq \mathcal{D}'(X_2)$ and the following holds:*

The mapping $\mathcal{D}'(X_2) \to \mathcal{D}'(X_1)$, $u \mapsto u \circ \Phi$ is linear and continuous in u.

In fact, for $\phi \in C_c^\infty(X_1)$, $u \circ \Phi$ is defined by the formula

$$\langle u \circ \Phi, \phi \rangle = \left\langle u, \frac{1}{|\det \Phi|} \phi \circ \Phi^{-1} \right\rangle. \tag{6.1}$$

Proof. Uniqueness is clear by density of $C_c^\infty(X_2)$ in $\mathcal{D}'(X_2)$. Let $u_j \rightharpoonup u$ be a sequence of u_j in $C_c^\infty(X_2)$ converging to u in the sense of distributions. Write $\Phi(x) = (y^1(x), \cdots, y^d(x))$ and

$$\frac{\partial(y^1, \cdots, y^d)}{\partial(x^1, \cdots, x^d)} = \det \Phi, \quad \text{and} \quad \frac{\partial(x^1, \cdots, x^d)}{\partial(y^1, \cdots, y^d)} = \det \Phi^{-1}.$$

For any $\phi \in C_c^\infty(X_1)$, we have

$$\begin{aligned}\langle u_j \circ \Phi, \phi \rangle &= \int u_j(\Phi(x))\phi(x) \, \mathrm{d}x \\ &= \int u_j(y)\phi(\Phi^{-1}(y)) \frac{\partial(x^1, \cdots, x^d)}{\partial(y^1, \cdots, y^d)} \, \mathrm{d}y.\end{aligned}$$

Since $\phi(\Phi^{-1}(y)) \frac{\partial(x^1, \cdots, x^d)}{\partial(y^1, \cdots, y^d)}$ is a test function on X_2 (**Exercise:** Verify!), it follows that the last line goes to

$$\left\langle u(y), \phi(\Phi^{-1}(y)) \frac{\partial(x^1, \cdots, x^d)}{\partial(y^1, \cdots, y^d)} \right\rangle_y = \left\langle u, \frac{1}{|\det \Phi|} \phi \circ \Phi^{-1} \right\rangle,$$

as desired. □

As an immediate corollary, we have the following linear change of variables for formula for the delta distribution.

Corollary 6.2. *Let $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ be a invertible linear transformation. Then we have*

$$\delta_0 \circ \Phi = \frac{1}{|\det \Phi|} \delta_0.$$

Similarly, we can define the *pullback* (or the *composition*) $u \circ f$ of a distribution $u \in \mathcal{D}'(\mathbb{R})$ by a smooth map $f : U \to \mathbb{R}$ as follows:

[12] In the actual lectures, we will directly pass to Section 7 and come back to this section whenever we need more distribution-theoretic tools.

Proposition 6.3. *Let $f : U \to J$ be a smooth surjective map with $\mathrm{d}f(x) \neq 0$ for every $x \in U$, where U is an open subsets of $\mathbb{R}^d$ and $J \subseteq \mathbb{R}$ is an open interval. To every distribution $u \in \mathcal{D}'(J)$, there exists a way to associate a unique distribution $u \circ f \in \mathcal{D}'(U)$ on U so that $f^* h$ agrees with the usual composition for $u \in C^\infty_c(J) \subseteq \mathcal{D}'(J)$ and the following holds:*

The mapping $\mathcal{D}'(J) \to \mathcal{D}'(U)$, $u \mapsto u \circ f$ is linear and continuous in u.

To define $u \circ f$, let us first consider the case when $u \in C^\infty_c(J)$. Fix $x_0 \in U$. Since $\mathrm{d}f(x_0) \neq 0$, after possibly relabeling the axes, the map $\Phi : (f(x), x^2, x^3, \dots, x^d)$ is a diffeomorphism in a neighborhood V_{x_0} of x_0. For $\phi \in C^\infty_c(V_{x_0})$, we have

$$\begin{aligned}
&\langle u \circ f, \phi \rangle \\
&= \int u(f(x)) \varphi(x) \, \mathrm{d}x \\
&= \int u(s) \varphi(\Phi^{-1}(s, x^2, \dots, x^d)) \frac{1}{|\det \Phi(\Phi^{-1}(s, x^2, \dots, x^d))|} \, \mathrm{d}s \mathrm{d}x^2 \cdots \mathrm{d}x^d \\
&= \int u(s) \left(\int \varphi(\Phi^{-1}(s, x^2, \dots, x^d)) \frac{1}{|\det \Phi(\Phi^{-1}(s, x^2, \dots, x^d))|} \, \mathrm{d}x^2 \cdots \mathrm{d}x^d \right) \mathrm{d}s \\
&= \left\langle u, \int \varphi(\Phi^{-1}(\cdot, x^2, \dots, x^d)) \frac{1}{|\det \Phi(\Phi^{-1}(\cdot, x^2, \dots, x^d))|} \, \mathrm{d}x^2 \cdots \mathrm{d}x^d \right\rangle .
\end{aligned}$$

The last line makes sense even if u is just a distribution.

Motivated by the above consideration, for $u \in \mathcal{D}'(J)$ and $\phi \in C^\infty_c(V_{x_0})$, we *define*

$$\langle u \circ f, \phi \rangle := \left\langle u, \int \varphi(\Phi^{-1}(\cdot, x^2, \dots, x^d)) \frac{1}{|\det \Phi(\Phi^{-1}(\cdot, x^2, \dots, x^d))|} \, \mathrm{d}x^2 \cdots \mathrm{d}x^d \right\rangle .$$

Since this can be done for each $x_0 \in U$, we may define $\langle u \circ f, \phi \rangle$ for $\phi \in C^\infty_c(U)$ by using a smooth partition of unity.

With the above definition in hand, the proof of Proposition 6.3 proceeds as in Proposition 6.1. We omit the details.

Remark 6.4. In fact, given open subsets $X_1 \subseteq \mathbb{R}^{d_1}$ and $X_2 \subseteq \mathbb{R}^{d_2}$, $u \circ f \in \mathcal{D}'(X_1)$ can be defined for $u \in \mathcal{D}'(X_2)$ and $f : X_1 \to X_2$ with $Df(x)$ surjective for every $x \in X_1$ by a similar procedure; see [HÖ3, Theorem 6.1.2].

6.2. **Classification of distributions supported at a point: Application of Taylor expansion.**

Our goal is to prove the following result.

Theorem 6.5 (Classification of distributions supported at a point)**.** *Suppose that $u \in \mathcal{D}'(U)$ and* $\operatorname{supp} u = \{x_0\}$. *Then there exists $N \geq 0$ and $c_\alpha \in \mathbb{R}$ for $|\alpha| \leq N$ such that*

$$u = \sum_{\alpha : |\alpha| \leq N} c_\alpha D^\alpha \delta_{x_0}.$$

As we shall see, that u has finite order is an obvious consequence of Lemma 3.8. Establishing the following lemma, which follows from Taylor expansion, is the key step.

Lemma 6.6. *Let $u \in \mathcal{D}'(U)$ satisfy* $\operatorname{supp} u = \{x_0\}$ *and have order at most N. If $\psi \in C^\infty_c(U)$ satisfies $D^\alpha \psi(x_0) = 0$ for $|\alpha| \leq N$, then $\langle u, \psi \rangle = 0$.*

Proof. Without loss of generality, assume that $x_0 = 0$ and $B(0,1) \subset U$ (the latter can be ensured by rescaling the coordinate axes). Since the order of u is $\leq N$, on the compact ball $K = \overline{B(0,1)}$ there exists $C > 0$ such that

$$|\langle u, \phi \rangle| \leq C \sum_{|\alpha| \leq N} \sup_{x \in K} |D^\alpha \phi(x)| \quad \text{for every } \phi \in C_c^\infty(K). \tag{6.2}$$

Let us now localize this estimate. Fix a smooth function χ such that $\chi = 1$ on $B(0, \frac{1}{2})$ and $\operatorname{supp} \chi \subset B(0,1)$. For $\delta > 0$, define $\chi_\delta(x) := \chi(\delta^{-1}x)$. Clearly $\chi_\delta u = u$ and there exists $C' > 0$ such that $|D^\alpha \chi_\delta| \leq C' \delta^{-|\alpha|}$ for every α with $|\alpha| \leq N$. Then for every $\phi \in C^\infty(U)$, using (6.2), we derive

$$\begin{aligned} |\langle u, \phi \rangle| = |\langle u, \chi_\delta \phi \rangle| &\leq C \sum_{|\alpha| \leq N} \sup_{x \in K} |D^\alpha (\chi_\delta \phi)(x)| \\ &\leq C'' \sum_{|\alpha| \leq N} \sup_{|x| \leq \delta} \delta^{-N+|\alpha|} |D^\alpha \phi(x)|, \end{aligned} \tag{6.3}$$

where C'' depends only on C, C' and N.

We claim that the function ψ satisfying the hypothesis of the lemma has the property that

$$\sup_{|x| \leq \delta} |D^\alpha \psi(x)| \leq C''' \delta^{N+1-|\alpha|} \quad \text{for every } |\alpha| \leq N \text{ and } 0 < \delta < 1. \tag{6.4}$$

Then by applying (6.3) and taking $\delta \to 0$, we would obtain $\langle u, \psi \rangle = 0$ as desired.

To prove (6.4), we use Taylor expansion. For $x \in \mathbb{R}^d$, let us apply (3.6) with $k = N+1$ to the function $\sigma \mapsto \psi(\sigma x)$ and take $\sigma = 1$. Then

$$\psi(x) = \sum_{j=0}^{N} \frac{1}{j!} \frac{\mathrm{d}^j}{\mathrm{d}\sigma^j} \psi(\sigma x) \Big|_{\sigma=0} + \frac{1}{(N+1)!} \int_0^1 \frac{\mathrm{d}^{N+1}}{\mathrm{d}\sigma^{N+1}} \psi(\sigma x) \sigma^{N+1} \, \mathrm{d}\sigma.$$

The first term involves at most N derivatives on ψ evaluated at $x = 0$, which all vanish by the hypothesis. Hence, the first term vanishes. Using $\frac{\mathrm{d}}{\mathrm{d}\sigma} f(\sigma x) = \sum_i x^i \partial_i f(\sigma x)$, we may compute the second term and arrive at[13]

$$\psi(x) = \sum_{i_1, \ldots, i_{N+1} = 1}^{d} \frac{1}{(N+1)!} x^{i_1} \cdots x^{i_{N+1}} \int_0^1 \partial_{i_1} \cdots \partial_{i_{N+1}} \psi(\sigma x) \sigma^{N+1} \, \mathrm{d}\sigma.$$

It is easy to check that each $\int_0^1 \partial_{i_1} \cdots \partial_{i_{N+1}} \psi(\sigma x) \sigma^{N+1} \, \mathrm{d}\sigma$ is C^∞ on $B(0,1)$. Since ψ is the sum of the product of such functions with the monomials $x^{i_1} \cdots x^{i_{N+1}}$ of order $N+1$, the desired estimate (6.4) clearly follows. □

Proof of Theorem 6.5. Without loss of generality, let $x_0 = 0$. Fix a function $\chi \in C_c^\infty(U)$ such that $\chi = 1$ in a neighborhood of 0. By the support condition, we have, for every $\phi \in C_c^\infty(U)$, that

$$|\langle u, \phi \rangle| = |\langle u, \chi \phi \rangle|.$$

[13]Using the multi-index notation, we can write $\psi(x)$ more cleanly as

$$\psi(x) = \sum_{|\alpha| = N+1} \frac{1}{\alpha!} x^\alpha \int_0^1 D_\alpha \psi(\sigma x) \sigma^{N+1} \, \mathrm{d}\sigma$$

after some simple combinatorics, but it is not necessary.

Applying Lemma 3.8 to $K = \operatorname{supp} \chi$, there exists $N \in \mathbb{Z}_{\geq 0}$ and $C > 0$ such that

$$\begin{aligned}|\langle u, \phi \rangle| = |\langle u, \chi\phi \rangle| &\leq C \sum_{\alpha : |\alpha| \leq N} \sup_{x \in K} |\partial^\alpha (\chi\phi)(x)| \\ &\leq C' \sum_{\alpha : |\alpha| \leq N} \sup_{x \in U} |\partial^\alpha \phi(x)|.\end{aligned}$$

In particular, u is order at most N.

For any $\phi \in C_c^\infty(U)$, we write

$$\chi\phi(x) = \chi(x) \sum_{\alpha : |\alpha| \leq N} \frac{1}{\alpha!} D^\alpha \phi(0) x^\alpha + \psi(x).$$

Then ψ has the property that $D^\alpha \psi(0) = 0$ for $|\alpha| \leq N$. By Lemma 6.6, $\langle u, \psi \rangle = 0$. Thus,

$$\begin{aligned}\langle u, \phi \rangle = \langle u, \chi\phi \rangle &= \langle u, \chi(x) \sum_{\alpha : |\alpha| \leq N} \frac{1}{\alpha!} D^\alpha \phi(0) x^\alpha \rangle \\ &= \sum_{\alpha : |\alpha| \leq N} \frac{1}{\alpha!} \langle u, \chi(x) x^\alpha \rangle D^\alpha \phi(0),\end{aligned}$$

so the theorem holds with $c_\alpha = \frac{(-1)^{|\alpha|}}{\alpha!} \langle u, \chi(x) x^\alpha \rangle$. □

6.3. **Homogeneous distributions.** We now introduce the concept of *homogeneity* for distributions. Here, we shall consider **complex-valued** distributions with $a \in \mathbb{C}$ (which will be introduced below). Alternatively, we may consider real-valued distributions with $a \in \mathbb{R}$.

6.3.1. *General theory.* As usual, we start with the case of functions: A smooth function h on $\mathbb{R}^d \setminus \{0\}$ is said to be *homogeneous of degree* a if

$$h(\lambda x) = \lambda^a h(x) \quad \text{for every } x \neq 0,\ \lambda > 0.$$

We will use the adjoint method to extend this notion to distributions. For this purpose, note the following computation: If $\phi \in C_c^\infty(\mathbb{R}^d \setminus \{0\})$, then on the one hand, by change of variables,

$$\int h(\lambda x) \phi(x) \, \mathrm{d}x = \lambda^{-d} \int h(z) \phi(\lambda^{-1} z) \, \mathrm{d}z,$$

and on the other hand, by homogeneity,

$$\int h(\lambda x) \phi(x) \, \mathrm{d}x = \lambda^a \int h(x) \phi(x) \, \mathrm{d}x.$$

This computation motivates the following definition:

Definition 6.7. We say that $h \in \mathcal{D}'(\mathbb{R}^{1+d})$ (resp. $\mathcal{D}'(\mathbb{R}^{1+d} \setminus \{0\})$) is homogeneous of degree $a \in \mathbb{C}$ if for every $\phi \in C_c^\infty(\mathbb{R}^{1+d})$ (resp. $\phi \in C_c^\infty(\mathbb{R}^{1+d} \setminus \{0\})$) and $\lambda > 0$,

$$\lambda^{-d} \langle h, \phi(\lambda^{-1} \cdot) \rangle = \lambda^a \langle h, \phi \rangle.$$

We denote by h_λ the distribution defined by the LHS of the above equation, i.e., $\langle h_\lambda, \phi \rangle = \lambda^{-d} \langle h, \phi(\lambda^{-1} \cdot) \rangle$.

As a simple but important example, we note that δ_0 on $\mathbb{R}^d$ is homogeneous of degree $-d$ (see Proposition 6.1). Some more basic properties of homogeneous distributions are:

- If h is homogeneous of degree a, then $D^\alpha h$ is homogeneous of degree $a - |\alpha|$;
- If h is homogeneous of degree a, then we have the *Euler identity*:

$$\lambda \frac{\mathrm{d}}{\mathrm{d}\lambda} \langle h_\lambda, \phi \rangle = a \langle h_\lambda, \phi \rangle. \tag{6.5}$$

If h is a homogeneous function on $\mathbb{R}^d \setminus \{0\}$ with degree $a > -d$, then it defines a unique locally integrable function on $\mathbb{R}^d$. Similarly, a homogeneous distribution h on $\mathbb{R}^d \setminus \{0\}$ can be extended uniquely to a homogeneous distribution on the whole space $\mathbb{R}^d$ provided that its degree is greater than $-d$. In fact, the following more general result holds:

Lemma 6.8 (Homogeneous extension to the origin)**.** *If $h \in \mathcal{D}'(\mathbb{R}^d \setminus \{0\})$ is homogeneous of degree a, and a is not an integer less than or equal to $-d$, then h has a unique extension to a homogeneous distribution $\dot{h} \in \mathcal{D}'(\mathbb{R}^d)$ of degree a, so that the map $h \mapsto \dot{h}$ is continuous.*

For a proof, see [HÖ3, Theorem 3.2.3]. When a is an integer such that $a \leq -d$, then there many not(!) exist a homogeneous extension to $\mathcal{D}'(\mathbb{R}^d)$ of a homogeneous distribution $h \in \mathcal{D}'(\mathbb{R}^d \setminus \{0\})$; see Theorem 6.13 below.

6.3.2. *Families of homogeneous distributions on* $\mathbb{R}$. We now discuss the case $d = 1$. In this case, there is a general classification result. We begin with the following uniqueness result:

Proposition 6.9. *Let $h \in \mathcal{D}'(\mathbb{R})$ be a homogeneous distribution of degree a. Then the following statements hold:*

(1) h agrees with a smooth homogeneous function of degree a on $\mathbb{R} \setminus \{0\}$.
(2) If a is not a negative integer, $h \in \mathcal{D}'(\mathbb{R})$ is uniquely determined by $h(1)$ and $h(-1)$.
(3) If $a = -k$ (a negative integer), then any two homogeneous distributions $h, h' \in \mathcal{D}'(\mathbb{R})$ of degree $-k$ with $h(1) = h'(1)$ and $h(-1) = h'(-1)$ differs by a multiple of $\delta_0^{(k-1)}$.

Proof. Statements (2) and (3) follow from Theorem 6.5, so let us focus on the proof of Statement (1).

If u is a smooth function on $(0, \infty)$, then the homogeneity condition says that $u(\lambda) = \lambda^a u(1)$, so $\langle u, \phi \rangle = u(1) \langle x^a, \phi \rangle$. Moreover, introducing $\psi \in C_c^\infty((0, \infty))$ such that $\int \lambda^a \psi(\lambda) \, \mathrm{d}\lambda = 1$, we may write $u(1)$ as the integral

$$u(1) = \int u(\lambda) \psi(\lambda) \, \mathrm{d}\lambda.$$

In conclusion, for $\phi \in C_c^\infty((0, \infty))$, wehave

$$\langle u, \phi \rangle = \langle u, \psi \rangle \langle x^a, \phi \rangle. \tag{6.6}$$

Observe that $\langle u, \psi \rangle \langle x^a, \phi \rangle$ is an expression that make sense even if u is merely a distribution. Hence, our goal is to justify the above for $u \in \mathcal{D}'(\mathbb{R})$.

Let $\phi \in C_c^\infty((0, \infty))$. By the homogeneity condition, we have

$$\langle u, \phi \rangle = \lambda^{-a-1} \langle u, \phi(\lambda^{-1} \cdot) \rangle.$$

Multiplying both sides by $\lambda^a \psi(\lambda)$ and integrating in λ, we obtain

$$\langle u, \phi \rangle = \int_0^\infty \lambda^{-1} \langle u, \phi(\lambda^{-1} \cdot) \rangle \psi(\lambda) \, \mathrm{d}\lambda.$$

Consider an approximation $u_n \rightharpoonup u$ where $u_n \in C_c^\infty(\mathbb{R})$. Then

$$\begin{aligned}\int_0^\infty \lambda^{-1}\langle u_n(x), \phi(\lambda^{-1}x)\rangle \psi(\lambda)\,\mathrm{d}\lambda &= \int_0^\infty \lambda^{-1}\int u_n(x)\phi(\lambda^{-1}x)\,\mathrm{d}x\psi(\lambda)\,\mathrm{d}\lambda \\ &= \int_0^\infty\int_0^\infty \lambda^{-1}u_n(x)\phi(\lambda^{-1}x)\psi(\lambda)\,\mathrm{d}\lambda\mathrm{d}x \\ &= \int_0^\infty\int_0^\infty \mu^{-1}u_n(x)\phi(\mu)\psi(\mu^{-1}x)\,\mathrm{d}\mu\mathrm{d}x \\ &= \int_0^\infty \langle u_n(x), \mu^{-1}\psi(\mu^{-1}x)\rangle\phi(\mu)\,\mathrm{d}\mu.\end{aligned}$$

where we changed the variable from λ to μ where $\lambda = \frac{x}{\mu}$ in the third identity. Taking $n \to \infty$, it follows that

$$\langle u, \phi\rangle = \int \langle u, \mu^{-1}\psi(\mu^{-1}\cdot)\rangle\phi(\mu)\,\mathrm{d}\mu.$$

But by homogeneity, we have

$$\langle u, \mu^{-1}\psi(\mu^{-1}\cdot)\rangle = \mu^a\langle u, \psi\rangle,$$

from which (6.6) follows. The case of $(-\infty, 0)$ is handled similarly. □

We now consider examples of homogeneous distributions on $\mathbb{R}$.

Example 6.10 (degree a distributions supported in $[0,\infty)$, where $a \in \mathbb{C}\setminus\{-1,-2,\dots\}$)**.** We look for a homogeneous distribution h of degree a with $h(-1) = 0$ (or equivalently, supported in $[0,\infty)$). When $\operatorname{Re} a > -1$, an obvious example would be the following:

$$x_+^a(x) := \mathbf{1}_{(0,\infty)}x^a, \quad \operatorname{Re} a > -1.$$

The condition $\operatorname{Re} a > -1$ makes $x_+^a(x)$ locally integrable; hence $x_+^a \in \mathcal{D}'(\mathbb{R})$. Clearly, any homogeneous distribution of degree a with $\operatorname{Re} a > -1$ with $h(-1) = 0$ is a multiple of x_+^a.

How do we construct examples with $\operatorname{Re} a \leq -1$? We can differentiate. For $\operatorname{Re} a > 0$, we have

$$\begin{aligned}\langle \partial_x x_+^a(x), \phi\rangle &= \langle x_+^a(x), -\partial_x\phi\rangle \\ &= \int_0^\infty x^a(-\partial_x\phi)\,\mathrm{d}x \\ &= -x^a\phi\Big|_0^\infty + \int_0^\infty ax^{a-1}\phi\,\mathrm{d}x \\ &= \langle a x_+^{a-1}(x), \phi\rangle,\end{aligned}$$

or more succinctly,

$$x_+^{a-1}(x) = \frac{1}{a}\partial_x x_+^a(x) \quad \text{for } \operatorname{Re} a > 0.$$

We can try to extend $a \mapsto x_+^a$ to more general values of $a \in \mathbb{C}$ based on this functional equation. Concretely, we *define* x_+^a by

$$x_+^a := \left(\prod_{i=1}^N \frac{1}{(a+i)}\right)\partial_x^N x_+^{a+N}$$

in the sense of distributions. In order for the product and x_+^{a+N} to make sense we require that:

$$\operatorname{Re} a > -1 - N, \quad \text{ and } a \text{ is not a negative integer.}$$

Indeed, given $a \in \mathbb{C} \setminus \{-1, -2, \ldots\}$, note that x_+^a is well-defined (i.e., it is independent of the choice of N as long as $\operatorname{Re} a > -1 - N$). As before, it can be easily checked that any homogeneous distribution h of order $a \in \mathbb{C} \setminus \{-1, -2, \ldots\}$ with $\operatorname{supp} h \subseteq [0, \infty)$ is equal to x_+^a up to a constant.

Example 6.11 (degree a distributions supported in $[0, \infty)$ for all $a \in \mathbb{C}$)**.** The previous construction of homogeneous distributions supported in $[0, \infty)$ was satisfactory except for one point, namely, it misses the case $a \in \{-1, -2, \ldots\}$. To fix this point, let us think of a different way to normalize the family x_+^a. We define

$$\chi_+^a(x) = c(a) x_+^a(x) \quad \text{ for } \operatorname{Re} a > -1, \tag{6.7}$$

with $c(a)$ to be determined below. Then the functional equation becomes

$$\chi_+^{a-1}(x) = \frac{c(a-1)}{a c(a)} \partial_x \chi_+^a(x),$$

and we have

$$\chi_+^a(x) = \left(\prod_{i=1}^{N} \frac{c(a+i-1)}{(a+i) c(a+i)} \right) \partial_x^N \chi_+^{a+N} \tag{6.8}$$

for a such that $\operatorname{Re} a > -1$.

To extend χ_+^a to all $a \in \mathbb{C}$, the idea is to choose $c(a)$ in (6.7) so that the factor on the RHS of (6.8) is 1, i.e., $c(a) = (a+1) c(a+1)$. Taking the reciprocals, note that we are trying to find c such that

$$\frac{1}{c(a+1)} = (a+1) \frac{1}{c(a)}.$$

This is, in fact, exactly the problem of defining the extension (more precisely, analytic continuation) of the factorial! A well-known solution to this problem is given by the *Gamma function* $\Gamma(s)$, which is defined by

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} \, \mathrm{d}x \quad \text{ for } \operatorname{Re} s > 0.$$

The key property of the Gamma function is that it obeys the functional equation

$$\Gamma(s+1) = s \Gamma(s), \tag{6.9}$$

for $\operatorname{Re} s > 0$. Therefore, choosing

$$c(a) := \frac{1}{\Gamma(a+1)} \text{ for all } \operatorname{Re} a > -1$$

implies

$$c(a) = (a+1) c(a+1) \quad \text{ for all } \operatorname{Re} a > -1.$$

as desired.

In conclusion, if we define

$$\chi_+^a(x) := \frac{1}{\Gamma(a+1)} x_+^a(x) \quad \text{ for } \operatorname{Re} a > -1, \tag{6.10}$$

then this family of distributions satisfy the functional equation

$$\chi_+^a(x) = \partial_x^N \chi_+^{a+N}, \tag{6.11}$$

for any $\operatorname{Re} a > -1$ and $N = 0, 1, \ldots,$. As before, we may then take the RHS of the functional equation as the *definition* of $\chi_+^a(x)$ for $a \in \mathbb{C}$, i.e.,

$$\chi_+^a(x) := \partial_x^N \chi_+^{a+N} \quad \text{for } \operatorname{Re} a > -1 - N. \tag{6.12}$$

It is not difficult to check that this definition is independent of N as long as $\operatorname{Re} a > -1 - N$.

How does $\chi_+^{-a}(x)$ look like when a is a negative integer? Note that $\chi_+^{-1}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\chi_+^0(x) = \frac{\mathrm{d}}{\mathrm{d}x}H(x) = \delta_0(x)$. Then by the preceding identity, we see that

$$\chi_+^{-k}(x) = \delta_0^{(k-1)}(x). \tag{6.13}$$

Moreover, for negative half-integers, we have

$$\chi_+^{-\frac{1}{2}-k}(x) = \frac{\mathrm{d}^k}{\mathrm{d}x^k}\chi_+^{-\frac{1}{2}} = \frac{1}{\sqrt{\pi}}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\Big(H(x)\frac{1}{x^{1/2}}\Big) \tag{6.14}$$

For this identity, we used $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, which in turn follows from integration of a Gaussian (**Exercise:** Prove this!).

Remark 6.12. When $-k$ is a negative integer, $\chi_+^{-k} = \delta_0^{(k-1)}$ turns out to be the unique (up to multiplication by a constant) homogeneous distribution of degree $-k$ on $\mathbb{R}$ supported in $[0, \infty)$. See Theorem 6.13 below.

We remark that the families x_-^a and χ_-^a of homogeneous distributions of degree a supported in $(-\infty, 0]$ can be constructed in an entirely analogous manner starting from

$$x_-^a := \mathbf{1}_{(-\infty,0)}(x)(-x)^a \quad \text{for } \operatorname{Re} a > -1,$$
$$\chi_-^a := \frac{1}{\Gamma(a+1)}x_-(a) \quad \text{for } \operatorname{Re} a > -1.$$

In fact, the following complete *classification* of homogeneous distributions on $\mathbb{R}$ holds:

Theorem 6.13. *Let $h \in \mathcal{D}'(\mathbb{R})$ be a homogeneous distribution of degree a.*

(1) If $a \notin \{-1, -2, \ldots\}$, then

$$h = c_+ x_+^a + c_- x_-^a$$

for some scalars c_+, c_-.

(2) If $a = -k \in \{-1, -2, \ldots\}$, then

$$h = c\,\underline{x}^{-k} + c_0 \delta_0^{(k-1)}$$

for some scalars c, c_0, where $\underline{x}^{-k}$ is given by the formula

$$\underline{x}^{-k} = \frac{(-1)^{k-1}}{(k-1)!}\partial_x^k \log|x|.$$

Note that $\underline{x}^{-k} = x^{-k}$ on $\mathbb{R} \setminus \{0\}$, and $\underline{x}^{-1}$ agrees with the principal value distribution $\mathrm{pv}\frac{1}{x}$ *introduced earlier.*

As a corollary of this result, we see that χ_+^a (resp. χ_-^a) is, up to multiplication by a constant, the unique homogeneous distribution in $\mathcal{D}'(\mathbb{R})$ with homogeneity a such that $\operatorname{supp}\chi_+^a \subseteq [0, \infty)$ (resp. $\operatorname{supp}\chi_-^a \subseteq (-\infty, 0]$.

Proof (Optional). Statement (1) is obvious from the above construction and Proposition 6.9, so let us concentrate on Statement (2). We need to verify two assertions:

(i) $\underline{x}^{-k} \in \mathcal{D}'(\mathbb{R})$ is homogeneous of degree $-k$;

(ii) If $h \in \mathcal{D}'(\mathbb{R})$ is homogeneous of degree $-k$ and $\operatorname{supp} h \subseteq [0,\infty)$, then $h = c_0 \delta_0^{(k-1)}$ for some c_0.

Indeed, if we know (i) and (ii), then we may consider $u := h - h(-1)\underline{x}^{-k}$, which must be homogeneous of degree $-k$ and $\operatorname{supp} u \subseteq [0,\infty)$ by (i), so that $u = c_0\delta_0^{(k-1)}$ for some c_0 by (ii).

To verify (i), we check the homogeneity condition for $\underline{x}^{-k}$ explicitly as follows:

$$\begin{aligned}
\langle \underline{x}^{-k}, \lambda^{-1}\phi(\lambda^{-1}\cdot)\rangle &= -\frac{1}{(k-1)!}\int_{-\infty}^{\infty} \log|x| \partial_x^k \left(\phi(\lambda^{-1}x)\right) \lambda^{-1} \, \mathrm{d}x \\
&= -\frac{1}{(k-1)!}\int_{-\infty}^{\infty} \log|x| (\partial_x^k \phi)(\lambda^{-1}x)\lambda^{-1-k} \, \mathrm{d}x \\
&= -\lambda^{-k}\frac{1}{(k-1)!}\int_{-\infty}^{\infty} \log \lambda|z| (\partial_x^k \phi)(z) \, \mathrm{d}z \\
&= -\lambda^{-k}\frac{1}{(k-1)!}\int_{-\infty}^{\infty} \log|z| (\partial_x^k \phi)(z) \, \mathrm{d}z \\
&\quad -\lambda^{-k}\log\lambda\frac{1}{(k-1)!}\int_{-\infty}^{\infty} (\partial_x^k \phi)(z) \, \mathrm{d}z \\
&= \lambda^{-k}\langle \underline{x}^{-k}, \phi\rangle - \lambda^{-k}\log\lambda\frac{1}{(k-1)!}\int_{-\infty}^{\infty} (\partial_x^k \phi)(z) \, \mathrm{d}z.
\end{aligned}$$

But observe that the integral in the last expression is zero since $\phi \in C_c^\infty(\mathbb{R})$, which proves that $\underline{x}^{-k}$ is homogeneous of degree $-k$.

Finally, to prove (ii), suppose that (ii) is false. Then, in view of Proposition 6.9, there exists $h \in \mathcal{D}'(\mathbb{R})$ that is homogeneous of degree $-k$ such that $\operatorname{supp} h \subseteq [0,\infty)$ and $h(1) \neq 0$. We introduce

$$\widetilde{x}_+^{-k} := \frac{(-1)^{k-1}}{(k-1)!}\partial_x^k(\log|x|\mathbf{1}_{(0,\infty)}(x)).$$

Note that $\operatorname{supp} \widetilde{x}_+^{-k} \subseteq [0,\infty)$ and $\widetilde{x}_+^{-k} = x^{-k}$ on $(0,\infty)$. It follows that $h - h(1)\widetilde{x}_+^{-k}$ is supported in $\{0\}$, and hence by Theorem 6.5 it follows that

$$h = h(1)\widetilde{x}_+^{-k} + \sum_{j=0}^{J} c_j \partial_x^j \delta_0.$$

To conclude the proof, we will show by computation that this is impossible. Indeed, proceeding as in the case of $\underline{x}^{-k}$, we obtain

$$\begin{aligned}
\langle \widetilde{x}_+^{-k}, \lambda^{-1}\phi(\lambda^{-1}\cdot)\rangle &= \lambda^{-k}\langle \widetilde{x}_+^{-k}, \phi\rangle - \lambda^{-k}\log\lambda\frac{1}{(k-1)!}\int_0^\infty (\partial_x^k\phi)(z)\,\mathrm{d}z \\
&= \lambda^{-k}\langle \widetilde{x}_+^{-k}, \phi\rangle + \lambda^{-k}\log\lambda\frac{1}{(k-1)!}\partial_x^{k-1}\phi(0).
\end{aligned}$$

Note that, in contrast to the case of $\underline{x}^{-k}$, we pick up a boundary term at $x = 0$! Moreover, since

$$\langle \partial_x^j\delta_0, \lambda^{-1}\phi(\lambda^{-1}\cdot)\rangle = \lambda^{-j-1}(-1)^{j-1}\partial_x^j\phi(0),$$

it follows that

$$\langle h, \lambda^{-1}\phi(\lambda^{-1}\cdot)\rangle = h(1)\lambda^{-k}\langle \widetilde{x}_+^{-k}, \phi\rangle + h(1)\lambda^{-k}\log\lambda\frac{1}{(k-1)!}\partial_x^{k-1}\phi(0)$$

$$+\sum_{j=0}^{J} \lambda^{-j-1} c_j (-1)^{j-1} \partial_x^j \phi(0).$$

Hence,

$$\langle h, \lambda^{-1}\phi(\lambda^{-1}\cdot)\rangle - \lambda^{-k}\langle h, \phi\rangle = h(1)\lambda^{-k}\log\lambda \frac{1}{(k-1)!}\partial_x^{k-1}\phi(0) + \sum_{j\in\{0,\dots,J\},\, j\neq k-1} (\lambda^{-j-1} - \lambda^{-k}) c_j (-1)^{j-1} \partial_x^j \phi(0).$$

By homogeneity, this expression must be zero for all $\phi \in C_c^\infty(\mathbb{R})$. But this implies that $h(1) = 0$ (and, in fact, $c_j = 0$ for $j \neq k-1$), which is a contradiction. □

Example 6.14 (**Optional, using complex analysis:** $(x \pm i0)^a$)**.** In fact, the distinguished feature of the two families of distributions x_+^a and χ_+^a are that they are *analytic continuations*, i.e., for every $\phi \in C_c^\infty(\mathbb{R})$,

$$a \mapsto \langle x_+^a, \phi\rangle \text{ is analytic in } \mathbb{C} \setminus \{-1, -2, \dots\}, \text{ and}$$
$$a \mapsto \langle \chi_+^a, \phi\rangle \text{ is analytic in } \mathbb{C},$$

and these are the unique extensions of the original family on $\{\operatorname{Re} a > -1\}$ with this property.

There is another way to naturally define an analytic family of homogeneous distributions on $\mathbb{R}$, which is to consider the limits

$$(x \pm i0)^a := \lim_{\epsilon\to 0+} (x \pm i\epsilon)^a = \lim_{\epsilon\to 0+} e^{a \log(x \pm i\epsilon)},$$

where log denotes the principal branch of the logarithm (i.e., $\log x$ is real for $x > 0$ and undefined for $x \leq 0$). As we will see, this approach has the advantage of recovering $\underline{x}^{-k}$ and $\partial_x^{k-1}\delta_0$ (which are the only homogeneous distributions of degree $-k$ with $k = 1, 2, \dots$ by Theorem 6.13.(2)) in a natural way.

We begin by discussing the existence of this limit in the sense of distributions. In fact, we have the following general result:

Lemma 6.15. *Let I be an open interval in $\mathbb{R}$ and consider $\Omega = \{z \in \mathbb{C} : \operatorname{Re} z \in I,\, 0 < \operatorname{Im} z < a\}$ for some $a > 0$. If f is a holomorphic function in Ω such that, for some $N \in \mathbb{Z}_{\geq 0}$ and $C > 0$, we have*

$$|f(z)| \leq C(\operatorname{Im} z)^{-N} \text{ in } \Omega,$$

then $f(x + i\epsilon)$ has a limit $f_0 \in \mathcal{D}'(I)$ as $\epsilon \to 0+$, i.e.,

$$\lim_{\epsilon\to 0+} \langle f(x + i\epsilon), \phi(x)\rangle = \langle f_0(x), \phi(x)\rangle \quad \text{for all } \phi \in C_c^\infty(I).$$

Moreover, f_0 is of order at most $N + 1$.

Proof. When $N = 0$, this proposition is obvious, so suppose that $N \geq 1$. Fix $z_0 \in \Omega$ and consider $F(z) := \int_{z_0}^{z} f(z')\,\mathrm{d}z'$ (contour integral). Then F is holomorphic on Ω, $\frac{\mathrm{d}}{\mathrm{d}z}F(z) = f(z)$, and $|F(z)| \leq C'(\operatorname{Im} z)^{-N+1}$ if $N > 1$, and $|F(z)| \leq C'(1+\log(\operatorname{Im} z))$ if $N = 1$. When $N = 1$, the contour integral G of F is integrable, and we can define $f_0 := \lim_{\epsilon\to 0} G''(x + i\epsilon)$. In general, we consider the N-th anti-derivative G of F (i.e., $\frac{\mathrm{d}^N}{\mathrm{d}z^N}G = F$, so that $\frac{\mathrm{d}^{N+1}}{\mathrm{d}z^{N+1}}G = f$), and define $f_0 := \lim_{\epsilon\to 0} \frac{\mathrm{d}^{N+1}}{\mathrm{d}z^{N+1}}G(x+i\epsilon)$. □

Applying the lemma with $f(z) = z^a = e^{a \log z}$, we see that $(x \pm i0)^a$ is well-defined for all $a \in \mathbb{C}$. Moreover, observe that for every $\phi \in C_c^\infty(\mathbb{R})$, $a \mapsto \langle (x \pm i0)^a, \phi \rangle$ is an *entire* (i.e., analytic on $\mathbb{C}$) function since it is the limit of entire analytic functions.

How is $(x \pm i0)^a$ related with the previously considered families of homogeneous distributions? We begin with the easy observation that

$$(x \pm i0)^a = x_+^a + e^{\pm i\pi a} x_-^a \text{ for } \operatorname{Re} a > 0.$$

By the analyticity of both sides (with respect to a), it follows that

$$(x \pm i0)^a = x_+^a + e^{\pm i\pi a} x_-^a \text{ for } a \in \mathbb{C} \setminus \{-1, -2, \ldots\}.$$

Moreover, that $\langle (x \pm i0)^a, \phi \rangle$ is entire also tells us the interesting fact that, for every negative integer $-k$, the singular parts of x_+^a and $e^{\pm \pi a} x_-^a$ must cancel as $a \to -k$. In fact, we have the so-called *Plemelj relations*:

$$(x \pm i0)^{-k} = \underline{x}^{-k} \pm \pi i \frac{(-1)^k}{(k-1)!} \delta_0^{(k-1)}. \tag{6.15}$$

Let us verify these relations. We begin by observing that

$$\frac{\mathrm{d}}{\mathrm{d}a}(x \pm i0)^a = \lim_{\epsilon \to 0+} \frac{\mathrm{d}}{\mathrm{d}a}(x + i\epsilon)^a = \lim_{\epsilon \to 0+} a(x + i\epsilon)^{a-1} = a(x \pm i0)^{a-1},$$

and thus it suffices to verify (6.15) for $k = -1$. We write

$$(x \pm i\epsilon)^{-1} = \frac{x}{x^2 + \epsilon^2} \mp i \frac{\epsilon}{x^2 + \epsilon^2}.$$

Concerning the real part, observe that

$$\frac{x}{x^2 + \epsilon^2} = \partial_x \left(\frac{1}{2} \log(x^2 + \epsilon^2) \right),$$

and that $\lim_{\epsilon \to 0} \frac{1}{2} \log(x^2 + \epsilon^2) = \log |x|$ in the sense of distributions, while $\partial_x \log |x| = \underline{x}^{-1} = \text{p.v.} \frac{1}{x}$. Concerning the imaginary part, observe that (**Exercise:** Verify!)

$$\int \frac{\epsilon}{x^2 + \epsilon^2} \phi(x) \, \mathrm{d}x \to \pi \phi(0) \text{ as } \epsilon \to 0,$$

or in other words, $\frac{\epsilon}{x^2 + \epsilon^2} \rightharpoonup \pi \delta_0(x)$ in the sense of distributions.

6.4. **General structure theorems for distributions (optional).** We continue the theme of Section 6.2 and present more general structure theorems for distributions; morally, they tell us that distributions are given locally by the derivatives of continuous functions. Here, we will only cover the statement of the theorems and simply cite references for proofs.

From the very definition of the topology of $C_c^\infty(U)$, the following result is straightforward to derive:

Proposition 6.16. *If $u \in \mathcal{D}'(U)$ has a compact support, then u has a finite order.*

Proof. Since $\operatorname{supp} u$ is a compact subset of U, there exists a smooth function χ that equals 1 on $\operatorname{supp} u$ and $\operatorname{supp} \chi \subset U$. Clearly, $\chi u = u$. Applying Lemma 3.8 to the compact set $K = \operatorname{supp} \chi$, we see that there exists N and C such that for every $\phi \in C_c^\infty(U)$, we have

$$|\langle u, \phi \rangle| = |\langle u, \chi \phi \rangle| \leq C \sum_{|\alpha| \leq N} \sup_{x \in K} |D^\alpha (\chi \phi)(x)|.$$

Since $\sum_{|\alpha|\leq N}\sup_{x\in K}|D^{\alpha}(\chi\phi)(x)| \leq C'\sum_{|\alpha|\leq N}\sup_{s\in U}|D^{\alpha}\phi(x)|$ with C' depending only on $\sum_{|\alpha|\leq N}\sup_{x\in K}|D^{\alpha}\chi(x)|$, we have

$$\langle u,\phi\rangle \leq CC'\sum_{|\alpha|\leq N}\sup_{s\in U}|D^{\alpha}\phi(x)|,$$

which implies that u has order $\leq N$. □

Theorem 6.17 (Structure theorem for distributions with compact support)**.** *Suppose that $u\in\mathcal{D}'(U)$ and* supp u *is compact. Then there exist finitely many continuous functions f_α in U such that*

$$u=\sum_{\alpha}D^{\alpha}f_{\alpha}.$$

We note that, in general, supp f_α does *not* coincide with supp u; a simple example is the computation $\delta_0=\frac{\mathrm{d}^2}{\mathrm{d}x^2}x_+$ on $\mathbb{R}$, where $x_+=\max\{0,x\}$.

For a proof, see [Rud91, Theorems 6.26, 6.27].

Theorem 6.18 (Structure theorem for distributions)**.** *Suppose that $u\in\mathcal{D}'(U)$. Then for every multi-index α, there exists $g_\alpha\in C(U)$ such that*

- *each compact subset K of U intersects the support of only finitely many g_α; and*
- *we have*

$$u=\sum_{\alpha}D^{\alpha}g_{\alpha}.$$

If u has finite order, then the functions g_α can be chosen so that only finitely many are non-zero.

Theorem 6.18 makes precise the sense in which distribution theory is the completion of differential calculus: Every continuous function is differentiable, and every distribution is given locally by a finite sum of derivatives of continuous functions.

For a proof, see [Rud91, Theorem 6.28].

7. The wave equation

The subject of this section is the d'Alembertian on $\mathbb{R}^{1+d}$,

$$\Box\varphi=-\partial_t^2\varphi+\Delta\varphi,$$

and the associated wave equation,

$$\Box\varphi=f.$$

Our goals are as follows:

- to find an explicit (forward) fundamental solution for $\Box$;
- to find a representation formula for the Cauchy problem for $\Box$:

$$\begin{cases}\Box\phi=f & \text{in } \mathbb{R}^{1+d}_+=\{(t,x^1,\dots,x^d)\in\mathbb{R}^{1+d}: t>0\},\\ \phi=g & \text{on } \partial\mathbb{R}^{1+d}_+=\{0\}\times\mathbb{R}^d,\\ \partial_t\phi=h & \text{on } \partial\mathbb{R}^{1+d}_+=\{0\}\times\mathbb{R}^d;\end{cases}\tag{7.1}$$

- to prove the existence and uniqueness of a solution u to (7.1) under suitable conditions on f,g,h.

These three goals are, of course, related. As we have seen in Sections 3.11, 4.2 and 4.4, once we have a fundamental solution, there is a systematic procedure for deriving solution and representation formulae.

Other important ways to study the wave equation, namely the *Fourier and energy methods*, will be discussed later.

Remarks on the notation. In this section, we will use φ, ψ to refer to solutions to the wave equation instead of u, v, since we wish to reserve the letters u, v for the *null coordinates* $t - r$ and $t + r$, as is standard in the field. We also define

$$\Box = -\partial_t^2 + \Delta$$

which differs from the definition used by Evans by a sign. We also write x^0 and t interchangeably.

7.1. **Fundamental solutions on** $\mathbb{R}^{1+1}$**.** As a warm-up, we first consider the $(1+1)$-dimensional case. This case is simple to analyze, but nevertheless gives us intuition about what to expect in the more difficult case of $\mathbb{R}^{1+d}$ for $d \geq 2$.

d'Alembert's formula. In $\mathbb{R}^{1+1}$, the d'Alembertian takes the form

$$\Box = -\partial_t^2 + \partial_x^2. \tag{7.2}$$

We can formally factor $\partial_t^2 - \partial_x^2 = (\partial_t - \partial_x)(\partial_t + \partial_x)$. It will be convenient if we find a different coordinate system in which $\partial_t - \partial_x$ and $\partial_t + \partial_x$ are coordinate derivatives. To this end, we consider the *null coordinates*

$$u = t - x, \quad v = t + x. \tag{7.3}$$

Then we have

$$(t, x) = \left(\frac{u+v}{2}, \frac{v-u}{2} \right), \quad \partial_u = \frac{1}{2}(\partial_t - \partial_x), \quad \partial_v = \frac{1}{2}(\partial_t + \partial_x).$$

Hence the d'Alembertian (7.2) takes the simple form

$$\Box = -4\partial_u \partial_v. \tag{7.4}$$

Using this idea, now let us solve the equation

$$\Box E_0 = \delta_0.$$

We start by making the change of variables into $(u, v) = (t - x, t + x)$. The LHS becomes $4\partial_u \partial_v E_0(u, v)$. We need to be careful about the RHS; even though $(t, x) = (0, 0)$ if and only if $(u, v) = (0, 0)$, the delta distribution transforms as

$$\delta_0(t, x) = 2\delta_0(u, v). \tag{7.5}$$

A quick way to see this[14] is to use the approximation method and Lemma 3.19: Given $\chi \in C_c^\infty(\mathbb{R}^2)$ with $\int \chi = 1$,

$$\begin{aligned} \delta_0(t, x) &= \lim_{\epsilon \to 0+} \epsilon^{-2} \chi(\epsilon^{-1} t, \epsilon^{-1} x) \\ &= \lim_{\epsilon \to 0+} \epsilon^{-2} \chi \left(\epsilon^{-1} \frac{u+v}{2}, \epsilon^{-1} \frac{v-u}{2} \right) \\ &= \delta_0(u, v) \int \chi \left(\frac{u+v}{2}, \frac{v-u}{2} \right) \mathrm{d}u \mathrm{d}v, \end{aligned}$$

[14]For a more systematic way that doesn't use Lemma 3.19, see Proposition 6.1.

where we used Lemma 3.19. But by the change of variables formula for integrals,

$$\int \chi\left(\frac{u+v}{2}, \frac{v-u}{2}\right) \, \mathrm{d}u\mathrm{d}v = \int \chi(x,y) \left|\det \frac{\partial(u,v)}{\partial(x,y)}\right| \, \mathrm{d}x\mathrm{d}y = 2\int \chi(x,y) \, \mathrm{d}x\mathrm{d}y = 2.$$

In conclusion, we want to solve the equation

$$\partial_u \partial_v E_0(u,v) = -\frac{1}{2}\delta_0(u,v). \tag{7.6}$$

In view of the factorization $\partial_u \partial_v$, we can impose the ansatz that $E_0(u,v)$ is of the form $-\frac{1}{2}E_1(u)E_2(v)$, where

$$\partial_u E_1 = \delta_0(u), \quad \partial_v E_2 = \delta_0(v),$$

where $\delta_0(u), \delta_0(v)$ are delta distributions on $\mathbb{R}$, so that $\delta_0(u)\delta_0(v) = \delta_0(u,v)$. We know solutions to $\partial_u E_1 = \delta_0(u)$ on $\mathbb{R}$ are of the form

$$E_1(u) = \mathbf{1}_{(0,\infty)}(u) + c_1$$

where H is the Heaviside function and a constant $c_u \in \mathbb{R}$. Similar statement applies to $E_2(v)$. Hence

$$E_0(u,v) = -\frac{1}{2}(\mathbf{1}_{(0,\infty)}(u) + c_1)(\mathbf{1}_{(0,\infty)}(v) + c_2).$$

Luckily, $E_0(u,v)$ is a function, so we can change the variables back to (t,x) in the usual way and arrive at

$$E_0(t,x) = -\frac{1}{2}(\mathbf{1}_{(0,\infty)}(t-x) + c_1)(\mathbf{1}_{(0,\infty)}(t+x) + c_2).$$

Any choice of the constants $c_1, c_2 \in \mathbb{R}$ gives a fundamental solution for $\Box$ at 0. However, if we look for a *forward fundamental solution* (i.e., $\operatorname{supp} E_+ \subseteq \{t \geq 0\}$), we are forced to choose $c_1 = c_2 = 0$. Hence, we finally arrive at the following expression for the *forward fundamental solution* for $\Box$:

$$E_+(t,x) = -\frac{1}{2}\mathbf{1}_{(0,\infty)}(t-x)\mathbf{1}_{(0,\infty)}(t+x). \tag{7.7}$$

Note that for every compact interval $I \subseteq (-\infty,\infty)_t$, $E_+(t,\cdot)$ for all $t \in I$ is supported in a common bounded set; in fact, $\operatorname{supp}_x E_+(t,\cdot) \subseteq [-t,t]$ for $t \geq 0$ and empty for $t < 0$. This reflects the *finite speed of propagation* for the wave equation, i.e., the disturbance at $x = 0$ at time $t = 0$ can reach at best $|x| \leq t$ at time t.

Using the forward fundamental solution E_+, we can derive representation and solution formulae for (7.1).

Theorem 7.1 (d'Alembert's formula)**.** *For any $\phi \in C^\infty(\mathbb{R}^{1+1}_+)$ and $(t,x) \in \mathbb{R}^{1+1}_+$, we have the formula*

$$\phi(t,x) = -\frac{1}{2}\int_0^t \int_{x-t+s}^{x+t-s} \Box\phi(s,y) \, \mathrm{d}y\mathrm{d}s + \frac{1}{2}(\phi(0,x-t)+\phi(0,x+t)) + \frac{1}{2}\int_{x-t}^{x+t} \partial_t \phi(0,y) \, \mathrm{d}y. \tag{7.8}$$

Conversely, given any initial data $(g,h) \in C^\infty(\mathbb{R})$ and $f \in C^\infty(\mathbb{R}^{1+1})$, there exists a unique solution ϕ to the initial value problem (7.1) *defined by the RHS of the formula* (7.8) *with $\Box\phi(s,y) = f(s,y)$, $\phi(0,x) = g(x)$ and $\partial_t \phi(0,x) = h(x)$.*

Since $\Box$ is symmetric under time reversal $t \mapsto -t$, the same results applies to $\mathbb{R}^{1+1}_- = \{(t,x) \in \mathbb{R}^{1+1} : t < 0\}$. Of course, the regularity assumptions can be improved, but let us not worry about it for now.

As in the preceding sections, we will prove Theorem 7.1 as a consequence of 1) the existence of a forward fundamental solution and 2) the symmetry of $\Box$. We defer the proof of Theorem 7.1 until Section 7.3, after we find E_+ for $d > 1$.

Remark 7.2. The uniqueness of the choices of c_1, c_2 is no coincidence. Note that Theorem 7.1 implies that any solution to $\Box\phi = 0$ that is supported in $\{t \geq 0\}$ is zero; thus, the uniqueness of the forward fundamental solution follows. This property is analogous to the symmetry and uniqueness of the Green's function in Section 4.4.

7.2. **Forward fundamental solution for $d > 1$.** Our goal now is to construct the forward fundamental solution E_+ to the d'Alembertian on $\mathbb{R}^{1+d}$ for every $d \geq 1$. We will make an educated guess of the form of E_+, based on the symmetries of the d'Alembertian $\Box$.

Symmetries of the d'Alembertian. As we have seen in our study of the Laplace equation (Section 4.1), symmetries of the operator plays a key role in finding an explicit fundamental solution, which then opens up the door to a myriad of further applications. So let us begin our study of $\Box$ by discussing its symmetries.

Clearly, since $\Box$ is a constant coefficient partial differential operator, it is invariant under *translations*. Other types of symmetries can be found by requiring that the space-time origin remains fixed. Note that these symmetries will be useful for the purpose of finding a solution to $\Box E_0 = \delta_0$, since δ_0 will be invariant under those.

The symmetries of $\Box$ that fixes the space-time origin turn out to be precisely the linear transformations $L : \mathbb{R}^{1+d} \to \mathbb{R}^{1+d}$ which leave invariant the scalar quantity[15]

$$s^2(t,x) := t^2 - |x|^2. \tag{7.9}$$

These transformations are called *Lorentz transformations*. (**Exercise:** From the defining property $s^2(t,x) = s^2(L(t,x))$, show that $\Box(\phi \circ L) = (\Box\phi) \circ L$.) The Lorentz transformations form a group (by composition), which we will denote by $O(1,d)$. The group $O(1,d)$ is generated by the following elements:

(1) **Rotations.** Linear transformation $R : \mathbb{R}^{1+d} \to \mathbb{R}^{1+d}$ represented by the matrix

$$R = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \widetilde{R} & \\ 0 & & & \end{pmatrix} \tag{7.10}$$

where $\widetilde{R} \in O(d)$ is a $d \times d$ orthonormal matrix.

(2) **Reflection.** Linear transformation $\rho_k : \mathbb{R}^{1+d} \to \mathbb{R}^{1+d}$ $(k = 0, \ldots, d)$ defined by

$$(x^0, \cdots, x^d) \mapsto (x^0, \cdots, -x^k, \cdots, x^d). \tag{7.11}$$

In particular, the reflection of the $t = x^0$ variable is the *time reversal* symmetry of $\Box$.

[15] This quantity, of course, has a geometric meaning. It is precisely the 'space-time distance' from the origin to the event (t,x) in special relativity.

(3) **Lorentz boosts.** These symmetries correspond to choosing another frame of reference, which travels at a constant velocity compared to the original frame. If the new frame moves at speed $\gamma \in (0,1)$ in the x^1 direction, then its matrix representation is

$$\Lambda_{01}(\gamma) = \begin{pmatrix} \frac{1}{\sqrt{1-\gamma^2}} & -\frac{\gamma}{\sqrt{1-\gamma^2}} & 0 & \cdots & 0 \\ \frac{-\gamma}{\sqrt{1-\gamma^2}} & \frac{1}{\sqrt{1-\gamma^2}} & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & \mathrm{Id}_{d-1\times d-1} & \\ 0 & 0 & & & \end{pmatrix} \tag{7.12}$$

All Lorentz boosts then take the form $\Lambda(\gamma) = cR\Lambda_{01}(\gamma)R^{-1}$ for some constant $c \neq 0$ and rotation R.

For more on Lorentz transformations, we refer to [O'N83, Chapter 9].

Observe that all of the above symmetries are linear maps with determinant ± 1 (**Exercise:** Prove this statement!). Therefore, δ_0 is also *invariant* under these symmetries. For this reason, it is natural to look for E_+ which is invariant under these symmetries.

Scaling. Although it is not exactly a symmetry of $\Box$, we also point out that $\Box$ transforms in a simple way under *scaling*, i.e.,

$$\Box(\phi(t/\lambda, x/\lambda)) = \lambda^{-2}(\Box\phi)(t/\lambda, x/\lambda) \quad \text{for } \lambda > 0.$$

In view of this property, it is natural to look for E_+ which is *homogeneous.* From the equation

$$\Box E_+ = \delta_0,$$

observe that the right-hand side, being a delta distribution on $\mathbb{R}^{1+d}$, is homogeneous of degree $-d-1$. Since $\Box$ lowers the degree of homogeneity by 2, we see that

$$\text{If } E_+ \text{ is homogeneous, then it must be of degree } -d+1. \tag{7.13}$$

A nice feature of assuming E_+ to be homogeneous of degree $-d+1$, which is larger than $-d-1$, is that then E_+ is uniquely determined by its restriction to $\mathbb{R}^{1+d} \setminus \{(0,0)\}$; see Lemma 6.8.

Heuristic derivation. Motivated by the above consideration, we shall look for E_+ that is invariant under rotations, reflections and Lorentz boosts, and which is also homogeneous of degree $-d+1$. Recall that the rotations, reflections and Lorentz transformations are precisely the linear transformations which leave the scalar quantity $s^2(t,x) := t^2 - |x|^2$ invariant. Note, moreover, that $t^2 - |x|^2$ is homogeneous of degree 2. Combined with the earlier observation (7.13), we see that a reasonable guess is

$$\widetilde{E}_+(t,x) = \mathbf{1}_{(0,\infty)}(t) h^{-\frac{d-1}{2}}(t^2 - |x|^2) \quad \text{in } \mathbb{R}^{1+d} \setminus \{(0,0)\}.$$

where

$$h^{-\frac{d-1}{2}} \in \mathcal{D}'(\mathbb{R}) \text{ is homogeneous of degree } -\tfrac{d-1}{2}, \tag{7.14}$$

and $\mathbf{1}_{(0,\infty)}(t)$ has been multiplied to ensure the forward support condition. Here, the composition of the distribution $h^{\frac{d-1}{2}}$ with $t^2 - |x|^2$ on $\mathbb{R}^{1+d} \setminus \{(0,0)\}$ is to be interpreted as in Section 6.1. Observe that, since $\mathbf{1}_{(0,\infty)}$ is also homogeneous of degree 0 and Lorentz invariant, $\tilde{E}_+$ is still homogeneous of degree $-d+1$ and Lorentz invariant.

To pin down the homogeneous distribution $h^{-\frac{d-1}{2}}$, we note that, in order to avoid introducing jump discontinuities of $\widetilde{E}_+$ across $\{t=0\} \times \mathbb{R}^d \setminus \{(0,0)\}$ (which would make it impossible for $\widetilde{E}_+$ to satisfy $\Box E_+ = 0$ in a neighborhood of that region), we need

$$\operatorname{supp} h^{-\frac{d-1}{2}} \subseteq [0,\infty). \tag{7.15}$$

As discussed in Section 6.3, the family of homogeneous of distributions on $\mathbb{R}$ with the above support property that is well-defined for all negative half-integers would be $h^a = \chi_+^a$ (x_+^a will *not* be well-defined when a is a negative integer). Hence, we choose

$$h^{-\frac{d-1}{2}} = \chi_+^{-\frac{d-1}{2}}, \tag{7.16}$$

which is unique choice up to multiplication by a constant; see Section **??**.

Let us finally try to compute $\Box\widetilde{E}_+$. By construction, $\Box\widetilde{E}_+$ is a homogeneous distribution of degree $-d-1$. Restricted to $\mathbb{R}^{1+d} \setminus \{(0,0)\}$, we first note that

$$\begin{aligned}\Box\widetilde{E}_+ &= (-\partial_t^2 + \Delta)(\mathbf{1}_{\{t\geq 0\}}\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2)) \\ &= \mathbf{1}_{\{t\geq 0\}}(-\partial_t^2+\Delta)\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2),\end{aligned}$$

since Δ easily commutes with $1_{\{t\geq 0\}}$, and whenever ∂_t falls on $1_{\{t\geq 0\}}$ the result is zero thanks to the support property of $\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2)$. Using the chain rule, which is easily justified by approximation by C_c^∞ functions, on $\mathbb{R}^{1+d}\setminus(0,0)$ we have

$$\begin{aligned}&(-\partial_t^2+\Delta)\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2) \\ &= -\partial_t(2t(\chi_+^{-\frac{d-1}{2}})'(t^2-|x|^2)) - \nabla_x\cdot(2x(\chi_+^{-\frac{d-1}{2}})'(t^2-|x|^2)) \\ &= -(\chi_+^{-\frac{d-1}{2}})''(t^2-|x|^2)4t^2 - 2(\chi_+^{-\frac{d-1}{2}})'(t^2-|x|^2) \\ &\quad +(\chi_+^{-\frac{d-1}{2}})''(t^2-|x|^2)4|x|^2 - d(\chi_+^{-\frac{d-1}{2}})'(t^2-|x|^2) \\ &= -4(t^2-|x|^2)(\chi_+^{-\frac{d-1}{2}})''(t^2-|x|^2) - 2(d+1)(\chi_+^{-\frac{d-1}{2}})'(t^2-|x|^2).\end{aligned}$$

Now by the Euler identity, $(\chi_+^{-\frac{d-1}{2}})'$ satisfies the identity

$$s(\chi_+^{-\frac{d-1}{2}})''(s) = -\frac{d+1}{2}(\chi_+^{-\frac{d-1}{2}})'(s). \tag{7.17}$$

Therefore, the last line equals 0. In conclusion, $\Box\widetilde{E}_+$ is a distribution on $\mathbb{R}^{1+d}$ that is supported in $\{(0,0)\}$. By Theorem 6.5, $\Box\widetilde{E}_+$ is the (finite) linear combination of the delta distribution and its derivatives. Recalling that $\widetilde{E}_+$ is homogeneous of degree $-d-1$, it follows that

$$\widetilde{E}_+ = c\delta_0, \tag{7.18}$$

for some $c \in \mathbb{R}$, which is almost what we want!

To complete the derivation, it remains to show that, with the above choice of $\chi_+^{-\frac{d-1}{2}}$, (7.18) holds with $c \neq 0$; in fact,

$$\Box\left((\mathbf{1}_{(0,\infty)}\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2)\right) = -\frac{2}{\pi^{\frac{1-d}{2}}}\delta_0.$$

However, since the precise computation of this constant is rather detached from our discussion, we will leave its proof as an optional reading (see Section 7.4).

Remark 7.3. Computing the exact constant c requires explicit computation, but the fact that $c \neq 0$ (and hence that an appropriate c_d exists) can be seen by much softer methods. For example, it is sufficient to establish the following uniqueness statement: If $E \in \mathcal{D}'(\mathbb{R}^{1+d})$ is a solution to $\Box E = 0$ with $\operatorname{supp} E \subseteq \{|x| \leq t\}$, then $E = 0$. This statement can be proved by the Fourier or energy methods, which will be discussed later and which are independent of the existence of E_+.

In conclusion, the homogeneous distribution E_+ of degree $-d+1$ on $\mathbb{R}^{1+d}$, which takes the form

$$E_+ = -\frac{1}{2\pi^{\frac{d-1}{2}}} \mathbf{1}_{(0,\infty)} \chi_+^{-\frac{d-1}{2}} (t^2 - |x|^2) \quad \text{in } \mathbb{R}^{1+d} \setminus \{(0,0)\}, \tag{7.19}$$

where $\chi_+^{-\frac{d-1}{2}}$ is given by either (6.13) or (6.14), is the forward fundamental solution for $\Box$.

Applications of (7.19). We now discuss applications of the explicit formula (7.19) for E_+. Note that

$$\operatorname{supp} E_+ \subseteq \{(t,x) \in \mathbb{R}^{1+d} : |x| \leq t\},$$

so (7.25) immediately follows. Moreover, we have the following corollary of the representation formula (Theorem 7.11):

Corollary 7.4 (Finite speed of propagation)**.** *Suppose that a forward fundamental solution E_+ with the properties* (7.22), (7.23) *exists. Let $\phi \in C^\infty(\mathbb{R}^{1+d})$ solve the inhomogeneous wave equation $\Box\phi = f$ with initial data $(\phi, \partial_t \phi)|_{\{t=0\}} = (g,h)$, and consider a point $(t,x) \in \mathbb{R}^{1+d}$ such that $t > 0$. If*

$$\begin{aligned} f(s,y) =& 0 \quad \textit{in } \{(s,y) : 0 < s < t,\ |y-x| \leq t-s\}, \\ (g,h)(y) =& (0,0) \quad \textit{in } \{y : |y-x| \leq t\} \end{aligned}$$

then $\phi(t,x) = 0$.

Next, note that for $d \geq 3$ an odd integer, we have (in $\mathbb{R}^{1+d} \setminus \{(0,0)\}$)

$$\begin{aligned} E_+(t,x) &= -\frac{1}{2\pi^{\frac{d-1}{2}}} \mathbf{1}_{(0,\infty)} \chi_+^{-\frac{d-1}{2}} (t^2 - |x|^2) \\ &= -\frac{1}{2\pi^{\frac{d-1}{2}}} \mathbf{1}_{(0,\infty)} \delta_0^{(\frac{d-3}{2})} (t^2 - |x|^2) \end{aligned}$$

which is supported only on the boundary $\{(t,x) : |x| = t\}$ of the cone $\{(t,x) : |x| \leq t\}$. Hence a sharper version of Corollary 7.4 holds in this case. This phenomenon is called the *strong Huygens principle*; we record the precise statement in the following corollary.

Corollary 7.5 (Strong Huygens principle)**.** *Let $d \geq 3$ be an odd integer. Let $\phi \in C^\infty(\mathbb{R}^{1+d})$ solve the inhomogeneous wave equation $\Box\phi = f$ with initial data $(\phi, \partial_t\phi)|_{\{t=0\}} = (g,h)$, and consider a point $(t,x) \in \mathbb{R}^{1+d}$ such that $t > 0$. If*

$$\begin{aligned} f(s,y) =& 0 \quad \textit{in } \{(s,y) : 0 < s < t,\ |y-x| = t-s\}, \\ (g,h)(y) =& (0,0) \quad \textit{in } \{y : |y-x| = t\} \end{aligned}$$

then $\phi(t,x) = 0$.

It turns out that this property does *not* hold when $d \geq 2$ is even. Indeed, then by (6.14)

$$\begin{aligned} E_+(t,x) &= -\frac{1}{2\pi^{\frac{d-1}{2}}}\mathbf{1}_{(0,\infty)}\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2) \\ &= -\frac{1}{2\pi^{d/2}}\mathbf{1}_{(0,\infty)}\left(\frac{\mathrm{d}}{\mathrm{d}x^k}H(x)\frac{1}{x^{1/2}}\right)(t^2-|x|^2), \end{aligned}$$

where $\operatorname{supp} \frac{\mathrm{d}}{\mathrm{d}x^k}\frac{1}{x_+^{1/2}} = [0,\infty)$, so that $\operatorname{supp} E_+ = \{(t,x) \in \mathbb{R}^{1+d} : |x| \leq t\}$.

We also note that the continuity assumption (**??**) can be verified for E_+ using its homogeneity property (the point being that its degree $-d+1$ is greater than $-d$, so that $\delta_{t=t_0}E_+$ is well-defined for all t_0). We leave the straightforward task of verifying this property as an exercise.

Finally, we specialize to the cases $d = 1, 2, 3$ and derive classical representation formulae for the wave equation. Let us use the notation

$$c_d = -\frac{1}{2\pi^{\frac{d-1}{2}}}.$$

Explicit computation for $d = 1$. We now compute the form of the forward fundamental solution E_+ explicitly in dimension $d = 1$. When $d = 1$, we have

$$E_+(t,x) = c_1\mathbf{1}_{(0,\infty)}\,\chi_+^0(t^2-|x|^2) = c_1\mathbf{1}_{\{t>0\}}H(t^2-|x|^2) = c_1\mathbf{1}_{\{(t,x):0\leq|x|\leq t\}}.$$

As $c_1 = -\frac{1}{2}$, we recover the previous computation.

Explicit computation for $d = 2$. Next, we compute the form of the forward fundamental solution E_+ explicitly in dimension $d = 2$. Recalling the definition of $\chi_+^{-\frac{1}{2}}$, we have

$$\begin{aligned} E_+(t,x) =& c_2\mathbf{1}_{(0,\infty)}\,\chi_+^{-\frac{1}{2}}(t^2-|x|^2) \\ =& \frac{c_2}{\Gamma(\frac{1}{2})}\mathbf{1}_{(0,\infty)}\frac{1}{(t^2-|x|^2)_+^{\frac{1}{2}}} \\ =& -\frac{1}{2\pi}\mathbf{1}_{\{(t,x):0\leq|x|\leq t\}}\frac{1}{(t^2-|x|^2)^{\frac{1}{2}}}, \end{aligned}$$

outside the origin, and at the origin E_+ is determined by homogeneity.

Explicit computation for $d = 3$. Finally, we compute the form of the forward fundamental solution E_+ explicitly in dimension $d = 3$. Recall that $\chi_+^{-1} = \delta_0$; hence

$$E_+(t,x) = c_3\mathbf{1}_{(0,\infty)}(t)\,\delta_0(t^2-|x|^2) = -\frac{1}{2\pi}\mathbf{1}_{(0,\infty)}(t)\,\delta_0(t^2-|x|^2),$$

outside the origin, and at the origin E_+ is determined by homogeneity. Concretely, $\langle E_+, \phi\rangle$ can be computed in the following way.

Lemma 7.6. *On $\mathbb{R}^{1+3} \setminus \{0\}$, we have the identity*

$$\delta_0(t^2-|x|^2) = \frac{1}{2\sqrt{2}t}\mathrm{d}\sigma_{C_0^+}(t,x) \tag{7.20}$$

where $C_0^+ := \{(t,x) : t = |x|, t > 0\}$ is a forward cone and $\mathrm{d}\sigma_{C_0^+}$ is the induced measure on C_0^+.

Moreover, for $\phi \in C^{\infty}((0,\infty)\times\mathbb{R}^d)$, we have

$$\langle \mathbf{1}_{(0,\infty)}(t)\delta_0(t^2-|x|^2), \phi\rangle = \int_0^\infty \frac{1}{2t}\langle \mathrm{d}\sigma_{S_t}, \phi\rangle \,\mathrm{d}t, \tag{7.21}$$

where S_{t_0} is the sphere $\{(t,x) : t = t_0, |x| = t_0\}$ and $\mathrm{d}\sigma_{S_{t_0}}$ is the induced measure on S_{t_0}.

Proof. Let (r,ω) be the standard polar coordinates on $\mathbb{R}^3 \setminus \{0\}$, i.e.,

$$(r,\omega) = (|x|, \tfrac{x}{|x|}) \in (0,\infty)\times\mathbb{S}^2.$$

We employ the coordinate transform $\Phi(t,r,\omega) := (t, s(t,r), \omega)$ on $\mathbb{R}^{1+3}$, which is defined by

$$s = t^2 - |x|^2 = t^2 - r^2.$$

Note that $\Phi : (t,r,\omega) \mapsto (t,s,\omega)$ is a diffeomorphism in $\{t > 0, r > 0\}$ and the change of variables formula reads

$$\int f(t,r,\omega) r^{d-1}\,\mathrm{d}t\mathrm{d}r\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\omega) = \int \frac{f(t,\sqrt{t^2-s},\omega)}{2\sqrt{t^2-s}}\sqrt{t^2-s}^{\,d-1}\,\mathrm{d}t\mathrm{d}s\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\omega).$$

Then according to the definition of $\delta_0(t^2-|x|^2) = \delta_0 \circ s$ in Proposition 6.3, the following holds: for $\phi \in C^\infty_c(\{t>0, r>0\})$, we have

$$\begin{aligned}\langle \delta_0(t^2-|x|^2), \phi\rangle &= \left\langle \delta_0(s), \int \frac{\phi(t,\sqrt{t^2-s},\omega)}{2\sqrt{t^2-s}}\sqrt{t^2-s}^{\,d-1}\,\mathrm{d}t\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\omega)\right\rangle \\ &= \int \frac{\phi(t,t,\omega)}{2t} t^{d-1}\,\mathrm{d}t\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\omega).\end{aligned}$$

Note that the formula clearly extends to $\phi \in C^\infty_c(\{t>0\})$ since the distribution is supported away from $\{t>0, r=0\}$. In view of the fact that

$$\int f \mathrm{d}\sigma_{C_0^+} = 2\sqrt{2}\int f(t,t,\omega)t^{d-1}\mathrm{d}t\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\omega),$$

we have

$$\langle \delta_0(t^2-|x|^2), \phi\rangle = \left\langle \frac{1}{2\sqrt{2}t}\mathrm{d}\sigma_{C_0^+}, \phi\right\rangle,$$

which is exactly (7.20). Similarly, since

$$\int f\mathrm{d}\sigma_{S_t} = \int f(t,t,\omega)t^{d-1}\mathrm{d}\sigma_{\mathbb{S}^{d-1}}(\omega),$$

we have

$$\langle \delta_0(t^2-|x|^2), \phi\rangle = \int_0^\infty \left\langle \frac{1}{2t}\mathrm{d}\sigma_{S_t}\phi\right\rangle \mathrm{d}t,$$

which is precisely (7.21). □

7.3. **Uses of the forward fundamental solution.** Finally, we derive representation and solutions formulas for the wave equation using the forward fundamental solution E_+ that we just found. Unlike in Sections 4 and 5, E_+ is not smooth on $\mathbb{R}^{1+d} \setminus \{(0,0)\}$, so we need to be a bit more careful.

Recall that a forward fundamental solution E_+ satisfies the following properties:

$$\Box E_+ = \delta_0, \tag{7.22}$$

$$\mathrm{supp}\ E_+ \subseteq \{(t,x) \in \mathbb{R}^{1+d} : t \geq 0\}. \tag{7.23}$$

Moreover, by inspecting the form of E_+ derived in Sections 7.1 and 7.2, we may verify the following property: there exists a continuous family of (spatial) distributions $(0,\infty) \to \mathcal{D}'(\mathbb{R}^d)$, $t \mapsto E_+(t)$ such that

$$\langle E_+, \phi \rangle = \int_0^\infty \langle E_+(t), \phi(t,\cdot) \rangle \, \mathrm{d}t \tag{7.24}$$

for every $\phi \in C_c^\infty(\mathbb{R}^{1+d})$. Moreover,

$$\mathrm{supp}\, E_+(t) \subseteq B_t(0) \text{ for every } t > 0, \tag{7.25}$$

which is the *finite speed of propagation.*

For example, we have

$$\langle E_+(t), \psi \rangle = \begin{cases} -\frac{1}{2} \int_{|x|<t} \psi(x) \, \mathrm{d}x & d = 1, \\ -\frac{1}{2\pi} \int_{|x|<t} \frac{\psi(x)}{\sqrt{t^2 - |x|^2}} \, \mathrm{d}x & d = 2, \\ -\frac{1}{4\pi t} \int_{S_t} \psi \, \mathrm{d}\sigma_{S_t} & d = 3. \end{cases} \tag{7.26}$$

Indeed, these follow from the explicit computations discussed in the previous subsection. Using these formulae, Properties (7.24) and (7.25) may be easily verified for $d = 1, 2, 3$. For the general case, see [H03, Section 6.2]

We begin by observing the following consequences of the above properties.

Lemma 7.7. *The following statements hold.*

(1) Define $\partial_t^j E_+(t)$ by the following requirement: for every $\phi \in C_c^\infty((0,\infty) \times \mathbb{R}^d)$, we have

$$\int_0^\infty \langle \partial_t^j E_+(t), \varphi(t,\cdot) \rangle \, \mathrm{d}t = (-1)^j \int_0^\infty \langle E_+(t), \partial_t^j \varphi(t,\cdot) \rangle \, \mathrm{d}t.$$

Then $\partial_t E_+(t)$ and $\partial_t^2 E_+(t)$ are belong to the class $C_t((0,\infty); \mathcal{D}(\mathbb{R}^d))$ (i.e., $(0,\infty) \to \mathcal{D}'(\mathbb{R}^d)$, $t \mapsto E_+$ is C^2 on $(0,\infty)$).

(2) We have

$$\lim_{t \to 0+} E_+(t) = 0, \quad \lim_{t \to 0+} \partial_t E_+(t) = -\delta_0(x). \tag{7.27}$$

Proof. By (7.22), $\partial_t^2 E_+(t) = \Delta E_+(t)$ for every $t > 0$, and therefore $\partial_t^2 E_+(t) \in C_t((0,\infty); \mathcal{D}(\mathbb{R}^d))$. By integrating in t, it follows that $\partial_t E_+(t) \in C_t((0,\infty); \mathcal{D}(\mathbb{R}^d))$ as well. This proves Statement (1).

Next, we prove Statement (2). By (7.22), $\langle \Box E_+, \phi \rangle = \phi(0,0)$ for any $\phi \in C_c^\infty(\mathbb{R}^{1+d})$. Using the definition of E_+, we also have

$$\begin{aligned} \phi(0,0) = \langle E_+, \Box\phi \rangle &= \int_0^\infty \langle E_+(t), -\partial_t^2 + \Delta)\phi(t,\cdot) \rangle \, \mathrm{d}t \\ &= \int_0^\infty \langle (-\partial_t^2 + \Delta) E_+(t), \phi(t,\cdot) \rangle \, \mathrm{d}t \end{aligned}$$

$$+\lim_{t\to 0+}\langle E_+(t), \partial_t\phi(0,\cdot)\rangle - \lim_{t\to 0+}\langle \partial_t E_+(t), \phi(0,\cdot)\rangle$$
$$= \lim_{t\to 0+}\langle E_+(t), \partial_t\phi(0,\cdot)\rangle - \lim_{t\to 0+}\langle \partial_t E_+(t), \phi(0,\cdot)\rangle.$$

Since this identity holds for every $\phi \in C_c^\infty(\mathbb{R}^{1+d})$, (7.27) follows. □

Remark 7.8. Observe that (7.27) allows us to interpret E_+ as the solution to the homogeneous equation $\Box E_+ = 0$ in $(0,\infty)_t \times \mathbb{R}^d$ with Cauchy data $\lim_{t\to 0+} E_+(t) = 0$ and $\lim_{t\to 0+} \partial_t E_+(t) = -\delta_0$ similar to what we have seen in Sections 5.1 and 5.2.

The following lemma, which is a consequence of (7.23) and (7.25), will be our basic technical tool.

Lemma 7.9. *Let f be any distribution with* $\operatorname{supp} f \subseteq \{t \in [L,\infty)\}$ *for some* $L \in \mathbb{R}$. *Then the convolution $E_+ * f$ is well-defined.*

Moreover, if $f \in C^\infty(\overline{(0,\infty)\times\mathbb{R}^d})$ and $h \in C^\infty(\mathbb{R}^d)$, then

$$E_+ * f = \int_0^t E_+(t-s) *_{\mathbb{R}^d} f(s)\,\mathrm{d}s, \tag{7.28}$$

$$E_+ * (\delta_0(t) h) = E_+(t) *_{\mathbb{R}^d} h, \tag{7.29}$$

where $*_{\mathbb{R}^d}$ refers to the convolution on $\mathbb{R}^d$.

The idea behind Lemma 7.9 is simple to understand when f and E_+ are both assumed to be functions; then by the support property of f and $\operatorname{supp} E_+ \subseteq [0,\infty)$,

$$\begin{aligned} E_+ * f(t,x) \text{``}=\text{''} &\iint E_+(t-s, x-y) f(s,y)\,\mathrm{d}s\mathrm{d}y \\ &= \int_L^\infty \left(\int E_+(t-s,x-y) f(s,y)\,\mathrm{d}y\right)\mathrm{d}s \\ &= \iint \mathbf{1}_{(0,t-L)}(t-s) E_+(t-s,x-y) f(s,y)\,\mathrm{d}s\mathrm{d}y, \end{aligned}$$

where the last line is well-defined thanks to (7.25). Note also that (7.28) and (7.29) readily follow from (7.24) and the preceding identity.

Proof. Identities (7.28) and (7.29) are straightforward consequences of (7.24); we omit the proofs. Let us focus on the proof that $E_+ * f$ is well-defined.

Without loss of generality, let us set $L = 1$. For any interval $I = (a,b) \subseteq \mathbb{R}$ (where a, b could be $\pm\infty$), denote by χ_I a smooth function on $\mathbb{R}^{1+d}$ such that $\chi_I(t,x) = 1$ if $tinI$ and 0 if either $t \leq a-1$ or $t \geq b+1$. We will show that

$$\chi_I(f * E_+)$$

is well-defined for any bounded interval I; then by approximation, $f * E_+$ can then be defined.

We will use the adjoint method to define $\chi_I(f * E_+)$. Given $\varphi \in C_c^\infty(\mathbb{R}^{1+d})$, let us first formally compute:

$$\begin{aligned} \langle \chi_I(f * E_+), \varphi\rangle &\text{``}=\text{''} \langle f, E_+ *' \chi_I \varphi\rangle \\ &\text{``}=\text{''} \langle f, \mathbf{1}_{(0,\infty)}(E_+ *' \chi_I\varphi)\rangle, \end{aligned}$$

where on the second line we used the support property of f. Thus, our task is to show that

$$T'[\varphi] := \mathbf{1}_{(0,\infty)}(E_+ *' \chi_I \varphi)$$

is a test function (to be pedantic, we also need to show that $T' : C_c^\infty(\mathbb{R}^{1+d}) \to C_c^\infty(\mathbb{R}^{1+d})$ is continuous, but this property will be evident). Writing $I = (a, b)$, introduce the half-open interval $J = (b + 10, \infty)$. Since

$$\operatorname{supp}(g *' h) \subseteq -\operatorname{supp} g + \operatorname{supp} h,$$

we have

$$\mathbf{1}_{(0,\infty)}(\chi_J E_+ *' \chi_I \varphi) = 0.$$

Thus,

$$T'[\varphi] = \mathbf{1}_{(0,\infty)}((1-\chi_J)E_+ *' \chi_I\varphi) = \mathbf{1}_{(0,\infty)}((1-\chi_J)\chi_{(-1,\infty)}E_+ *' \chi_I\varphi)$$

where for the last equality, we used (7.23). By (7.25), $(1-\chi_J)\chi_{(-1,\infty)}E_+$ is compactly supported. Since $\chi_I\varphi \in C_c^\infty(\mathbb{R}^{1+d})$, it follows that its adjoint convolution with the compactly supported distribution $(1-\chi_J)\chi_{(-1,\infty)}E_+$ is smooth and compactly supported; thus $T'[\varphi] \in C_c^\infty(\mathbb{R}^{1+d})$, as desired. □

We are now ready to prove the main results of this subsection. We begin with the solution formula.

Theorem 7.10 (Solvability of the wave equation)**.** *Suppose that a forward fundamental solution E_+ with the properties* (7.22), (7.23) *exists. Given $g, h \in C^\infty(\mathbb{R}^d)$ and $f \in C^\infty(\mathbb{R}^{1+d})$, there exists a unique solution ϕ to the initial value problem* (7.1) *defined by*

$$\begin{aligned}\phi &:= -E_+ * (h\delta_0(t)) - \partial_t\left(E_+ * (g\delta_0(t))\right) + E_+ * (f\mathbf{1}_{(0,\infty)}(t)) \\ &= -E_+(t) *_{\mathbb{R}^d} h - \partial_t E_+(t) *_{\mathbb{R}^d} g + \int_0^t E_+(t-s) *_{\mathbb{R}^d} f(s)\,\mathrm{d}s.\end{aligned} \tag{7.30}$$

Proof. When $g = h = 0$, it is clear from (7.22) that $E_+ * (f\mathbf{1}_{(0,\infty)}(t))$ defines a forward solution to $\Box\phi = f\mathbf{1}_{(0,\infty)}(t)$. That $-E_+(t) *_{\mathbb{R}^d} h - \partial_t E_+(t) *_{\mathbb{R}^d} g$ (which equals $-E_+ * (h\delta_0(t)) - \partial_t\left(E_+ * (g\delta_0(t))\right)$) satisfies the homogeneous wave equation in $(0,\infty)\times\mathbb{R}^d$ with $(\phi, \partial_t\phi)(0) = (g, h)$ follows from the properties of $E_+(t)$. □

Next, let us derive a *representation formula.*

Theorem 7.11 (Representation formula)**.** *Given any $\phi \in C^\infty(\mathbb{R}^{1+d})$, we have the formula*

$$\begin{aligned}\phi &= -E_+ * \partial_t\phi|_{\{t=0\}}\delta_0(t) - \partial_t\left(E_+ * \phi|_{\{t=0\}}\delta_0(t)\right) + E_+ * (\Box\phi\mathbf{1}_{(0,\infty)}(t)) \\ &= -E_+(t) *_{\mathbb{R}^d} \partial_t\phi|_{\{t=0\}} - \partial_t E_+(t) *_{\mathbb{R}^d} \phi|_{\{t=0\}} + \int_0^t E_+(t-s) *_{\mathbb{R}^d} \Box\phi(s)\,\mathrm{d}s.\end{aligned} \tag{7.31}$$

Proof. The second identity is a simple consequence of (7.24), so let us focus on establishing the first identity. For $(t,x) \in (0,\infty)\times\mathbb{R}^d$, we compute

$$\begin{aligned}\phi(t,x) &= \delta_0 * \phi\mathbf{1}_{(0,\infty)}(t,x) \\ &= \Box E_+ * \phi\mathbf{1}_{(0,\infty)} \\ &= -\partial_t^2 E_+ * \phi\mathbf{1}_{(0,\infty)}(t) + E_+ * \Delta\phi\mathbf{1}_{(0,\infty)} \\ &= -\partial_t^2 E_+ * \phi\mathbf{1}_{(0,\infty)} + E_+ * (\partial_t^2\phi)\mathbf{1}_{(0,\infty)} + E_+ * \Box\phi\mathbf{1}_{(0,\infty)},\end{aligned}$$

which are justified thanks to Lemma 7.9. We then compute

$$-\partial_t^2 E_+ * \phi\mathbf{1}_{(0,\infty)} + E_+ * (\partial_t^2\phi)\mathbf{1}_{(0,\infty)}$$

$$
\begin{aligned}
&= -\partial_t E_+ * (\partial_t \phi)\mathbf{1}_{(0,\infty)} - \partial_t E_+ * \phi\delta_{t=0} + E_+ * (\partial_t^2 \phi)\mathbf{1}_{(0,\infty)} \\
&= -E_+ * (\partial_t^2 \phi)\mathbf{1}_{(0,\infty)} - E_+ * (\partial_t \phi)\delta_{t=0} - \partial_t E_+ * \phi\delta_{t=0} + E_+ * (\partial_t^2 \phi)\mathbf{1}_{(0,\infty)} \\
&= -E_+ * (\partial_t \phi)\delta_{t=0} - \partial_t E_+ * \phi\delta_{t=0} \\
&= -E_+ * (\partial_t \phi)\delta_{t=0} - \partial_t \left(E_+ * \phi\delta_{t=0} \right),
\end{aligned}
$$

where all computations are justified again thanks to Lemma 7.9. Finally, for $\phi \in C^\infty(\mathbb{R}^{1+d})$, note that

$$\phi(t,x)\delta_{t=0} = \phi|_{\{t=0\}}(x)\delta_{t=0}, \quad \partial_t\phi(t,x)\delta_{t=0} = \partial_t\phi|_{\{t=0\}}(x)\delta_{t=0}. \qquad \square$$

Let us note that the regularity hypothesis for ϕ in Theorem 7.11 can be weakened considerably, although we will not pursue the details. Moreover, analogous statements can be proved in the negative time direction, simply by reversing the time coordinate $t \mapsto -t$.

We also note that, by a simple variant of the proof of the representation theorem, we have the uniqueness of the forward fundamental solution E_+:

Proposition 7.12. *Suppose that a forward fundamental solution E_+ with the properties* (7.22), (7.23) *exists. Then it is the unique forward fundamental solution, i.e., any fundamental solution E with* $\operatorname{supp} E \subseteq \{t \geq 0\}$ *equals E_+.*

Proof. Let E be a forward fundamental solution, i.e., $\Box E = \delta_0$ and $\operatorname{supp} E \subseteq \{t \in [0,\infty)\}$. By Lemma 7.9, the convolution $E * E_+$ is well-defined, so that we have

$$E_+ = \delta_0 * E_+ = (\Box E) * E_+ (= \Box(E * E_+)) = E * \Box E_+ = E * \delta_0 = E. \qquad \square$$

We end this subsection by considering some (important) examples.

d'Alembert's formula for $d = 1$. Recall that $E_+(t,x) = \frac{1}{2}H(t-x)H(t+x)$. We compute

$$
\begin{aligned}
-E_+ * h\delta_0(t) =& \frac{1}{2}\langle H(t-s-(x-y))H(t-s+x-y), h(y)\delta_0(s)\rangle_{y,s} \\
=& \frac{1}{2}\langle H(t-(x-y))H(t+x-y), h(y)\rangle_y \\
=& \frac{1}{2}\int_{x-t}^{x+t} h(y)\,\mathrm{d}y,
\end{aligned}
$$

$$
\begin{aligned}
-\partial_t(E_+ * g\delta_0(t)) =& \partial_t\Big(\frac{1}{2}\int_{x-t}^{x+t} g(y)\,\mathrm{d}y\Big) \\
=& \frac{1}{2}(g(x+t) + g(x-t)),
\end{aligned}
$$

and

$$
\begin{aligned}
E_+ * f\mathbf{1}_{(0,\infty)} =& -\frac{1}{2}\langle H(t-s-(x-y))H(t-s+x-y), f(s,y)H(s)\rangle_{y,s} \\
=& -\frac{1}{2}\int_0^t \int_{x-t+s}^{x+t-s} f(s,y)\,\mathrm{d}y\,\mathrm{d}s.
\end{aligned}
$$

Hence we obtain *d'Alembert's formula* in $\mathbb{R}^{1+1}$ (Theorem 7.1), which is both the solution and representation formula in this case.

Explicit computation for $d = 2$. Recall that

$$E_+(t,x) = -\frac{1}{2\pi}\mathbf{1}_{\{(t,x):0\leq|x|\leq t\}}\frac{1}{(t^2-|x|^2)^{\frac{1}{2}}},$$

outside the origin, and at the origin E_+ is determined by homogeneity. We may easily compute

$$\begin{aligned}
-E_+ * (h\delta_{t=0})(t,x) =& -\langle E_+(t-s, x-y), h(y)\delta_0(s)\rangle_{y,s} \\
=& \frac{1}{2\pi}\int_{\{|x|\leq t\}} \frac{h(y)}{(t^2-|x-y|^2)^{\frac{1}{2}}}\,\mathrm{d}y \\
E_+ * (f\,\mathbf{1}_{(0,\infty)})(t,x) =& \langle E_+(t-s,x-y), f(s,y)\mathbf{1}_{(0,\infty)}(s,y)\rangle_{y,s} \\
=& -\frac{1}{2\pi}\int_0^t\int_{\{|x|\leq s\}}\frac{f(s,y)}{((t-s)^2-|x-y|^2)^{\frac{1}{2}}}\,\mathrm{d}y\,\mathrm{d}s.
\end{aligned}$$

Combined with Theorems 7.11 and 7.10, we recover *Poisson's formula*:

Theorem 7.13 (Poisson's formula)**.** *Let ϕ be a solution to the equation $\Box\phi = F$ with $\phi, F \in C^\infty(\mathbb{R}^{1+2})$. Then we have the formula*

(7.32)
$$\begin{aligned}
\phi(t,x) = \partial_t\Big(\frac{1}{2\pi}\int_{\{|x|\leq t\}}\frac{g(y)}{(t^2-|x-y|^2)^{\frac{1}{2}}}\,\mathrm{d}y\Big) + \frac{1}{2\pi}\int_{\{|x|\leq t\}}\frac{h(y)}{(t^2-|x-y|^2)^{\frac{1}{2}}}\,\mathrm{d}y \\
-\frac{1}{2\pi}\int_0^t\int_{\{|x|\leq s\}}\frac{f(s,y)}{((t-s)^2-|x-y|^2)^{\frac{1}{2}}}\,\mathrm{d}y\,\mathrm{d}s.
\end{aligned}$$

where $(\phi_0,\phi_1) = (\phi, \partial_t\phi)|_{\{t=0\}}$ and $B_{0,t}(x)$ is the ball $\{(0,y) : |x| \leq t\}$.

Conversely, given any initial data $(\phi_1,\phi_2) \in C^\infty(\mathbb{R}^2)$ and $F \in C^\infty(\mathbb{R}^{1+2})$, there exists a unique solution ϕ to the initial value problem (7.1) *defined by the formula* (7.32)*.*

Explicit computation for $d = 3$. Recall that

$$E_+(t,x) = -\frac{1}{2\pi}\mathbf{1}_{(0,\infty)}\,\delta_0(t^2-|x|^2),$$

outside the origin, and at the origin E_+ is determined by homogeneity. Using Theorems 7.11, 7.10 and Lemma 7.6, now it is not difficult to prove *Kirchhoff's formula*:

Theorem 7.14 (Kirchhoff's formula)**.** *Let ϕ be a solution to the equation $\Box\phi = F$ with $\phi, F \in C^\infty(\mathbb{R}^{1+3})$. Then we have the formula*

$$\begin{aligned}
\phi(t,x) =& \partial_t\Big(\frac{1}{4\pi t}\int_{S_{0,t}(x)} g(y)\,\mathrm{d}\sigma(y)\Big) + \frac{1}{4\pi t}\int_{S_{0,t}(x)} h(y)\,\mathrm{d}\sigma(y) \\
&-\int_0^t \frac{1}{4\pi(t-s)}\int_{S_{0,t-s}(x)} f(s,y)\,\mathrm{d}\sigma(y)
\end{aligned} \tag{7.33}$$

where $(\phi_0,\phi_1) = (\phi,\partial_t\phi)|_{\{t=0\}}$ and $S_{0,t}(x)$ is the sphere $\{(0,y) : |y-x| = t\}$.

Conversely, given any initial data $(\phi_1,\phi_2) \in C^\infty(\mathbb{R}^3)$ and $F \in C^\infty(\mathbb{R}^{1+3})$, there exists a unique solution ϕ to the initial value problem (7.1) *defined by the formula* (7.33)*.*

Remark 7.15. For an alternative approach to derivation of the classical representation formulae, which does not use the theory of distributions, we refer the reader to [Eva10, Chapter 2].

7.4. **Computation of precise constant for E_+ (Optional).** Here we give a precise computation of the constant in the forward fundamental solution for the d'Alembertian. We recall the formula here for the convenience of the reader:

$$E_+ = c_d \mathbf{1}_{(0,\infty)} \chi_+^{-\frac{d-1}{2}}(t^2 - |x|^2), \qquad \text{where } c_d = -\frac{1}{2\pi^{\frac{d-1}{2}}}. \tag{7.34}$$

This formula can be read off from [Hö3, Theorem 6.2.1], which in fact applies to more general constant coefficient second order differential operators. We present another argument[16] here, which is based on the use of the null coordinates (u, v, ω).

We need to recall the following well-known functional equations for the Gamma function $\Gamma(a)$:

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 s^{a-1}(1-s)^{b-1} \, \mathrm{d}s. \tag{7.35}$$

$$\Gamma(a)\Gamma(a+\tfrac{1}{2}) = 2^{1-2a}\sqrt{\pi}\Gamma(2a). \tag{7.36}$$

The function defined by the RHS of (7.35) is called the *Beta function* $B(a, b)$; it can be easily proved by writing out $\Gamma(a)\Gamma(b) = \int_0^\infty \int_0^\infty e^{-(s+t)} s^{a-1} t^{b-1} \, \mathrm{d}s\mathrm{d}t$ and making the change of variables $s = uv$, $t = u(1 - v)$. Equation (7.36), called *Legendre's duplication formula*, can be derived by using (7.35) twice, with an appropriate change of variables (**Exercise:** Prove these formulae!).

We also record the following formula concerning the homogeneous distribution χ_+^a:

$$\chi_+^a * \chi_+^b = \chi_+^{a+b+1} \tag{7.37}$$

This identity is in fact equivalent to (7.35).

Proof of (7.34). First, given $g, h \in C_c^\infty$, note that we have the simple formula (by integration by parts)

$$\langle \Box g, h \rangle = -\int \partial_t g \partial_t h \, \mathrm{d}t\mathrm{d}x + \int \nabla_x g \cdot \nabla_x h \, \mathrm{d}t\mathrm{d}x$$

Suppose that g, h is rotationally invariant. Then in the polar coordinates (t, r, ω), we see that

$$\langle \Box g, h \rangle = -d\alpha(d) \iint (\partial_t g \partial_t h - \int \partial_r g \cdot \partial_r h) r^{d-1} \, \mathrm{d}t\mathrm{d}r$$

where $d\alpha(d) = \int_{\mathbb{S}^{d-1}} \mathrm{d}\sigma$ is the $d-1$-dimensional volume of the unit sphere $\mathbb{S}^{d-1}$. Making another change of variables to the null coordinates[17] $(u, v, \omega) = (t - r, t + r, \omega)$, we then have the formula

$$\langle \Box g, h \rangle = -d\alpha(d) \iint (\partial_v f \partial_u g + \partial_u f \partial_v g) \Big(\frac{v-u}{2}\Big)_+^{d-1} \mathrm{d}u\mathrm{d}v. \tag{7.38}$$

Now recall that E_+ is a function of $t^2 - |x|^2$, which equals uv in the null coordinates. Using $g = \Box E_+ = \delta_0$ and $h = H(v) = 1_{\{t+|x|\leq 1\}}$, the identity (7.38) can then be used to deduce

$$1 = \langle \Box E_+, 1_{\{|x|+t\leq 1\}} \rangle = d\alpha(d) \iint \partial_u E_+(uv) \partial_v 1_{\{v\leq 1\}} \Big(\frac{v-u}{2}\Big)_+^{d-1} \mathrm{d}u\mathrm{d}v \tag{7.39}$$

[16] This proof is due to P. Isett.

[17] Note that the normalization is slightly different from Section 7.1, but the idea is the same.

where the integral is interpreted suitably. (**Exercise:** Using the support properties of E_+ and $1_{\{v\leq 1\}}$, show that (7.39) makes sense. Indeed, show that the right-hand side is the limit

$$\omega_d \iint 1_{v+u\geq 0}\partial_u g_j(uv)\partial_v h_j(1-v)(\frac{v-u}{2})_+^{d-1}\,\mathrm{d}u\mathrm{d}v \qquad \text{as } j\to\infty,$$

where $g_j, h_j \in C_c^\infty(\mathbb{R})$, $g_j(x) \rightharpoonup c_d\chi_+^{-\frac{d-1}{2}}(x)$ and $h_j(x) \rightharpoonup \mathbf{1}_{\{x>0\}}$, both in the sense of distributions.)

Now note that

$$\begin{aligned}
\partial_u E_+ =& -\mathbf{1}_{\{v+u>0\}} c_d v \chi_+^{-\frac{d+1}{2}}(uv),\\
\left(\frac{v-u}{2}\right)_+^{d-1} =& 2^{-d+1}(d-1)!\chi_+^{d-1}(v-u),\\
\partial_v \mathbf{1}_{v<1} =& -\delta(1-v).
\end{aligned}$$

Substituting these identities into (7.39), it follows that

$$\begin{aligned}
c_d^{-1} =& -2^{-d+1}(d-1)!\, d\alpha(d) \iint \chi_+^{-\frac{d+1}{2}}(u)\chi_+^{d-1}(v-u)\,\mathrm{d}u\\
=& -2^{-d+1}(d-1)!\, d\alpha(d)\chi_+^{-\frac{d+1}{2}} * \chi_+^{d-1}(1).
\end{aligned}$$

Using the identities

$$\begin{aligned}
\chi_+^a * \chi_+^b =& \chi_+^{a+b+1}\\
\chi_+^a(1) =& \frac{1}{\Gamma(a+1)}
\end{aligned}$$

and the formula $d\alpha(d) = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}$, we see that

$$c_d^{-1} = -2^{-d+1}(d-1)!\,\frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}\,\frac{1}{\Gamma(\frac{d+1}{2})}. \tag{7.40}$$

By Legendre's duplication formula (7.36) with $a = \frac{d}{2}$, we have

$$\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{d+1}{2}\right) = 2^{-d+1}\sqrt{\pi}(d-1)!$$

Substituting the preceding computation into (7.40), we obtain (7.34). $\square$

8. The Fourier transform

This section is a quick introduction to the *Fourier transform*, which is a fundamental tool not only in the study of PDEs, but also in many other fields in Mathematics, Science and Engineering.

8.1. **Motivation.** The Fourier transform is, essentially, a "change-of-basis" transformation in the "space of functions" on $\mathbb{R}^d$, in the following sense. When we express a function f in terms of its pointwise values $\{f(y)\}_{y\in\mathbb{R}^d}$, we can think as if we are using $\{\delta_0(x-y)\}_{y\in\mathbb{R}^d}$ as the "basis" (of course, we are making a formal discussion here, not caring about the fact that $\delta_0(x-y)$ themselves are *not* functions). Indeed, if f is a continuous function on $\mathbb{R}^d$, then $f(x) = \int f(y)\delta_0(x-y)\,\mathrm{d}y$ in the sense of distributions (i.e., f lies in the "span" of $\{\delta(x-y)\}_{y\in\mathbb{R}^d}$) and the coefficients $\{f(y)\}_{y\in\mathbb{R}^d}$ uniquely determine f (i.e., $\{\delta(x-y)\}_{y\in\mathbb{R}^d}$ is "linearly independent"); this point of view underlied the ideas behind the fundamental solutions (Section 3.11). The "basis" $\{\delta(x-y)\}_{y\in\mathbb{R}^d}$ is nice in that it simultaneously diagonalizes multiplication by any (nice enough) functions, i.e., for $m \in C^\infty(\mathbb{R}^d)$,

$$m\delta_0(x-y) = m(y)\delta_0(x-y) \quad \text{for every } y \in \mathbb{R}^d,$$

so operations such as multiplication by a smooth function is easy to understand with this "basis." However, differentiation, which is a central operation in the study of differential equations for obvious reasons, is more difficult to understand.

It turns out that another "basis" is more suitable to understand the operation of differentiation, namely, $\{e^{i\xi\cdot x}\}_{\xi\in\mathbb{R}^d}$. An important property of these objects is:

$$\partial_j e^{i\xi\cdot x} = i\xi_j e^{i\xi\cdot x},$$

so if $\{e^{i\xi\cdot x}\}_{\xi\in\mathbb{R}^d}$ were really a "basis" in a similar sense in which $\{\delta_0(x-y)\}_{y\in\mathbb{R}^d}$ is a "basis" (i.e., any function can be written in the form $\int a(\xi)e^{i\xi\cdot x}\,\mathrm{d}\xi$ in a unique way), then all constant coefficient partial differential operators would be simultaneously diagonalized in $\{e^{i\xi\cdot x}\}_{\xi\in\mathbb{R}^d}$. We are led to the question: Given a function f on $\mathbb{R}^d$, can we write f uniquely in the form

$$f(x) \left(= \int f(y)\delta_0(x-y)\,\mathrm{d}y \right) = \int a(\xi)e^{i\xi\cdot x}\,\mathrm{d}\xi? \tag{8.1}$$

Remarkably, the theory of Fourier transform tells us that the answer is *yes.* The *Fourier transform* $\mathcal{F}$ is precisely the "change-of-basis" formula that links $\{f(y)\}_{y\in\mathbb{R}^d}$ with $\{a(\xi)\}_{\xi\in\mathbb{R}^d}$, i.e.,

$$a(\xi) \propto \mathcal{F}[f](\xi),$$

where $f \propto g$ for functions f, g means that $f = cg$ for some non-zero constant $c \in \mathbb{R}$.

Let us continue this heuristic discussion to derive the form of $\mathcal{F}$. Experience from linear algebra tells us that the "change-of-basis" formula should be

$$\mathcal{F}[f](\xi) \propto \int f(y)m(y,\xi)\,\mathrm{d}y$$

where the "matrix" $m(y,\xi)$ is characterized by

$$\delta_0(x-y) \text{ ``}={\text{''}} \int m(y,\xi)e^{i\xi\cdot x}\,\mathrm{d}\xi.$$

Let us derive some formal properties of $m(y,\xi)$. By the translation symmetry, we can easily see that

$$\int m(y,\xi)e^{i\xi\cdot x}\,\mathrm{d}\xi \text{``}=\text{''} \delta_0(x-y) \text{``}=\text{''} \int m(0,\xi)e^{i\xi\cdot(x-y)}\,\mathrm{d}\xi = \int m(0,\xi)e^{-i\xi\cdot y}e^{i\xi\cdot x}\,\mathrm{d}\xi,$$

so if we believe that $\{e^{i\xi\cdot x}\}_{\xi\in\mathbb{R}^d}$ forms a basis, we should have

$$m(y,\xi) = m(0,\xi)e^{-i\xi\cdot y}.$$

In particular, it suffices to consider the case $y=0$. Note that $\delta_0(x)$ has the property that $x^j\delta_0(x)=0$; on the other hand, we have

$$0 = x^j\delta_0 \text{``}=\text{''} \int m(0,\xi)x^je^{i\xi\cdot x}\,\mathrm{d}\xi = \int m(0,\xi)\frac{1}{i}\partial_{\xi_j}e^{i\xi\cdot x}\,\mathrm{d}\xi \text{``}=\text{''} i\int \partial_\xi m(0,\xi)e^{i\xi\cdot x}\,\mathrm{d}\xi.$$

If we believe that $\{e^{i\xi\cdot x}\}_{\xi\in\mathbb{R}^d}$ forms a "basis", then $\partial_{\xi_j}m(0,\xi)$ should be zero. Thus $m(0,\xi)$ must be a non-zero constant, i.e.,

$$\delta_0(x) \text{``}=\text{''} c\int e^{i\xi\cdot x}\,\mathrm{d}\xi.$$

Note that, amusingly, at this point we already deduced that $m(y,\xi)\propto e^{-i\xi\cdot y}$, so we are led to

$$\mathcal{F}[f](\xi) \propto \int f(y)e^{-i\xi\cdot y}\,\mathrm{d}y,$$

which is the correct form of the Fourier transform (as some of you may have already learned)!

Let us finally nail down the constant c. We claim that $c=\frac{1}{(2\pi)^d}$, i.e.,

$$\delta_0(x) \text{``}=\text{''} \frac{1}{(2\pi)^d}\int e^{i\xi\cdot x}\,\mathrm{d}\xi. \tag{8.2}$$

An informal derivation is as follows. In view of the decompositions $\delta_0(x) = \delta_0(x^1)\cdots\delta_0(x^d)$ and $\int e^{i\xi\cdot x}\,\mathrm{d}\xi = \int e^{i\xi_1x^1}\,\mathrm{d}\xi_1\cdots\int e^{i\xi_1x^1}\,\mathrm{d}\xi_1$, where the delta distributions on the RHS are on $\mathbb{R}$, we see that $c=c_1^d$, where c_1 is the constant in dimension $d=1$:

$$\delta_0(x) \text{``}=\text{''} c_1\int_{-\infty}^{\infty} e^{i\xi x}\,\mathrm{d}\xi \text{ on } \mathbb{R}.$$

We present two approaches for determining c_1 (which is $\frac{1}{2\pi}$), one using an approximation procedure to make sense of $\int_{-\infty}^{\infty} e^{i\xi x}\,\mathrm{d}\xi$, and another using the formal algebraic properties of $e^{i\xi x}$ and the Gaussian.

- *An approach using approximation.* To make sense of the integral $\int_{-\infty}^{\infty} e^{i\xi x}\,\mathrm{d}\xi$, we "temper" the integrand by multiplying by $e^{-\epsilon|\xi|}$, integrate in ξ and take the limit $\epsilon\to 0+$ (approximation method). This limit may be rewritten as

$$\begin{aligned}\lim_{\epsilon\to 0+}\int_{-\infty}^{\infty} e^{i\xi x-\epsilon|\xi|}\,\mathrm{d}\xi &= \lim_{\epsilon\to 0+}\left(\int_{-\infty}^{0} e^{i\xi(x-i\epsilon)}\,\mathrm{d}\xi + \int_0^{\infty} e^{i\xi(x+i\epsilon)}\,\mathrm{d}\xi\right)\\ &= \lim_{\epsilon\to 0+}\left(\frac{1}{i(x-i\epsilon)} - \frac{1}{i(x+i\epsilon)}\right).\end{aligned}$$

  Note that

$$\frac{1}{x\mp i\epsilon} = \frac{x}{x^2+\epsilon^2} \pm \frac{i\epsilon}{x^2+\epsilon^2},$$

so

$$\lim_{\epsilon\to 0+}\left(\frac{1}{i(x-i\epsilon)}-\frac{1}{i(x+i\epsilon)}\right)=\lim_{\epsilon\to 0+}\frac{2\epsilon}{x^2+\epsilon^2}=2\left(\int\frac{\mathrm{d}t}{1+t^2}\right)\delta_0=2\pi\delta_0,$$

which implies that $c_1=(2\pi)^{-1}$.

- *An approach using the Gaussian.* For a "nice" function ϕ on $\mathbb{R}$, we must have

$$\phi(0)=\int\phi(x)\overline{\delta_0(x)}\,\mathrm{d}x \text{``}=\text{''} c_1\iint\phi(x)e^{-i\xi x}\,\mathrm{d}x\,\mathrm{d}\xi. \tag{8.3}$$

Let us try to find a ϕ for which

$$\hat{\phi}(\xi):=\int\phi(x)e^{-i\xi x}\,\mathrm{d}x,$$

which will be the Fourier transform, can be computed. The idea is to exploit the properties $xe^{i\xi x}=i\partial_\xi e^{-i\xi x}$ and $\partial_x e^{-i\xi x}=-i\xi e^{i\xi x}$, which implies that $\widehat{\partial_x\phi}(\xi)=i\xi\widehat{\phi}(\xi)$ and $\widehat{x\phi}(\xi)=i\partial_\xi\widehat{\phi}(\xi)$. Thus,

$$(\partial_x+x)\phi=0\Leftrightarrow(\partial_\xi+\xi)\widehat{\phi}=0.$$

By separation of variables, a general solution of the ODE $(\partial_x+x)\phi=0$ is the Gaussian $de^{-\frac{1}{2}x^2}$, where $d\in\mathbb{C}$ is any constant. Therefore, $\widehat{e^{-\frac{1}{2}x^2}}=de^{-\frac{1}{2}\xi^2}$. To evaluate the constant d, we take $\xi=0$, which implies

$$d=\int e^{-\frac{1}{2}x^2}\,\mathrm{d}x=\sqrt{2\pi},$$

so

$$\widehat{e^{-\frac{1}{2}x^2}}=\sqrt{2\pi}e^{-\frac{1}{2}\xi^2}. \tag{8.4}$$

Now plugging in $\phi(x)=e^{-\frac{1}{2}x^2}$ in (8.3), we see that

$$1=c_1\int\sqrt{2\pi}e^{-\frac{1}{2}\xi^2}\,\mathrm{d}\xi=2\pi c_1.$$

8.2. **The Fourier transform.** Our goal now is to make the heuristic discussion in Section 8.1 precise.

In the remainder of this section, we will be working with *complex-valued* functions and distributions, for the obvious reason that the elements in $\{e^{ix\cdot\xi}\}_{\xi\in\mathbb{R}^d}$ are complex-valued. Given two complex-valued functions f,g on a domain U in $\mathbb{R}^d$, we define their Hermitian L^2-pairing $\langle f,g\rangle$ by

$$\langle f,g\rangle:=\int_U f(x)\overline{g}(x)\,\mathrm{d}x.$$

Note that $\langle f,g\rangle$ is ($\mathbb{C}$-)linear in f, but *conjugate*-linear in g; moreover, $\langle f,g\rangle=\overline{\langle g,f\rangle}$. The set of smooth compactly supported complex-valued functions is written as $C^\infty_c(U;\mathbb{C})$; the topology on $C^\infty_c(U;\mathbb{C})$ is given by declaring that $f_j\to f$ in $C^\infty_c(U;\mathbb{C})$ if and only if $\operatorname{Re}f_j\to\operatorname{Re}f$ and $\operatorname{Im}f_j\to\operatorname{Im}f$ in $C^\infty_c(U)$. In keeping with this convention $\langle f,g\rangle=\int f\overline{g}$, we define the *complex-valued distributions* to be the continuous *conjugate*-linear functions on $C^\infty_c(U;\mathbb{C})$, i.e.,

$$a\langle f,\phi\rangle=\langle af,\phi\rangle=\langle f,\overline{a}\phi\rangle,\quad\langle f,\phi+\psi\rangle=\langle f,\phi\rangle+\langle f,\psi\rangle.$$

We write $\mathcal{D}'(U;\mathbb{C})$ for the complex-valued distributions on U. It is not difficult to see that this characterization of a complex-valued distribution f is equivalent to saying that f is of the form $u+iv$, where $u,v\in\mathcal{D}'(U)$. Thus, our entire discussion about real-valued distributions carries over to the complex-valued case without much change.

Motivated by the discussion in Section 8.1, we make the following definition of the Fourier transform:

Definition 8.1. For a complex-valued function $f \in C_c^\infty(\mathbb{R}^d; \mathbb{C})$, the *Fourier transform of f* is

$$\mathcal{F}[f](\xi) = \hat{f}(\xi) := \int_{\mathbb{R}^d} f(y) e^{-i\xi\cdot y} \,\mathrm{d}y. \tag{8.5}$$

We will equip the space $\mathbb{R}^d_\xi$ with the measure $\frac{\mathrm{d}\xi}{(2\pi)^d}$ (we will see the reason why in a moment) and define the Hermitian L^2-pairing for two complex-valued functions a, b by

$$\langle a, b\rangle_{(2\pi)^{-1}\mathrm{d}\xi} := \int_{\mathbb{R}^d} a\bar{b} \frac{\mathrm{d}\xi}{(2\pi)^d}.$$

Let us compute the formal Hermitian adjoint of $\mathcal{F}$. For $f, a \in C_c^\infty(\mathbb{R}^d)$,

$$\begin{aligned}
\langle \mathcal{F}f, a\rangle_{(2\pi)^{-d}\mathrm{d}\xi} &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} e^{-i\xi\cdot x} f(x)\overline{a(\xi)}\,\mathrm{d}x \frac{\mathrm{d}\xi}{(2\pi)^d} \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x)\overline{e^{i\xi\cdot x} a}(\xi) \frac{\mathrm{d}\xi}{(2\pi)^d}\,\mathrm{d}x \\
&= \langle f, \mathcal{F}^* a\rangle,
\end{aligned}$$

where

$$\mathcal{F}^* a(x) = \check{a}(x) := \int_{\mathbb{R}^d} a(\xi) e^{i\xi\cdot x} \frac{\mathrm{d}\xi}{(2\pi)^d}. \tag{8.6}$$

Observe that, according to the heuristic discussion in Section 8.1 (see, in particular, (8.1)), $\mathcal{F}^*$ *must be the inverse* of $\mathcal{F}$; we will prove this statement soon. The factor $(2\pi)^d$ in the measure, of course, is from (8.2).

Remark 8.2. Determining where to put the factor 2π is a well-known nuisance in dealing with the Fourier transform. Our choice (putting 2π in the measure $\mathrm{d}\xi$) is the oft-used one in PDEs, because we get to keep the simple identity $\partial_x e^{ix\xi} = i\xi e^{ix\xi}$ for turning differentiation into multiplication. Another popular choice, often used by harmonic analysts, is to put 2π in the basis and work with $\{e^{i2\pi\xi x}\}$.

Note that

$$\mathcal{F}^*[a](x) = (2\pi)^{-d}\mathcal{F}[a](-x).$$

so statements that we prove for $\mathcal{F}$ usually applies (after minor modifications) to $\mathcal{F}^*$ as well.

Note that $\mathcal{F}[f]$ and $\mathcal{F}^*[a]$ are well-defined for $f, a \in L^1(\mathbb{R}^d)$. Moreover, we have the following simple but important lemma:

Lemma 8.3. *Let $f \in L^1(\mathbb{R}^d)$.*

(1) Then $\mathcal{F}[f]$ is well-defined by the formula $\mathcal{F}[f] = \int f(y) e^{-i\xi\cdot y}\,\mathrm{d}y$. Moreover,

$$\|\mathcal{F}[f]\|_{L^\infty} \leq \|f\|_{L^1}.$$

(2) For any $x, \eta \in \mathbb{R}^d$,

$$\mathcal{F}[f(\cdot - x)](\xi) = e^{-ix\cdot\xi}\mathcal{F}[f](\xi), \quad \mathcal{F}[f(\cdot)](\xi - \eta) = \mathcal{F}[e^{i\eta(\cdot)}f](\xi).$$

(3) If both f and $\partial_j f$ lie in $L^1(\mathbb{R}^d)$, then

$$\mathcal{F}\left[\partial_j f\right](\xi) = i\xi_j \mathcal{F}[f](\xi).$$

(4) If both f and $x^j f$ lie in $L^1(\mathbb{R}^d)$, then $\mathcal{F}[f]$ is continuously differentiable in ξ_j and

$$\mathcal{F}\left[x^j f\right](\xi) = i\partial_{\xi_j} \mathcal{F}[f](\xi).$$

Thus, we arrive an important maxim regarding the Fourier transform:

Regularity of f corresponds to decay of $\hat{f}$, and vice versa.

Next, we turn to the goal of deriving the *Fourier inversion formula*, i.e., to understand $\mathcal{F}^{-1}$. For this purpose, we would like to work with a space of function that is closed under the Fourier transform (i.e., if f belongs to the space, so does $\mathcal{F}f$). The space $C^\infty_c(\mathbb{R}^d;\mathbb{C})$ in Definition 8.1 is not adequate for this purpose; it can be shown that $C^\infty_c(\mathbb{R}^d;\mathbb{C})$ is *not* $C^\infty_c(\mathbb{R}^d;\mathbb{C})$ itself. (**Exercise:** Show that if f and $\mathcal{F}[f]$ are both compactly supported, then f must be zero. [Hint: First show that f extends to a complex-analytic function on $\mathbb{C}^d$.]) On the other hand, Lemma 8.3 motivates us to consider the following class of functions:

Definition 8.4 (Schwartz class)**.**

$$\mathcal{S}(\mathbb{R}^d;\mathbb{C}) = \{\phi \in C^\infty(\mathbb{R}^d;\mathbb{C}) : \sup_{x\in\mathbb{R}^d} |x^\alpha D^\beta \phi(x)| < \infty \text{ for every multi-index } \alpha, \beta.\}$$

A sequence ϕ_j in $\mathcal{S}(\mathbb{R}^d;\mathbb{C})$ converges to $\phi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$ if and only if

$$|x^\alpha D^\beta(\phi_j - \phi)(x)| \to 0 \quad \text{as } j \to \infty \text{ for every multi-index } \alpha, \beta.$$

Remark 8.5 (For those who are familiar with functional analysis)*.* We note that $\mathcal{S}(\mathbb{R}^d;\mathbb{C})$ is a Frechét space defined with the semi-norms

$$p_{\alpha,\beta}(\phi) = \sup_{x\in\mathbb{R}^d} |x^\alpha D^\beta \phi(x)|.$$

By Lemma 8.3, it follows that

$$\mathcal{F} : \mathcal{S}(\mathbb{R}^d;\mathbb{C}) \to \mathcal{S}(\mathbb{R}^d;\mathbb{C}).$$

Since $\mathcal{F}^*\phi(x) = (2\pi)^{-d}\mathcal{F}\phi(-x)$, Lemma 8.3 implies that

$$\mathcal{F}^* : \mathcal{S}(\mathbb{R}^d;\mathbb{C}) \to \mathcal{S}(\mathbb{R}^d;\mathbb{C}),$$

as well. Thus, $\mathcal{F}$ can be extended to the dual space $\mathcal{S}'(\mathbb{R}^d;\mathbb{C})$ by the adjoint method; the elements in $\mathcal{S}'(\mathbb{R}^d;\mathbb{C})$ are what are called *tempered distributions*. More precisely,

Definition 8.6 (Tempered distributions)**.**

$$\mathcal{S}'(\mathbb{R}^d;\mathbb{C}) = \{u : u \text{ is a continuous conjugate-linear functional on } \mathcal{S}(\mathbb{R}^d;\mathbb{C})\}.$$

By continuity, we mean

$$\langle u, \phi_j\rangle \to \langle u, \phi\rangle \text{ as } j \to \infty \text{ whenever } \phi_j \to \phi \text{ in } \mathcal{S}(\mathbb{R}^d;\mathbb{C}) \text{ as } j \to \infty.$$

Given $u \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C}$ and $\phi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$, we also introduce the notation

$$\langle u, \phi\rangle_{(2\pi)^{-d}\mathrm{d}\xi} := (2\pi)^{-d} u(\phi),$$

which coincides with $\int u\overline{\phi}\frac{\mathrm{d}\xi}{(2\pi)^d}$ when $u, \phi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$. The extension of the Fourier transform $\mathcal{F}$ to a map $\mathcal{S}'(\mathbb{R}^d;\mathbb{C}) \to \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$ by the adjoint method is defined as follows:

$$\langle \mathcal{F}u, \phi\rangle_{(2\pi)^{-d}\mathrm{d}\xi} := \langle u, \mathcal{F}^*\phi\rangle \text{ for } u \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C}),\ \phi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C}).$$

Similarly, $\mathcal{F}^* : \mathcal{S}'(\mathbb{R}^d;\mathbb{C}) \to \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$ is defined as

$$\langle \mathcal{F}^*a, \phi\rangle := \langle u, \mathcal{F}\phi\rangle_{(2\pi)^{-d}\mathrm{d}\xi} \text{ for } a \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C}),\ \phi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C}).$$

Remark 8.7. As we will see below, in practice the precise definition of $\mathcal{F}[u]$ for $u \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$ using $\mathcal{F}^*$ is of limited utility. Instead, the computation of $\mathcal{F}[u]$ often proceeds by first approximating u by $u_j \in L^1(\mathbb{R}^d;\mathbb{C})$ (where the convergence is in the sense of $\mathcal{S}'(\mathbb{R}^d;\mathbb{C})$), computing $\mathcal{F}[u_j]$ by the explicit definition (8.5) (which works if $u_j \in L^1(\mathbb{R}^d;\mathbb{C})$), then computing the limit $\mathcal{F}[u] = \lim_{j\to\infty}\mathcal{F}[u_j]$.

Before we continue, let us take a break from the main discussion and study some basic properties of $\mathcal{S}(\mathbb{R}^d;\mathbb{C})$ and $\mathcal{S}'(\mathbb{R}^d;\mathbb{C})$. Note that

$$C_c^\infty(\mathbb{R}^d;\mathbb{C}) \subset \mathcal{S}(\mathbb{R}^d;\mathbb{C}),\ \text{so } \mathcal{S}'(\mathbb{R}^d;\mathbb{C}) \subset \mathcal{D}'(\mathbb{R}^d;\mathbb{C}),$$

where both inclusions are strict, as we can see in the following examples:

- $e^{-|x|^2} \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$ but $\notin C_c^\infty(\mathbb{R}^d;\mathbb{C})$, which is not difficult to verify.
- $e^{|x|^2} \in \mathcal{D}'(\mathbb{R}^d;\mathbb{C})$ but $\notin \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$. To show this, it suffices to find a sequence $\phi_j \in C_c^\infty(\mathbb{R}^d;\mathbb{C})$ such that $\phi_j \to 0$ in $\mathcal{S}(\mathbb{R}^d;\mathbb{C})$ as $j \to \infty$, but

  $$\langle e^{|x|^2}, \phi_j\rangle \to \infty \text{ as } j \to \infty.$$

  For instance, we may take $\phi_j = \chi(x/2^j)e^{-\frac{1}{2}|x|^2}$, where $\chi \in C_c^\infty(\mathbb{R}^d)$ is supported in the annulus $\{x \in \mathbb{R}^d : \frac{1}{2} < |x| < 4\}$ and equals 1 for $1 < |x| < 2$ (**Exercise:** Verify!).
- Another example is

  $$u = \sum_{k=0}^{\infty} \delta^{(k)}(x-k) \in \mathcal{D}'(\mathbb{R};\mathbb{C}) \text{ but } \notin \mathcal{S}'(\mathbb{R};\mathbb{C}).$$

  In fact, that $u \notin \mathcal{S}'(\mathbb{R};\mathbb{C})$ is an instance of the general fact that *any tempered distribution* is of finite order. This statement follows from an argument similar to the proof of Lemma 3.8.

We also make a simple observation that $1 \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$, and for $u \in \mathcal{S}$,

$$\langle u, 1\rangle = \int u(y)\,\mathrm{d}y,$$

where the RHS is the usual Lebesgue integral.

Let us come back to the discussion on the Fourier transform. We are now ready to prove the Fourier inversion formula and the Plancherel theorem, i.e., $\mathcal{F}^{-1} = \mathcal{F}^*$.

Theorem 8.8. *The following statements hold.*

(1) **Fourier inversion in $\mathcal{S}$.** *For $f \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$,*

$$f = \mathcal{F}^*\mathcal{F}[f] = \mathcal{F}\mathcal{F}^*[f]. \tag{8.7}$$

(2) **Plancherel.** *For $f \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$,*

$$\int_{\mathbb{R}^d} f(x)\overline{g(x)}\,\mathrm{d}x = \int_{\mathbb{R}^d} \mathcal{F}[f](\xi)\overline{\mathcal{F}g(\xi)}\,\frac{\mathrm{d}\xi}{(2\pi)^d}. \tag{8.8}$$

In particular, $\|\mathcal{F}[f]\|_{L^2_{(2\pi)^{-d}\mathrm{d}\xi}} = \|f\|_{L^2}$, so $\mathcal{F}$ extends in a unique fashion to an isometry $L^2(\mathbb{R}^d;\mathbb{C}) \to L^2_{(2\pi)^{-d}\mathrm{d}\xi}(\mathbb{R}^d;\mathbb{C})$. Similarly, $\mathcal{F}^*$ extends in a unique fashion to an isometry $L^2_{(2\pi)^{-d}\mathrm{d}\xi}(\mathbb{R}^d;\mathbb{C}) \to L^2(\mathbb{R}^d;\mathbb{C})$.

(3) **Fourier inversion in $\mathcal{S}'$.** *For $f \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$,*

$$f = \mathcal{F}^*\mathcal{F}[f] = \mathcal{F}\mathcal{F}^*[f]. \tag{8.9}$$

Proof. The key assertions are (8.7) and (8.9), which are equivalent (more precisely, dual) to each other. Once (8.7) is known, then (8.8) follows by $f = \mathcal{F}^*\mathcal{F}[f]$ and the definition of the formal adjoint $\mathcal{F}^*$; the statement for $\mathcal{F}^*$ follows from $f = \mathcal{F}\mathcal{F}^*[f]$.

We give two proofs of (8.7) and (8.9), which make the two formal derivations of (8.2) in Section 8.1 rigorous.

- *An approach using approximation.* Here, we will prove (8.7). We may write

$$\begin{aligned}\mathcal{F}^*\mathcal{F}[f] &= \int e^{i\xi\cdot x}\mathcal{F}[f](\xi)\frac{\mathrm{d}\xi}{(2\pi)^d}\\ &= \lim_{\epsilon\to 0+}\int e^{-\epsilon(|\xi_1|+\cdots+|\xi_d|)}e^{i\xi\cdot x}\mathcal{F}[f](\xi)\frac{\mathrm{d}\xi}{(2\pi)^d},\end{aligned}$$

which can be easily justified using the dominated convergence theorem and the fact that $\mathcal{F}[f] \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$. Now expanding the definition of $\mathcal{F}[f]$ and using Fubini's theorem,

$$\begin{aligned}&\lim_{\epsilon\to 0+}\int e^{-\epsilon(|\xi_1|+\cdots+|\xi_d|)}e^{i\xi\cdot x}\mathcal{F}[f](\xi)\frac{\mathrm{d}\xi}{(2\pi)^d}\\ &= \lim_{\epsilon\to 0+}\int\left(\int e^{-\epsilon(|\xi_1|+\cdots+|\xi_d|)}e^{i\xi\cdot(x-y)}\frac{\mathrm{d}\xi}{(2\pi)^d}\right)f(y)\,\mathrm{d}y\\ &= \lim_{\epsilon\to 0+}\int\left(\int e^{-\epsilon|\xi_1|}e^{i\xi_1(x^1-y^1)}\frac{\mathrm{d}\xi_1}{2\pi}\right)\cdots\left(\int e^{-\epsilon|\xi_d|}e^{i\xi_d(x^d-y^d)}\frac{\mathrm{d}\xi_d}{2\pi}\right)f(y)\,\mathrm{d}y.\end{aligned}$$

Recall that in Section 8.1, we computed the distribution limit

$$\lim_{\epsilon\to 0+}\int e^{-\epsilon|\xi|}e^{i\xi x}\frac{\mathrm{d}\xi}{2\pi} = \delta_0(x),$$

so the last line is equal to (of course, f is only in $\mathcal{S}(\mathbb{R}^d;\mathbb{C})$, but this property is enough)

$$\int\cdots\int \delta_0(x^1-y^1)\cdots\delta_0(x^d-y^d)f(y^1,\dots,y^d)\,\mathrm{d}y^1\cdots\mathrm{d}y^d = f(x^1,\dots,x^d),$$

as desired. The identity $f = \mathcal{F}\mathcal{F}^*[f]$ follows from the previous case since $\mathcal{F}^*[a](x) = (2\pi)^{-d}\mathcal{F}[a](-x)$.

- *An approach using the Gaussian.* Here, we will prove (8.9). We claim that

$$\delta_0 = \mathcal{F}^*[1], \tag{8.10}$$

which can be thought of as the adjoint-method way of making the formal identity (8.2) rigorous. Then for any $f \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$,

$$\begin{aligned}f(x) &= \langle f, \delta_0(\cdot - x)\rangle = \langle f(\cdot + x), \mathcal{F}^*[1]\rangle\\ &= \langle \mathcal{F}[f(\cdot + x)](\xi), 1\rangle_{(2\pi)^{-d}\mathrm{d}\xi}\\ &= \int \mathcal{F}[f(\cdot + x)](\xi)\frac{\mathrm{d}\xi}{(2\pi)^d}\end{aligned}$$

$$= \int e^{ix\cdot\xi} \mathcal{F}[f](\xi) \frac{\mathrm{d}\xi}{(2\pi)^d}$$
$$= \mathcal{F}^* \mathcal{F}[f],$$

as desired. The identity $f = \mathcal{F}\mathcal{F}^*[f]$ follows from the previous case since $\mathcal{F}^*[a](x) = (2\pi)^{-d}\mathcal{F}[a](-x)$.

To show (8.10), we will use the algebraic properties of $\mathcal{F}^*$ to first show that

$$\mathcal{F}^*[1] = c\delta_0 \tag{8.11}$$

for some $c \in \mathbb{C}$. By Lemma 8.3 and a duality argument, we have $\mathcal{F}^*[\partial_{\xi_j} a] = -ix^j \mathcal{F}^*[a]$ for any $a \in \mathcal{S}'(\mathbb{R}^d;\mathbb{C})$. It follows that $ix^j\mathcal{F}^*[1] = \mathcal{F}^*[\partial_{\xi_j} 1] = 0$ for any $j = 1, \dots, d$, or equivalently,

$$\langle \mathcal{F}^*[1], x^j\phi\rangle = 0 \quad \text{ for all } \phi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C}).$$

Let us fix $\chi \in C^\infty_c(\mathbb{R}^d)$ such that $\chi(0) = 1$. By the fundamental theorem of calculus, any $\psi \in \mathcal{S}(\mathbb{R}^d;\mathbb{C})$ can be written in the form

$$\begin{aligned}\psi(x) &= \chi(x)\left(\psi(0) + \sum_{j=1}^d x^j \int_0^1 \partial_j \psi(\sigma x)\,\mathrm{d}\sigma\right) + (1-\chi(x))\psi(x)\\ &= \chi(x)\psi(0) + \sum_{j=1}^d x^j \phi_j(x),\end{aligned}$$

where

$$\phi_j = \chi(x)\int_0^1 \partial_j\psi(\sigma x)\,\mathrm{d}\sigma + \frac{1-\chi(x)}{x^j}\psi(x).$$

It is not difficult to see that each ϕ_j belongs to $\mathcal{S}(\mathbb{R}^d;\mathbb{C})$. It follows that

$$\begin{aligned}\langle \mathcal{F}^*[1], \psi\rangle &= \langle \mathcal{F}^*[1], \psi(0)\chi\rangle + \sum_{j=1}^d \langle \mathcal{F}^*[1], x^j\phi_j\rangle\\ &= \langle \mathcal{F}^*[1], \chi\rangle \psi(0),\end{aligned}$$

or equivalently, (8.11) holds with $c = \langle \mathcal{F}^*[1], \chi\rangle$.

Finally, to nail down the constant c, we test (8.11) against $e^{-\frac{1}{2}|x|^2}$. Then the RHS is equal to c, whereas the LHS equals

$$\langle \mathcal{F}^*[1], e^{-\frac{1}{2}|x|^2}\rangle = \langle 1, \mathcal{F}[e^{-\frac{1}{2}|x|^2}](\xi)\rangle_{(2\pi)^{-d}\mathrm{d}\xi} = \int \overline{\mathcal{F}[e^{-\frac{1}{2}|x|^2}](\xi)}\,\frac{\mathrm{d}\xi}{(2\pi)^d}.$$

By the one-dimensional computation (8.4), it follows that

$$\mathcal{F}[e^{-\frac{1}{2}|x|^2}](\xi) = \mathcal{F}[e^{-\frac{1}{2}(x^1)^2}](\xi_1)\cdots\mathcal{F}[e^{-\frac{1}{2}(x^d)^2}](\xi_d) = (2\pi)^{\frac{d}{2}} e^{-\frac{1}{2}|\xi|^2}.$$

Now $\int e^{-\frac{1}{2}|\xi|^2}\,\mathrm{d}\xi = (2\pi)^{\frac{d}{2}}$, so the desired conclusion $c = 1$ follows. □

Remark 8.9. Observe that in each proof, the heart of the matter is to make sense of the formal identity (8.2):

$$\delta_0(x) \text{“}=\text{''} \int e^{i\xi\cdot x}\,\frac{\mathrm{d}\xi}{(2\pi)^d}.$$

It is a nice exercise to try to come up with other ways to "derive" the formal identity (8.2) and turn it into a rigorous proof like the above.

As a quick corollary of Theorem 8.8, we can compute the Fourier transform of $\delta_0(\cdot - y)$ and $e^{i\xi\cdot(\cdot)}$:

Corollary 8.10. *For any $y, \eta \in \mathbb{R}^d$, we have*

$$\mathcal{F}[\delta_0(\cdot - y)](\xi) = e^{-i\xi\cdot y}, \quad \mathcal{F}[e^{i\eta\cdot(\cdot)}](\xi) = (2\pi)^d \delta_0(\xi - \eta),$$
$$\mathcal{F}^{-1}[\delta_0(\cdot - \eta)](x) = (2\pi)^{-d} e^{i\eta\cdot x}, \quad \mathcal{F}^{-1}[e^{-i(\cdot)\cdot y}](x) = \delta_0(x - y).$$

Proof. The assertions $\mathcal{F}[\delta_0(\cdot - y)](\xi) = e^{-i\xi\cdot y}$ and $\mathcal{F}^{-1}[\delta_0(\cdot - \eta)](x) = \mathcal{F}^*[\delta_0(\cdot - \eta)](x) = e^{i\eta\cdot x}$ are easy to compute using the direct (adjoint) definition. The other two assertions then follow from Theorem 8.8. □

Let us list a few more basic properties of the Fourier transform.

- **Convolution and product.** Suppose that one of f, g is in the Schwartz class, and the other is a tempered distribution, e.g., $f \in \mathcal{S}'(\mathbb{R}^d; \mathbb{C})$ and $g \in \mathcal{S}(\mathbb{R}^d; \mathbb{C})$. Then $f * g(x) = \langle f, \overline{g}(x - \cdot)\rangle$ is a well-defined smooth function. Moreover,

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]. \tag{8.12}$$

Indeed, when f, g are both in the Schwartz class,

$$\begin{aligned}
\mathcal{F}[f * g] &= \int \left(\int f(y - z) g(z) \, \mathrm{d}z \right) e^{-i\xi\cdot y} \, \mathrm{d}y \\
&= \int \int f(y - z) g(z) e^{-i\xi\cdot(y-z)} e^{-i\xi\cdot z} \, \mathrm{d}y\mathrm{d}z \\
&= \mathcal{F}[f](\xi) \int g(z) e^{-i\xi\cdot z} \, \mathrm{d}z \\
&= \mathcal{F}[f](\xi)\mathcal{F}[g](\xi).
\end{aligned}$$

The general case can be deduced either by the approximation or the adjoint method.

By a similar computation, for $a \in \mathcal{S}'(\mathbb{R}^d; \mathbb{C})$ and $b \in \mathcal{S}(\mathbb{R}^d; \mathbb{C})$, if we define

$$a *_{(2\pi)^{-1}\mathrm{d}\xi} b(\xi) := \int a(\xi - \eta) b(\eta) \, \frac{\mathrm{d}\eta}{(2\pi)^d} = (2\pi)^{-d} a * b(\xi),$$

then

$$\mathcal{F}^{-1}[a *_{(2\pi)^{-1}\mathrm{d}\xi} b] = \mathcal{F}^{-1}[a]\mathcal{F}^{-1}[b]. \tag{8.13}$$

To summarize,

The Fourier transform turns convolutions into products, and vice versa.

As an application of the preceding property, let us introduce and discuss the concept of a *Fourier multiplier.*

Definition 8.11. A linear operator $T : \mathcal{S}(\mathbb{R}^d; \mathbb{C}) \to \mathcal{S}'(\mathbb{R}^d; \mathbb{C})$ is called a *Fourier multiplier operator* if there exists $m \in \mathcal{S}'(\mathbb{R}^d; \mathbb{C})$, called the symbol of T, such that $\mathcal{F}[Tf] = m\mathcal{F}[f]$.

By the preceding property, a Fourier multiplier operator T takes the convolution form

$$Tf = K * f,$$

where $K = \mathcal{F}^{-1}[m]$. From this form, it is evident that T is *translation-invariant*, in the sense that

$$Tf(\cdot - y) = T(f(\cdot - y)) \quad \text{for any } y \in \mathbb{R}^d.$$

Moreover, by the Plancherel theorem, we see that if $m \in L^\infty$, then

$$\|Tf\|_{L^2} = \|m\hat{f}\|_{L^2_{(2\pi)^{-d}\mathrm{d}\xi}} \leq \|m\|_{L^\infty} \|\mathcal{F}[f]\|_{L^2_{(2\pi)^{-d}\mathrm{d}\xi}} = \|m\|_{L^\infty} \|f\|_{L^2}$$

In particular, T is a bounded operator on $L^2 = L^2(\mathbb{R}^d; \mathbb{C})$; with a bit more work, it is possible to show that $\|T\|_{L^2 \to L^2} = \|m\|_{L^\infty}$. Conversely, again by the Plancherel theorem, it is easy to see that if T is a Fourier multiplier operator that is bounded in L^2, then $m \in L^\infty$.

It turns out that any translation-invariant linear operator that is bounded on L^2 must be a Fourier multiplier operator:

Proposition 8.12. *A bounded linear operator $T : L^2(\mathbb{R}^d; \mathbb{C}) \to L^2(\mathbb{R}^d; \mathbb{C})$ is translation-invariant if and only if it is a Fourier multiplier operator with a symbol $m \in L^\infty(\mathbb{R}^d; \mathbb{C})$.*

Proof. Observe that if T is translation-invariant, then for any $f, g \in \mathcal{S}(\mathbb{R}^d; \mathbb{C})$ we have

$$\begin{aligned} f * T[g](x) &= \int f(y) T[g](x-y) \, \mathrm{d}y \\ &= \int T\left[\int f(y) g(\cdot - y) \, \mathrm{d}y\right](x) \\ &= T[f * g](x). \end{aligned}$$

After conjugation with the Fourier transform, this property should imply that the conjugated operator $S = \mathcal{F}^{-1} T \mathcal{F}$ commutes with multiplication (the sense in which this holds will be made precise below). Our goal is to use this property to show that the functional

$$a \mapsto \int S[a] \, \mathrm{d}\xi$$

is a well-defined bounded linear functional on $L^1(\mathbb{R}^d; \mathbb{C})$. Since $\mathbb{R}^d$ is σ-finite, it would then follow that $S[a] = m(\xi) a(\xi)$ for some $m \in (L^1(\mathbb{R}^d; \mathbb{C})' = L^\infty(\mathbb{R}^d; \mathbb{C})$ as desired.

The property $f * T[g] = T[f * g]$ for $f, g \in \mathcal{S}(\mathbb{R}^d; \mathbb{C})$ implies that

$$aS[b] = S[ab] \tag{8.14}$$

for $a, b \in \mathcal{S}(\mathbb{R}^d; \mathbb{C})$. Since $\mathcal{F}$, $\mathcal{F}^{-1}$ and T are bounded in L^2, S is also bounded in L^2; then by approximation, we can extend (8.14) to $a \in L^\infty$ and $b \in L^2$.

In order to proceed, let us introduce the space L^∞_{comp} of bounded measurable (complex-valued) functions on $\mathbb{R}^d$ with compact support. Then for any $a \in L^\infty_{comp}$, we have $S[a] = S[a]\mathbf{1}_{\operatorname{supp} a}$, so $\int S[a] \, \mathrm{d}\xi = \int S[a]\mathbf{1}_{\operatorname{supp} a} \, \mathrm{d}\xi$ is well-defined (here, we use $S[a]$, $\mathbf{1}_{\operatorname{supp} a} \in L^2$). Clearly, the functional $L^\infty_{comp} \ni a \mapsto \int S[a] \, \mathrm{d}\xi$ is linear. It remains to show that, for $a \in L^\infty_{comp}$,

$$|S[a]| \leq C\|a\|_{L^1}$$

for some $C > 0$ independent of a. By linearity, it suffices to justify this bound for a nonnegative function $a \in L^\infty_{comp}$. In this case, by (8.14) and the Plancherel theorem,

$$|\int S[a] \, \mathrm{d}\xi| = |\int S[\sqrt{a}]\sqrt{a} \, \mathrm{d}\xi| \leq (2\pi)^d \|\sqrt{a}\|^2_{L^2} = (2\pi)^d \|a\|_{L^1},$$

as desired. □

Remark 8.13. Fourier multiplier operators are important since constant coefficient differential operators and their fundamental solutions are such operators. The study of the boundedness property of Fourier multiplier operators in translation-invariant functions spaces (e.g., L^p spaces) is a central topic in harmonic analysis. As we have seen, the L^2-boundedness property of Fourier multiplier operators is easy to understand, thanks to the Plancherel theorem. Fourier multipliers that arise from the fundamental solution of an elliptic operator turn out to obey nice L^p-boundedness properties as well; these are the subject of *Calderón–Zygmund theory* (see, also, the *Mikhlin multiplier theorem*). On the other hand, boundedness properties among L^p spaces of the fundamental solution to (even) the wave equation is much less understood, and their study is a huge topic in harmonic analysis (the relevant keywords are the *local smoothing conjecture*, the *restriction conjecture* etc.).

- **Behavior under linear change of coordinates.** Let L be a non-degenerate linear map from $\mathbb{R}^d$ to $\mathbb{R}^d$. Then for $f \in \mathcal{S}(\mathbb{R}^d; \mathbb{C})$,

$$\mathcal{F}[f \circ L](\xi) = (\det L)^{-1} \mathcal{F}[f]((L^{-1})^\top \xi). \tag{8.15}$$

Indeed,

$$\begin{aligned} \mathcal{F}[f \circ L](\xi) &= \int f(Ly) e^{-i\xi \cdot y} \, \mathrm{d}y \\ &= (\det L)^{-1} \int f(z) e^{-i\xi \cdot L^{-1} z} \, \mathrm{d}z \\ &= (\det L)^{-1} \int f(z) e^{-i(L^{-1})^\top \xi \cdot z} \, \mathrm{d}z = (\det L)^{-1} \mathcal{F}[f]((L^{-1})^\top \xi). \end{aligned}$$

These properties extend to more general functions f, g, provided that the operations involved make sense; we leave such generalizations as exercises.

As an application of the preceding formula, we compute the Fourier transform of a general Gaussian in $\mathbb{R}^d$.

Proposition 8.14 (Fourier transform of Gaussians)**.** *Let A be a symmetric positive definite matrix. Then we have*

$$\mathcal{F}[e^{-\frac{1}{2} x^\top A x}] = (2\pi)^{\frac{d}{2}} |\det A|^{-\frac{1}{2}} e^{-\frac{1}{2} \xi^\top A^{-1} \xi}.$$

The idea is to diagonalize A by an orthonormal matrix to first reduce the problem to the case when A is a diagonal matrix, and then using the explicit computation

$$\mathcal{F}[e^{-\frac{1}{2}|x|^2}] = (2\pi)^{\frac{d}{2}} e^{-\frac{1}{2}|\xi|^2}$$

from Section 8.1.

Proof. Since A is symmetric, we may write

$$A = O^\top D O,$$

where O is an orthonormal matrix (i.e., $O^\top = O^{-1}$) and D is a diagonal matrix. Making the change of variables $(x, \xi) \mapsto (Ox, O\xi)$ and using the invariance of $x^\top A x$ and $\xi^\top A^{-1} \xi$ under such a variable change, we may assume that A is diagonal, i.e.,

$$A = \mathrm{diag}(\lambda_1, \dots, \lambda_d),$$

where $0 < \lambda_1 \le \cdots \le \lambda_d$. Define $L = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$, so that $A = L^\top L$. Then

$$e^{-\frac{1}{2}x^\top Ax} = e^{-\frac{1}{2}|Lx|^2}.$$

Recall from Section 8.1 that

$$\mathcal{F}[e^{-\frac{1}{2}|x|^2}] = (2\pi)^{\frac{d}{2}} e^{-\frac{1}{2}|\xi|^2}.$$

Thus, by (8.15),

$$\mathcal{F}\left[e^{-\frac{1}{2}|Lx|^2}\right] = (2\pi)^{\frac{d}{2}} (\det L)^{-1} e^{-\frac{1}{2}|L^{-1}\xi|^2}.$$

Since $\det L = \prod_{i=1}^d \lambda_i^{1/2} = (\det A)^{1/2}$ and $|L^{-1}\xi| = \xi^\top A^{-1}\xi$, the claimed formula follows. ☐

- **Fourier transform of homogeneous distributions.**

Proposition 8.15. *The Fourier transform of distribution $h \in \mathcal{S}'(\mathbb{R}^d)$ that is homogeneous of degree a is homogeneous of degree $-a-d$.*

We leave the proof as an exercise.

Remark 8.16 (Fourier inversion on $\mathbb{T}$ and the Poisson summation formula). As an application of our distribution-theoretic approach to the Fourier transform, let us give a short derivation of the *Poisson summation formula.*

By the Fourier inversion theorem on $\mathbb{T}$,

$$\delta_0(\cdot) = \sum_k e^{2\pi i k(\cdot)},$$

in the sense of distributions. Now pullback both sides by the projection $\pi : \mathbb{R} \to \mathbb{T}$:

$$\pi^*\delta_0(x) = \sum_{k\in\mathbb{Z}} \delta_0(x-k),$$

$$\pi^* \sum_k e^{2\pi i k(\cdot)}(x) = \sum_{k\in\mathbb{Z}} \pi^* e^{2\pi i k(\cdot)} = \sum_{k\in\mathbb{Z}} e^{2\pi i k x}.$$

So on $\mathbb{R}$,

$$\sum_{k\in\mathbb{Z}} \delta_0(x-k) = \sum_{k\in\mathbb{Z}} e^{2\pi i k x}.$$

Testing this identity against any $f \in C_c^\infty(\mathbb{R};\mathbb{C})$,

$$\sum_{k\in\mathbb{Z}} f(k) = \sum_{k\in\mathbb{Z}} \hat{f}(k),$$

which is the Poisson summation formula.

8.3. **Applications to the Laplace equation (optional).** As a warm-up for what is to come, we now apply the Fourier transform to the study of the Laplacian $-\Delta$.

Alternative derivation of E_0 for $d \geq 3$. As the first application, let us give an alternative derivation of the fundamental solution E_0 in Section 4.1 in the case $d \geq 3$, using the Fourier transform.

Note that
$$-\Delta E_0 = \delta_0 \Leftrightarrow |\xi|^2 \widehat{E_0} = 1.$$
When $d \geq 3$, $\widehat{E_0}$ can be defined as the unique (tempered) distribution of homogeneity -2 such that
$$\widehat{E_0}(\xi) = \frac{1}{|\xi|^2} \quad \text{in } \mathbb{R}^d \setminus \{0\}.$$
To compute the inverse Fourier transform, we use the Gaussian.
$$\widehat{E_0}(\xi) = \frac{1}{|\xi|^2} = \int_0^\infty e^{-s|\xi|^2} \, \mathrm{d}s.$$
It follows that E_0 is the homogeneous distribution of degree $d-2$ such that in $\mathbb{R}^d \setminus \{0\}$,
$$E_0(x) = \mathcal{F}^{-1}\left[\int_0^\infty e^{-s|\xi|^2} \, \mathrm{d}s\right] = (4\pi s)^{-\frac{d}{2}} \int_0^\infty e^{-\frac{|x|^2}{4s}} \, \mathrm{d}s.$$
Making the change of variables
$$t = \frac{|x|^2}{4s}, \qquad \text{so that } \frac{\mathrm{d}t}{t} = -\frac{\mathrm{d}s}{s},$$
we see that
$$(4\pi s)^{-\frac{d}{2}} \int_0^\infty e^{-\frac{|x|^2}{4s}} \, \mathrm{d}s = \frac{1}{(4\pi)^{\frac{d}{2}}} \frac{4^{\frac{d-2}{2}}}{|x|^{d-2}} \int_0^\infty t^{\frac{d-2}{2}} e^{-t} \, \frac{\mathrm{d}t}{t} = \frac{\Gamma(\frac{d-2}{2})}{4\pi^{\frac{d}{2}}} |x|^{-d+2}.$$
Now recall that
$$d\alpha(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})},$$
and $\frac{d-2}{2}\Gamma(\frac{d-2}{2}) = \Gamma(\frac{d}{2})$. Thus,
$$\frac{\Gamma(\frac{d-2}{2})}{4\pi^{\frac{d}{2}}} = \frac{\Gamma(\frac{d}{2})}{2(d-2)\pi^{\frac{d}{2}}} = \frac{1}{d(d-2)\alpha(d)}.$$
In conclusion, E_0 is the homogeneous distribution of degree $d-2$ such that in $\mathbb{R}^d \setminus \{0\}$,
$$E_0(x) = \frac{1}{d(d-2)\alpha(d)} |x|^{-d+2}.$$

Entire tempered harmonic functions. As another application, let us prove a generalization of Liouville's theorem (Theorem 4.6):

Proposition 8.17. *Let u be a harmonic function on $\mathbb{R}^d$ that is also a tempered distribution. Then u is a polynomial.*

Proof. Since $u \in \mathcal{S}'(\mathbb{R}^d)$, we can take the Fourier transform. Then $|\xi|^2 \hat{u}(\xi) = 0$, which implies that $\operatorname{supp} u \subseteq \{0\}$. By Theorem 6.5, $\hat{u} = \sum_{\alpha : |\alpha| \leq K} c_\alpha D^\alpha \delta$ for some finite K and $c_\alpha \in \mathbb{C}$; by taking the inverse Fourier transform, we obtain the proposition. □

From this result, Liouville's theorem (Theorem 4.6) immediately follows since the only bounded polynomials are the constant functions. We also note that, of course, there exist many harmonic functions on $\mathbb{R}^d$ with $d \geq 2$ that are not tempered. Take, for instance, the real part of any entire function (e.g., $e^z = e^x(\cos y + i \sin y)$) in $\mathbb{C} = \mathbb{R}^2$.

Poisson integral formula on the half-space. As a final application, let us give an alternative derivation of the Poisson integral formula on the half-space $\mathbb{R}^d_+$.

Let us use the notation $x' = (x^1, \ldots, x^{d-1})$ and $t = x^d$. Let $g \in \mathcal{S}(\mathbb{R}^{d-1})$ and let us look for a harmonic function u on $\mathbb{R}^d_+$ such that $\lim_{(x,t)\to(x_0,0)} u(x,t) = g(x_0)$. Clearly, to uniquely specify u, we need some condition on the growth of u as $t \to \infty$ (otherwise, for instance, $u = ct$ would be a solution even when $g = 0$). Let us require

$$\|u(\cdot, t)\|_{L^1(\mathbb{R}^{d-1})} \text{ is bounded as } t \to \infty. \tag{8.16}$$

This condition more stringent than just assuming $u \in L^\infty(\mathbb{R}^d_+) \cap C(\overline{\mathbb{R}^d_+})$, but it will be convenient for reading off the correct formula.

Denoting the Fourier transform of u in x' by $\widehat{u}(\xi, t)$, we have

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

For each $\xi \in \mathbb{R}^{d-1}$ such that $\xi \neq 0$, a general solution to this ODE is of the form

$$\widehat{u}(\xi, t) = a(\xi) e^{-t|\xi|} + b(\xi) e^{t|\xi|}.$$

Thanks to (8.16), we have the pointwise identity $\widehat{u}(\xi, t) = \int u(x', t) e^{-i\xi \cdot x'} \, \mathrm{d}x'$ for each $\xi \in \mathbb{R}^{d-1}$ and $t > 0$. So if $b(\xi) \neq 0$, then $\widehat{u}(\xi)(t) \to \infty$, while $|\widehat{u}(\xi)(t)| \leq \int |u(x', t) e^{i\xi \cdot \xi'}| \, \mathrm{d}x' = \|u(\cdot, t\|_{L^1(\mathbb{R}^{d-1})}$; this situation is impossible due to (8.16). Therefore, $b(\xi) = 0$ for all $\xi \neq 0$ and

$$\widehat{u}(\xi, t) = \widehat{u}(\xi, 0) e^{-t|\xi|}. \tag{8.17}$$

It follows that

$$u(x', t) = \mathcal{F}^{-1}[e^{-t|\xi|}] * g(x').$$

It remains to compute $\mathcal{F}^{-1}[e^{-t|\xi|}]$ in $\mathbb{R}^n$ where $n = d - 1$. We proceed in several steps.

- *Step 1: The case $n = 1$.* This case can be handled easily by a direct computation.

$$\begin{aligned}
\mathcal{F}^{-1}[e^{-t|\xi|}] &= \int_{-\infty}^0 e^{t\xi} e^{ix\xi} \frac{\mathrm{d}\xi}{2\pi} + \int_0^\infty e^{-t\xi} e^{ix\xi} \frac{\mathrm{d}\xi}{2\pi} \\
&= \frac{1}{2\pi} \left(\frac{1}{t + ix} + \frac{1}{t - ix} \right) \\
&= \frac{1}{\pi} \frac{t}{t^2 + x^2}.
\end{aligned}$$

- *Step 2: Writing $e^{-t|\xi|}$ as an integral of rescaled Gaussians.* To compute $\mathcal{F}^{-1}[e^{-t|\xi|}$ on $\mathbb{R}^n$ for $n > 1$, we use a similar idea as in our derivation of the fundamental solution E_0 using the Fourier transform, i.e., we try to look for an identity of the form

$$e^{-t|\xi|} = \int_0^\infty g(s) e^{-s|\xi|^2} \, \mathrm{d}s.$$

Note that, even though we are interested in this formula for $\xi \in \mathbb{R}^n$, the identity itself only involves $|\xi|$. Therefore, we may look for such an identity assuming that $\xi \in \mathbb{R}$, which is much easier!

In the remainder of this step, $\mathcal{F}$ **refers to the Fourier transform on** $\mathbb{R}$ **and** $x, \xi \in \mathbb{R}$. From the previous step, we saw that

$$\mathcal{F}^{-1}[e^{-t|\xi|}](x) = \frac{1}{\pi} \frac{t}{t^2 + x^2}.$$

On the other hand,

$$\frac{1}{\pi} \frac{t}{t^2 + x^2} = \frac{1}{\pi} \int_0^\infty t e^{-s(t^2+x^2)} \, \mathrm{d}s$$

Taking the Fourier transform, it follows that

$$e^{-t|\xi|} = \frac{1}{\pi^{\frac{1}{2}}} \int_0^\infty t s^{-\frac{1}{2}} e^{-st^2} e^{-\frac{\xi^2}{4s}} \, \mathrm{d}s = \frac{1}{(4\pi)^{\frac{1}{2}}} \int_0^\infty t s^{-\frac{3}{2}} e^{-\frac{t^2}{4s}} e^{-s\xi^2} \, \mathrm{d}s, \tag{8.18}$$

which is the desired formula.

- *Step 3: Computation of* $\mathcal{F}^{-1}[e^{-t|\xi|}]$ *in* $\mathbb{R}^n$ Now we take (8.18), but interpret ξ as a point in $\mathbb{R}^n$ for $n > 1$. In this step, $\mathcal{F}$ **refers to the Fourier transform on** $\mathbb{R}^n$**, and** $x, \xi \in \mathbb{R}^d$.

Using (8.18), we compute

$$\begin{aligned}
\mathcal{F}^{-1}[e^{-t|\xi|}] &= \frac{1}{(4\pi)^{\frac{1}{2}}} \int_0^\infty t s^{-\frac{3}{2}} e^{-\frac{t^2}{4s}} \mathcal{F}^{-1}[e^{-s|\xi|^2}] \, \mathrm{d}s \\
&= \frac{1}{(4\pi)^{\frac{1}{2}}} \int_0^\infty t s^{-\frac{3}{2}} e^{-\frac{t^2}{4s}} (4\pi s)^{-\frac{n}{2}} e^{-\frac{|x|^2}{4s}} \, \mathrm{d}s \\
&= \frac{1}{\pi^{\frac{n+1}{2}}} \int_0^\infty t s^{\frac{n+1}{2}} e^{-s(t^2+|x|^2)} \, \frac{\mathrm{d}s}{s} \\
&= \frac{1}{\pi^{\frac{n+1}{2}}} \frac{t}{(t^2+|x|^2)^{\frac{n+1}{2}}} \int_0^\infty s^{\frac{n+1}{2}} e^{-s} \, \frac{\mathrm{d}s}{s}.
\end{aligned}$$

Recalling the definition of the Gamma function, we arrive at the formula

$$\mathcal{F}^{-1}[e^{-t|\xi|}] = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}} \frac{t}{(t^2+|x|^2)^{\frac{n+1}{2}}}. \tag{8.19}$$

9. Application of the Fourier transform to evolution equations: The heat, Schrödinger and wave equations

In this section, we will apply the Fourier transform to study the evolutionary constant coefficient linear scalar PDEs. Our primary example will be the *heat equation*, but the ideas in this section are fairly general and apply (almost) equally well to the *Schrödinger* and the *wave equations*, as we will outline. The main ideas that we will cover in this section are as follows:

- Using the Fourier transform to solve the homogeneous initial value problem;
- Using Duhamel's principle for solving the inhomogeneous initial value problem;
- computation of the spacetime Fourier transform of the forward fundamental solution.

In fact, for the equations that we consider (heat, Schödinger and wave), we will be able to invert the Fourier transform explicitly to obtain an expression for the forward fundamental solution! As we have seen in Sections 4, 5 and 7, the forward fundamental solution can then be used as the starting point for derivation of representation formulas (as well as a host of other properties) for general solutions to the initial value problem.

As we will see, the strength of the Fourier-analytic approach is that it is more systematic than the explicit computation of the forward fundamental solution as in Section 4 and 7 (remember that in each case, we had to find ad-hoc ways to exploit the symmetry properties of the partial differential operator). Moreover, through the Plancherel theorem, it easily yields very detailed information about the L^2-type norms of the solution, that are not as transparent in the fundamental-solution approach.

On the other hand, one drawback of the Fourier-analytic approach is that it is less clear to read off what happens in the physical space compared to the fundamental-solution approach, since the formula for the Fourier transform of the solution is often difficult to invert (as mentioned earlier, for the particular examples we consider it will be possible to invert the Fourier transform, but you will see that it is no simple task!). A more serious shortcoming is that the Fourier-analytic approach depends crucially on *linearity* and *translation-invariance* (i.e., that the coefficients are constant) of the partial differential operator, and ceases to work as nicely (although it is still useful!) when either of the two properties are lost.

This last point should be compared with the *energy method* (which goes hand-in-hand with the machinery of *Sobolev spaces*), which is less explicit but more robust so that it is readily applicable to nonlinear and/or variable-coefficient settings.

9.1. **The heat equation.**

Let us apply the Fourier transform to study the *heat equation*

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

We are interested in the *initial value problem* for the heat equation:

$$\begin{cases} (\partial_t - \Delta)u = f & \text{in } \mathbb{R}_+ \times \mathbb{R}^d = \mathbb{R}^{1+d}_+, \\ u = g & \text{on } \{0\} \times \mathbb{R}^d = \partial\mathbb{R}^{1+d}_+. \end{cases} \tag{9.1}$$

Representation formula in the homogeneous case. Taking the Fourier transform in space, (9.1) becomes

$$\begin{cases} \partial_t \widehat{u}(t,\xi) + |\xi|^2 \widehat{u}(t,\xi) = \widehat{f}(t,\xi), \\ \widehat{u}(0,\xi) = \widehat{g}(\xi). \end{cases} \tag{9.2}$$

Let us first consider the homogeneous case $f = 0$; then the problem has become a homogeneous first-order ODE for each fixed $\xi \in \mathbb{R}^d$. It follows that

$$\widehat{u}(t,\xi) = e^{-t|\xi|^2} \widehat{g}(\xi). \tag{9.3}$$

From this formula, it is not difficult to prove the following result.

Proposition 9.1. *The following statements hold:*

(1) **Existence.** *For $g \in L^2$, there exists a solution $u \in C_t([0,\infty); L^2)$ to* (9.1) *with $f = 0$ such that $u(0) = g$, such that*
$$\|u(t)\|_{L^2} \leq \|g\|_{L^2} \quad \textit{for every } t \geq 0.$$
(2) **Uniqueness.** *If u and v are solutions to* (9.1) *in $C_t([0,\infty); L^2)$ with the same f and g, then $u = v$.*

Recall that given a topological vector space X, $C_t(I; X)$ is the space of functions $u(t,x)$ such that
$$I \ni t \mapsto u(t,\cdot) \in X \quad \text{is continuous.}$$
When X is a normed vector space, we equip $C_t(I;X)$ with the norm $\|u\|_{C_t X} = \sup_{t \in I} \|u(t)\|_X$.

Proof. Part (1) is easily proved by defining $\widehat{u}$ via (9.3) and appealing to Theorem 8.8. To prove Part (2), note that $w := u - v \in C_t([0,\infty); L^2)$ solves the homogeneous equation with $w(0) = 0$. Thus
$$(\partial_t + |\xi|^2) \widehat{w}(t,\xi) = 0$$
in the sense of distributions, which implies
$$\partial_t (e^{t|\xi|^2} \widehat{w}(t,\xi)) = 0$$
in the sense of distributions. Since $\widehat{w}(0,\xi) = 0$, it follows that $\widehat{w}(t,\xi) = 0$, or equivalently, $w = 0$ as desired. □

Remark 9.2. We note that the condition $u \in C_t([0,\infty); L^2)$ in the uniqueness statement implies that the solution $u(t,x)$ is, in particular, bounded as $|x| \to \infty$. Such a condition on the growth of u at the spatial infinity is necessary for uniqueness, due to a classical counterexample of Tychonoff [Tyc35].

In the case of the heat equation, it is easy to take the inverse Fourier transform of (9.3). Then we obtain the formula

$$u(t,\cdot) = K_t * g, \tag{9.4}$$

for the solution to (9.1) with $f = 0$, where

$$K_t(x) = \mathcal{F}^{-1}[e^{-t|\xi|^2}] = \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} \quad \text{for } t > 0,\ x \in \mathbb{R}^d, \tag{9.5}$$

by Proposition 8.14. Note that $K_t(x) = (4t)^{-\frac{d}{2}} K_1\left(\frac{x}{2\sqrt{t}}\right)$, so $K_t \rightharpoonup \delta_0$ as $t \to 0+$ via Lemma 3.19. It follows that we have the convergence $u(t,\cdot) \to g$ as $t \to 0+$ in any space X where the usual approximation of the identity argument applies (e.g., any L^p for $1 \leq p < \infty$).

Representation formula in the inhomogeneous case. Let $f \in L^1_t((0,T);L^2)$. Applying Duhamel's formula to the ODE (9.2) in the Fourier space (i.e., we take $\mathcal{A}_t = |\xi|^2$ and $F = \mathbf{1}_{(0,T)}(t)\widehat{f}$), we see that

$$\widehat{u_f}(t,\xi) = \int_0^t e^{-(t-s)|\xi|^2}\widehat{f}(s,\xi)\,\mathrm{d}s$$

obeys $(\partial_t + |\xi|^2)\widehat{u_f}(t,\xi) = \mathbf{1}_{(0,T)}(t)\widehat{f}$, $\widehat{u_f}(t,\xi) \in C_t([0,T];L^2)$ and $\widehat{u_f}(0,\xi) = 0$. Thus, $\widehat{u} - \widehat{u_f}$ solves the homogeneous equation with initial data $\widehat{g}$. Combined with (9.3), we arrive at the formula

$$\widehat{u}(t,\xi) = e^{-t|\xi|^2}\widehat{g}(\xi) + \int_0^t e^{-(t-s)|\xi|^2}\widehat{f}(s,\xi)\,\mathrm{d}s \quad \text{for } 0 \leq t \leq T. \tag{9.6}$$

From this formula, we obtain the following strengthening of Proposition 9.1.

Theorem 9.3. *The following statements hold:*

(1) **Existence.** *For $g \in L^2$ and $f \in L^1_t((0,T);L^2)$, there exists a solution $u \in C_t([0,T];L^2)$ to (9.1) such that*

$$\|u(t)\|_{L^2} \leq \|g\|_{L^2} + \|f\|_{L^1_t((0,t);L^2)} \quad \text{for every } t \geq 0.$$

(2) **Uniqueness.** *If u and v are solutions to (9.1) in $C_t([0,\infty);L^2)$ with the same f and g, then $u = v$.*

The proof is straightforward, so we will not go through the details.

Remark 9.4. We also note that Duhamel's principle can also be applied directly to the heat equation in the physical space. More specifically, we take $\mathcal{A}_t = -\Delta$ and $F = \mathbf{1}_{(0,T)}(t)f$ where $f \in L^1_t((0,T);X)$, where X is any normed vector space in which the homogeneous equation is well-posed. In view of (9.4) for homogeneous solutions, we arrive at the formula

$$u(t,x) = K_t * g(x) + \int_0^t K_{t-s} * f(s)(x)\,\mathrm{d}s,$$

where K_t is given by (9.5).

Forward fundamental solution for the heat equation. Recall that a *forward fundamental solution* E_+ may be constructed by solving

$$\begin{cases} (\partial_t - \Delta)E_+(t,x) = 0 & \text{in } \{t > 0\}, \\ E_+(0,x) = \delta_0(x) & \text{on } \{t = 0\}, \end{cases}$$

and setting $E_+(t,x) = 0$ in $\{t < 0\}$. By (9.3) and the fact that $\mathcal{F}[\delta_0] = 1$, we have

$$E_+(t,x) = \mathbf{1}_{(0,\infty)}(t)\mathcal{F}^{-1}[e^{-t|\xi|^2}](x). \tag{9.7}$$

By (9.5), we are led to the expression

$$E_+(t,x) = \mathbf{1}_{(0,\infty)}(t)\frac{1}{(4\pi t)^{\frac{d}{2}}}e^{-\frac{|x|^2}{4t}}. \tag{9.8}$$

Indeed, (9.8) coincides with the previous computation in Section 5!

The spacetime Fourier transform of the forward fundamental solution. Finally, we study the spacetime Fourier transform $\mathcal{F}_{t,x}[E_+]$ of the forward fundamental solution E_+ for $\partial_t - \Delta$. As we will see, this road will lead to yet another venue for the computation of E_+.

A naive first attempt to compute $\mathcal{F}_{t,x}[E_+]$ may be to take the spacetime Fourier transform of the equation $(\partial_t - \Delta)E_+ = \delta_0$, which gives

$$(i\tau + |\xi|^2)\mathcal{F}_{t,x}[E_+] = 1.$$

This computation immediately tells us that $\mathcal{F}_{t,x}[E_+] = (i\tau + |\xi|^2)^{-1}$ away from the zero set of $(i\tau + |\xi|^2)$ (which is just $\{0\}$), but it does not tell us the precise behavior of $\mathcal{F}_{t,x}[E_+]$ on the zero set.

In a sense, we should have foreseen this problem, because we did not use the forward support condition (i.e., $\operatorname{supp} E_+ \subseteq \mathbb{R}^{1+d}_+$) at all! The preceding consideration proves that *all fundamental solutions have the same spacetime Fourier transform away from the zero set of* $i\tau + |\xi|^2$; their differences are all due to the subtle structure of the distributions near the zero set. In fact, this property holds for a general constant coefficient partial differential operator.

To compute $\mathcal{F}_{t,x}[E_+]$ rigorously, the idea is to introduce a natural approximation to E_+ that takes advantage of the forward support condition. More precisely, since $\operatorname{supp} E_+ \subseteq \mathbb{R}^{1+d}_+$, we have

$$\lim_{\epsilon \to 0+} e^{-\epsilon t} E_+ = E_+ \quad \text{in } \mathcal{S}'(\mathbb{R}^{1+d}; \mathbb{C}).$$

Note that

$$(\partial_t - \Delta)(e^{-\epsilon t} E_+) = -\epsilon e^{-\epsilon t} E_+ + e^{-\epsilon t}(\partial_t - \Delta) E_+ = -\epsilon e^{-\epsilon t} E_+ + \delta_0,$$

or equivalently,

$$(\partial_t + \epsilon - \Delta)(e^{-\epsilon t} E_+) = \delta_0.$$

Taking the spacetime Fourier transform,

$$(i\tau + \epsilon + |\xi|^2)\mathcal{F}_{t,x}[e^{-\epsilon t} E_+] = 1,$$

so that

$$\mathcal{F}_{t,x}[e^{-\epsilon t} E_+] = \frac{1}{i\tau + \epsilon + |\xi|^2} = \frac{1}{i(\tau - i\epsilon) + |\xi|^2}.$$

For each $\epsilon > 0$, the RHS is clearly locally integrable; one can also check that it is in $\mathcal{S}'(\mathbb{R}^{1+d}; \mathbb{C})$. Thus, we arrive at the following conclusion:

$$\mathcal{F}_{t,x}[E_+] = \lim_{\epsilon \to 0+} \mathcal{F}_{t,x}[e^{-\epsilon t} E_+] = \lim_{\epsilon \to 0+} \frac{1}{i(\tau - i\epsilon) + |\xi|^2}. \tag{9.9}$$

Remark 9.5. If we started with the backward fundamental solution, i.e., $(\partial_t - \Delta)E_- = \delta_0$ with $\operatorname{supp} E_- \subseteq (-\infty, 0] \times \mathbb{R}^d$, then the same procedure gives

$$\mathcal{F}_{t,x}[E_-] = \lim_{\epsilon \to 0+} \frac{1}{i(\tau + i\epsilon) + |\xi|^2}.$$

Remark 9.6. The correspondence of the analytic continuation property of the Fourier transform to the lower [resp. upper] half-space and the forward [resp. backward] support property of the original function is a special instance of classes of results the so-called *Paley–Wiener-type theorems*.

Observe that our derivation of (9.9) did *not* rely on any specific properties of E_+, except for the forward support condition and that $E_+ \in \mathcal{S}'(\mathbb{R}^{1+d})$. In fact, the expression (9.9) provides another (independent) starting point for the derivation of the forward fundamental solution for the heat equation. We need the following lemma:

Lemma 9.7. *Let* $a \in \overline{\mathbb{H}} = \{a \in \mathbb{C} : \operatorname{Im} a \geq 0\}$. *Then*

$$\lim_{\epsilon \to 0+} \mathcal{F}_t^{-1}\left[\frac{1}{\tau - i\epsilon - a}\right] = i\mathbf{1}_{(0,\infty)}(t)e^{iat}.$$

For those who are familiar with complex analysis, it is a nice exercise to try to prove this identity directly using the Cauchy integral formula (see also Remark 9.6), at least when $t \neq 0$. Here, we take a short cut and compute the Fourier transform of the RHS, and then appeal to Theorem 8.8.

Proof. By Theorem 8.8, it suffices to show that

$$\mathcal{F}[-i\mathbf{1}_{(0,\infty)}(t)e^{iat}](\tau) = \lim_{\epsilon \to 0+} \frac{1}{\tau - i\epsilon - a}.$$

To compute the Fourier transform on the LHS, we use the approximation method. Note that since $\operatorname{Im} a \geq 0$,

$$i\mathbf{1}_{(0,\infty)}(t)e^{-\epsilon t}e^{iat} \to -i\mathbf{1}_{(0,\infty)}(t)e^{iat}$$

uniformly, and thus also in the sense of tempered distributions. Moreover, for each fixed $\epsilon > 0$, the LHS is in L^1. Thus,

$$\begin{aligned}
\mathcal{F}[i\mathbf{1}_{(0,\infty)}(t)e^{iat}](\tau) &= \lim_{\epsilon \to 0+} \int i\mathbf{1}_{(0,\infty)}(t)e^{-\epsilon t}e^{iat}e^{-i\tau t}\,\mathrm{d}t \\
&= \lim_{\epsilon \to 0+} i\int_0^\infty e^{-(\epsilon - ia + i\tau)t}\,\mathrm{d}t \\
&= \lim_{\epsilon \to 0+} i\frac{1}{\epsilon - ia + i\tau} \\
&= \lim_{\epsilon \to 0+} \frac{1}{\tau - i\epsilon - a},
\end{aligned}$$

as desired. □

By (9.9) and the preceding lemma, we have

$$\begin{aligned}
\mathcal{F}[E_+](t, \xi) &= \lim_{\epsilon \to 0+} \mathcal{F}_t^{-1}\left[\frac{1}{i(\tau - i\epsilon) + |\xi|^2}\right] \\
&= \mathbf{1}_{(0,\infty)}(t)e^{-t|\xi|^2}.
\end{aligned}$$

Now inverting the space Fourier transform $\mathcal{F}$ using Proposition 8.14, we again obtain (9.8).

9.2. **The Schrödinger equation.** Next, we consider the *initial value problem* for the Schrödinger equation:

$$(9.10) \qquad \begin{cases} (i\partial_t - \Delta)u = f & \text{in } \mathbb{R}_+ \times \mathbb{R}^d = \mathbb{R}^{1+d}_+, \\ u = g & \text{on } \{0\} \times \mathbb{R}^d = \partial\mathbb{R}^{1+d}_+. \end{cases}$$

We will closely follow our discussion of the heat equation. As we will see, there are a lot of algebraic similarities, but the actual natures are quite different.

Representation formula in the homogeneous case. Taking the Fourier transform in space, (9.10) becomes

$$\begin{cases} i\partial_t \widehat{u}(t,\xi) + |\xi|^2 \widehat{u}(t,\xi) = \widehat{f}(t,\xi), \\ \widehat{u}(0,\xi) = \widehat{g}(\xi). \end{cases} \tag{9.11}$$

As before, we begin with the homogeneous case $f = 0$. Solving the resulting homogeneous first-order ODE for each fixed $\xi \in \mathbb{R}^d$, we obtain

$$\widehat{u}(t,\xi) = e^{it|\xi|^2} \widehat{g}(\xi). \tag{9.12}$$

This formula looks very similar to (9.3) except for the factor of i in the exponential; but of course, this makes a world of difference. For instance, in (9.3), each Fourier coefficient (for $\xi \neq 0$) decreases exponentially to the future. In particular, for a general element $g \in L^2$, $\widehat{u}(t,\xi)$ according to (9.3) is not even a tempered distribution in $t < 0$ (this is related to the time-*irreversility* of the heat equation). On the other hand, in (9.12), the amplitude of each Fourier coefficient remains same for all time. Indeed, by the Plancherel theorem, we have the *conservation of the total probability*

$$\|u(t,\cdot)\|_{L^2} = \|\widehat{u}(t,\cdot)\|_{L^2_{(2\pi)^{-d}\mathrm{d}\xi}} = \|\widehat{g}(\cdot)\|_{L^2_{(2\pi)^{-d}\mathrm{d}\xi}} = \|g\|_{L^2}.$$

Moreover, unlike (9.3), (9.12) makes perfect sense when $t < 0$ for any $g \in L^2$. Indeed, the equation $(i\partial_t - \Delta)u = 0$ is time-reversible, in the sense that it is invariant under the *time-reversal* symmetry $u(t,x) \mapsto \overline{u}(-t,x)$.

Remark 9.8. With a bit of complex analysis, we can also compute the formula for u in the physical space. We may write

$$u(t,\cdot) = K_t^{(Sch)} * g,$$

where $K_t^{(Sch)}(x) = \mathcal{F}^{-1}[e^{it|\xi|^2}]$. By Proposition 8.14 and analytic continuation, we have

$$K_t^{(Sch)}(x) = \mathcal{F}^{-1}[e^{it|\xi|^2}] = \frac{1}{(-4\pi it)^{\frac{d}{2}}} e^{\frac{|x|^2}{4it}} \qquad \text{for } t > 0,\ x \in \mathbb{R}^d,$$

where $(-4\pi it)^{\frac{1}{2}}$ is the square root of $-4\pi it$ with the positive real part. Unlike (9.5), note that $K_t^{(Sch)}$ is not absolutely integrable, so the convergence $K_t^{(Sch)} \rightharpoonup \delta_0$ in $\mathcal{S}'(\mathbb{R}^d;\mathbb{C})$ as $t \to 0+$ (which can be deduced from the Fourier transform) does *not* follow from a usual approximation of the identity argument.

Representation formula in the inhomogeneous case. Applying Duhamel's principle to the ODE in the Fourier space (i.e., $\mathcal{A}_t = \frac{1}{i}|\xi|^2$ and $F = \mathbf{1}_{(0,\infty)}(t)\widehat{f}$ for $f \in L^1((0,\infty);L^2)$) as well as (9.12), we obtain

$$\widehat{u}(t,\xi) = e^{it|\xi|^2}\widehat{g}(\xi) + \frac{1}{i}\int_0^t e^{i(t-s)|\xi|^2} \widehat{f}(s,\xi)\,\mathrm{d}s.$$

Based on this formula, it is not difficult to prove the following result:

Theorem 9.9. *The following statements hold:*

(1) **Existence.** *For $g \in L^2$ and $f \in L^1_t((0,T);L^2)$, there exists a solution $u \in C_t([0,T];L^2)$ to (9.10) such that*

$$\|u(t)\|_{L^2} \leq \|g\|_{L^2} + \|f\|_{L^1_t((0,t);L^2)} \qquad \textit{for every } t \geq 0.$$

(2) **Uniqueness.** *If u and v are solutions to (9.10) in $C_t([0,\infty);L^2)$ with the same f and g, then $u = v$.*

The proof is very similar to that for Proposition 9.1 and Theorem 9.3, so we omit the details.

Remark 9.10. Applying Duhamel's principle in the physical space, we obtain the formula

$$u(t,x) = K_t^{(Sch)} * g(x) + \frac{1}{i}\int_0^t K_{t-s}^{(Sch)} * f(s)(x)\,\mathrm{d}s.$$

Forward fundamental solution for the Schrödinger equation. As in the case of the heat equation, (**??**), which is a basic computation behind Duhamel's principle, suggests that the forward fundamental solution is given by

$$E_+(t,x) = \mathbf{1}_{(0,\infty)}(t)\mathcal{F}^{-1}[e^{it|\xi|^2}](x). \tag{9.13}$$

If we use Remark 9.8, we arrive at the formula

$$E_+(t,x) = \mathbf{1}_{(0,\infty)}(t)\frac{-i}{(-4\pi i t)^{\frac{d}{2}}}e^{\frac{|x|^2}{4it}}. \tag{9.14}$$

As remarked earlier, (9.14) algebraically resembles (9.8), but the nature of the two forward fundamental solutions is very different.

- In contrast to the heat case, the Schrödinger forward fundamental solution $E_+(t,x)$ is singular along the hyperplane $\{t=0\}$. As a consequence, no regularity theorem like [Eva10, Section 2.3, Theorem 8] is available.
- The Schrödinger forward fundamental solution $E_+(t,x)$ does not have a definite sign, so no maximum principle like [Eva10, Section 2.3, Theorem 4] is available.
- Finally, the Schrödinger forward fundamental solution $E_+(t,x)$ is *not* integrable in x for each fixed $t > 0$.

Because of the first and third properties, it takes much more work to justify taking the convolution $E_+ * u$ (where $\operatorname{supp} u \subseteq \{t \geq L\}$ for some $L \in \mathbb{R}$) if we work purely in the physical space.

The spacetime Fourier transform of the forward fundamental solution. As in the case of the heat equation, the spacetime Fourier transform of the forward fundamental solution takes the form

$$\mathcal{F}_{t,x}[E_+] = \lim_{\epsilon\to 0+}\mathcal{F}_{t,x}[e^{-\epsilon t}E_+] = \lim_{\epsilon\to 0+}\frac{1}{-(\tau - i\epsilon) + |\xi|^2}. \tag{9.15}$$

We note that an application of Lemma 9.7 leads to an alternative derivation of (9.13).

9.3. **The wave equation (optional).** Finally, we re-consider the *initial value problem* for the wave equation (7.1) using the Fourier transform.

Representation formula in the homogeneous case. Taking the Fourier transform in space, (7.1) becomes

$$\begin{cases} -\partial_t^2\widehat{u}(t,\xi) - |\xi|^2\widehat{u}(t,\xi) = \widehat{f}(t,\xi), \\ \widehat{u}(0,\xi) = \widehat{g}(\xi), \\ \partial_t\widehat{u}(0,\xi) = \widehat{h}(\xi). \end{cases} \tag{9.16}$$

As before, we begin with the homogeneous case $f = 0$. Solving the resulting homogeneous second-order ODE $-\partial_t^2 \widehat{u}(t,\xi) - |\xi|^2 \widehat{u}(t,\xi) = 0$ for each fixed $\xi \in \mathbb{R}^d$, we obtain

$$\widehat{u}(t,\xi) = \cos t|\xi| \widehat{g}(\xi) + \frac{\sin t|\xi|}{|\xi|} \widehat{h}(\xi). \tag{9.17}$$

Duhamel's principle for second-order evolutionary equations. To handle the inhomogeneous problem, we need to adapt our derivation of Duhamel's principle to the second-order time derivative[18].

Consider now the abstract second-order evolutionary equation

$$(\partial_t^2 + \partial_t \mathcal{A}_t + \mathcal{B}_t) u = F, \tag{9.18}$$

where $\mathcal{A}_t = \sum_\alpha a_\alpha(t,x) D^\alpha$ and $\mathcal{B}_t = \sum_\alpha b_\alpha(t,x) D^\alpha$ do *not* involve any time derivatives. For each fixed $s \in \mathbb{R}$ and (g,h) in some normed vector space $X_0 \times X_1$ of pairs of functions on $\mathbb{R}^d$, suppose that there exists a solution $S(t,s)[g,h]$ to the initial value problem

$$(\partial_t^2 + \partial_t \mathcal{A}_t + \mathcal{B}_t)(S(t,s)[g,h]) = 0, \quad (S(s,s)[g,h], \partial_t S(s,s)[g,h]) = (g,h),$$

such that $(S(t,s)[g,h]\partial_t S(t,s)[g,h] \in C_t([s,\infty); X_0 \times X_1)$. Then we have

$$\begin{aligned} &(\partial_t^2 + \partial_t \mathcal{A}_t + \mathcal{B}_t)(\mathbf{1}_{(s,\infty)}(t) S(t,s)[g,h]) \\ &= \partial_t(\delta_0(t-s) g) + \delta_0(t-s)\mathcal{A}_0 g + \delta_0(t-s) h. \end{aligned} \tag{9.19}$$

Accordingly, *Duhamel's formula for a second-order evolutionary equation* takes the form

$$\begin{aligned} u_F(t,x) &= \int_{-\infty}^{\infty} \mathbf{1}_{(s,\infty)}(t) S(t,s)[0, F(s)](x) \, \mathrm{d}s \\ &= \int_{-\infty}^{s} S(t,s)[0,F(s)](x) \, \mathrm{d}s. \end{aligned} \tag{9.20}$$

At least formally, we then have

$$\begin{aligned} (\partial_t^2 + \partial_t \mathcal{A}_t + \mathcal{B}_t) u(t,x) &= \int_{-\infty}^{\infty} (\partial_t^2 + \partial_t \mathcal{A}_t + \mathcal{B}_t)(\mathbf{1}_{(s,\infty)}(t) S(t,s)[0, f(s)])(x) \, \mathrm{d}s \\ &= \int_{-\infty}^{\infty} \delta_0(t-s) f(s,x) \, \mathrm{d}s = f(t,x). \end{aligned}$$

Moreover, if $F \in L^1_t((a,b); X_1)$, then provided that we have a quantitative estimate for $S(t,s)[g,h]$ of the form

$$\|S(t,x)[g,h]\|_{X_0 \times X_1} \leq C \|(g,h)\|_{X_0 \times X_1}, \tag{9.21}$$

where C is independent of $t, s \in [a,b]$ and $(g,h) \in X_0 \times X_1$, we have $(u_F, \partial_t u_F) \in C_t([a,b]; X_0 \times X_1)$,

$$u_F(t,x) = \int_a^t S(t,s)[0,F(s)](x) \, \mathrm{d}s \quad \text{for } a \leq t \leq b,$$

and $(u_F, \partial_t u_F)(a,x) = 0$.

[18]For an alternative approach that still relies on Duhamel's principle for first-order evolutionary equations, see Remark 9.12.

9.3.1. *Representation formula in the inhomogeneous case.* Let $f \in L^1_t((0,T);L^2)$. Applying Duhamel's formula to the ODE (9.16) in the Fourier space (i.e., we take $\mathcal{A}_t = 0$ and $\mathcal{B}_t = |\xi|^2$ and $F = \mathbf{1}_{(0,T)}(t)\widehat{f})$ to take case of the inhomogeneity f, and using (9.3) for the remainder, we arrive at the formula

$$\widehat{u}(t,\xi) = \cos t|\xi|\widehat{g}(\xi) + \frac{\sin t|\xi|}{|\xi|}\widehat{h}(\xi) - \int_0^t \frac{\sin(t-s)|\xi|}{|\xi|}\widehat{f}(s,\xi)\,\mathrm{d}s. \tag{9.22}$$

Based on this formula, we obtain the following result:

Theorem 9.11. *Define the (L^2-Sobolev) space $H^1 = \{g : g \in L^2,\ \partial_j g \in L^2$ for all $j\}$ equipped with the norm $\|g\|_{H^1}^2 = \|g\|_{L^2}^2 + \sum_{j=1}^d \|\partial_j g\|_{L^2}^2$.*

(1) **Existence.** *For $(g,h) \in H^1 \times L^2$ and $f \in L^1_t((0,T);L^2)$, there exists a solution u to* (7.1) *such that $(u, \partial_t u) \in C_t([0,T]; H^1 \times L^2)$ and for every $0 \le t \le T$,*

$$\begin{aligned}
\|Du(t)\|_{L^2} &\le \|Dg\|_{L^2} + \|h\|_{L^2} + \|f\|_{L^1_t((0,t);L^2)} \\
\|\partial_t u(t)\|_{L^2} &\le \|Dg\|_{L^2} + \|h\|_{L^2} + \|f\|_{L^1_t((0,t);L^2)} \\
\|u(t)\|_{L^2} &\le \|g\|_{L^2} + t\|h\|_{L^2} + t\|f\|_{L^1_t((0,t);L^2)}.
\end{aligned}$$

(2) **Uniqueness.** *If u and v are solutions to* (7.1) *such that $(u,\partial_t u), (v, \partial_t v) \in C_t([0,\infty); H^1 \times L^2)$ with the same f, g and h, then $u = v$.*

Proof. Part (1) is easily proved by defining $\hat{u}$ via (9.22) and using Theorem 8.8. For Part (2), note that $w = u - v$ solves the homogeneous equation with $(w, \partial_t w) \in C_t(H^1 \times L^2)$ and $(w, \partial_t w)(0) = 0$. We have

$$(\partial_t^2 + |\xi|^2)\widehat{w}(t,\xi) = 0$$

in the sense of distributions. By factoring $(\partial_t^2 + |\xi|^2) = (\partial_t - |\xi|)(\partial_t + |\xi|)$, we see that

$$\partial_t \left(e^{it|\xi|}(\partial_t + |\xi|)\widehat{w}(t,\xi) \right) = 0$$

in the sense of distributions. By $(w, \partial_t w)(0) = 0$, it follows that the expression inside the parenthesis is zero, i.e.,

$$e^{it|\xi|}(\partial_t + |\xi|)\widehat{w}(t,\xi) = 0.$$

But then we have

$$\partial_t \left(e^{-it|\xi|}\widehat{w}(t,\xi) \right) = 0$$

in the sense of distributions. Again using $(w, \partial_t w)(0) = 0$, it follows that the expression inside the parenthesis is zero, which implies $\widehat{w} = 0$ as desired. □

Remark 9.12 (The first-order formulation of the wave equation). Taking a cue from the ODE theory, an alternative way to prove the preceding result is to reformulate the wave equation as a first-order evolutionary system. We briefly sketch the key computations, which are often useful in practice.

One begins by introducing the variables $(u_0, u_1) = (u, \partial_t u)$ and rewriting the wave equation $\Box u = f$ in the following fashion:

$$\partial_t \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \Delta & 0 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} - \begin{pmatrix} 0 \\ f \end{pmatrix}.$$

Taking the space Fourier transform, we obtain

$$\partial_t \begin{pmatrix} \widehat{u}_0 \\ \widehat{u}_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -|\xi|^2 & 0 \end{pmatrix} \begin{pmatrix} \widehat{u}_0 \\ \widehat{u}_1 \end{pmatrix} - \begin{pmatrix} 0 \\ \widehat{f} \end{pmatrix}.$$

Now this is a system of first-order ODEs. If we diagonalize the above 2×2-matrix, we arrive at

$$\partial_t \begin{pmatrix} \widehat{u}_+ \\ \widehat{u}_- \end{pmatrix} = \begin{pmatrix} i|\xi| & 0 \\ 0 & -i|\xi| \end{pmatrix} \begin{pmatrix} \widehat{u}_+ \\ \widehat{u}_- \end{pmatrix} - \begin{pmatrix} \frac{1}{2i|\xi|} \widehat{f} \\ -\frac{1}{2i|\xi|} \widehat{f} \end{pmatrix}.$$

where

$$\widehat{u}_\pm = \frac{1}{2} \widehat{u}_0 \pm \frac{1}{2i|\xi|} \widehat{u}_1.$$

Each (decoupled) first-order equation for $\widehat{u}_\pm$ closely resembles the Schödinger equation. Formula (9.22) and Theorem 9.11 can be alternatively proved by studying these equations, proceeding similarly as in the case of the Schrödinger equation.

Forward fundamental solution for the wave equation via the Fourier transform. Here, we give an alternative derivation of the forward fundamental solution for the wave equation using the Fourier transform when $d \geq 2$. For this computation, we need the formula (8.19) and a little bit of complex analysis.

As it can be read off from the derivation of Duhamel's principle, the forward fundamental solution takes the form

$$E_+(t, x) = -\mathbf{1}_{(0,\infty)}(t) \mathcal{F}^{-1} \left[\frac{\sin t|\xi|}{|\xi|} \right]. \tag{9.23}$$

Let us write $\widehat{K}_t(\xi) = \frac{\sin t|\xi|}{|\xi|}$, and consider the approximation

$$\widehat{K}^\epsilon_t(\xi) = e^{-\epsilon|\xi|} \frac{\sin t|\xi|}{|\xi|}.$$

Clearly, $\widehat{K}^\epsilon_t \to \widehat{K}_t$ uniformly, and thus also as tempered distributions, as $\epsilon \to 0$. Therefore,

$$\mathcal{F}^{-1} \left[\frac{\sin t|\xi|}{|\xi|} \right] = \lim_{\epsilon \to 0+} \mathcal{F}^{-1} \left[\widehat{K}^\epsilon_t \right].$$

A key advantage of the new RHS is that for each fixed $\epsilon > 0$, $\widehat{K}^\epsilon_t \in L^1$, so we can use the pointwise definition of the inverse Fourier transform, i.e.,

$$\begin{aligned} \mathcal{F}^{-1} \left[\widehat{K}^\epsilon_t \right](x) &= \int e^{-\epsilon|\xi|} \frac{\sin t|\xi|}{|\xi|} e^{i\xi \cdot x} \frac{\mathrm{d}\xi}{(2\pi)^d} \\ &= \int \frac{e^{-(\epsilon - it)|\xi|} - e^{-(\epsilon + it)|\xi|}}{2i|\xi|} e^{i\xi \cdot x} \frac{\mathrm{d}\xi}{(2\pi)^d} \\ &= \frac{1}{2i} \int_{\epsilon - it}^{\epsilon + it} \int e^{-s|\xi|} e^{i\xi \cdot x} \, \mathrm{d}\xi \, \mathrm{d}s. \end{aligned}$$

For the inner integral, for $s \in \mathbb{C}$ such that $\mathrm{Re}\, s > 0$, if we let $(s^2 + |x|^2)^{\frac{1}{2}}$ be the square root of $s^2 + |x|^2$ with the positive real part, then

$$\int e^{-s|\xi|} e^{i\xi \cdot x} \, \mathrm{d}\xi = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}} \frac{s}{(s^2 + |x|^2)^{\frac{d+1}{2}}}.$$

Indeed, this identity is exactly (8.19) when s lies on the positive real axis. Moreover, both sides define holomorphic functions on $\{\mathrm{Re}\, s > 0\}$ that agree on the positive real axis; hence the identity follows. It follows that

$$\mathcal{F}^{-1} \left[\widehat{K}^\epsilon_t \right](x) = \frac{1}{2i} \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}} \int_{\epsilon - it}^{\epsilon + it} \frac{s}{(s^2 + |x|^2)^{\frac{d+1}{2}}} \, \mathrm{d}s.$$

For $d > 1$, we have

$$\mathcal{F}^{-1}\left[\widehat{K}_t^{\epsilon}\right](x) = -\frac{1}{4i}\frac{\Gamma\left(\frac{d-1}{2}\right)}{\pi^{\frac{d+1}{2}}}\left(\frac{1}{(-(t-i\epsilon)^2+|x|^2)^{\frac{d-1}{2}}} - \frac{1}{(-(t+i\epsilon)^2+|x|^2)^{\frac{d-1}{2}}}\right).$$

Thus,

$$E_+(t,x) = \lim_{\epsilon\to 0+}\mathbf{1}_{(0,\infty)}(t)\frac{1}{4i}\frac{\Gamma\left(\frac{d-1}{2}\right)}{\pi^{\frac{d+1}{2}}}\left(\frac{1}{(-(t-i\epsilon)^2+|x|^2)^{\frac{d-1}{2}}} - \frac{1}{(-(t+i\epsilon)^2+|x|^2)^{\frac{d-1}{2}}}\right).$$

Finally, note that the RHS defines a distribution on $\mathbb{R}^{1+d}$ that is homogeneous of degree $-d+1$ that clearly vanishes in the open set $\{|x|^2 > t^2\} \cup \{t < 0\}$. From these properties, as well as Lemma 6.8, we see that

$$E_+(t,x) = \lim_{\epsilon\to 0+}\mathbf{1}_{(0,\infty)}(t)\frac{1}{4i}\frac{\Gamma\left(\frac{d-1}{2}\right)}{\pi^{\frac{d+1}{2}}}\left(\frac{1}{(-t^2+|x|^2+i\epsilon)^{\frac{d-1}{2}}} - \frac{1}{(-t^2+|x|^2-i\epsilon)^{\frac{d-1}{2}}}\right).$$

We now use the identity

$$\lim_{\epsilon\to 0+}((s+i\epsilon)^a - (s-i\epsilon)^a) = 2i\sin(a\pi)\Gamma(1+a)\chi_-^a(s),$$

which follows by first verifying both sides for $\operatorname{Re} a > 0$, and then observing that both sides are entire in a. (To see that $\sin(a\pi)\Gamma(1+a)$ is analytic, use $\Gamma(-a)\Gamma(1-(-a)) = \frac{\pi}{\sin((-a)\pi)}$.) It follows that

$$\begin{aligned} E_+(t,x) &= \mathbf{1}_{(0,\infty)}(t)\frac{\sin(-\frac{d-1}{2}\pi)}{2}\frac{\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(1-\frac{d-1}{2}\right)}{\pi^{\frac{d+1}{2}}}\chi_-^{-\frac{d-1}{2}}(-t^2+|x|^2) \\ &= -\mathbf{1}_{(0,\infty)}(t)\frac{1}{2\pi^{\frac{d-1}{2}}}\chi_+^{-\frac{d-1}{2}}(t^2-|x|^2). \end{aligned}$$

The spacetime Fourier transform of the forward fundamental solution. Proceeding as in the case of the heat equation, it is not difficult to see that the spacetime Fourier transform of the forward fundamental solution takes the form

$$\begin{aligned} \mathcal{F}_{t,x}[E_+] = \lim_{\epsilon\to 0+}\mathcal{F}_{t,x}[e^{-\epsilon t}E_+] &= \lim_{\epsilon\to 0+}\frac{1}{(\tau-i\epsilon)^2-|\xi|^2} \\ &= \lim_{\epsilon\to 0+}\frac{1}{2|\xi|}\left(\frac{1}{\tau-i\epsilon-|\xi|} - \frac{1}{\tau-i\epsilon+|\xi|}\right). \end{aligned}$$

Let us note that, by this formula and Lemma 9.7, we obtain an alternative derivation of the formula for $\mathcal{F}[E_+](t,\xi)$ that does not involve solving the second-order ODE in t. Indeed,

$$\begin{aligned} \mathcal{F}[E_+](t,\xi) &= \mathcal{F}_t^{-1}\mathcal{F}_{t,x}[E_+](t,\xi) \\ &= \lim_{\epsilon\to 0+}\frac{1}{2|\xi|}\mathcal{F}_t^{-1}\left[\left(\frac{1}{\tau-i\epsilon-|\xi|} - \frac{1}{\tau-i\epsilon+|\xi|}\right)\right](t) \\ &= \frac{i}{2|\xi|}\mathbf{1}_{(0,\infty)}(t)(e^{i|\xi|t} - e^{-i|\xi|t}) \\ &= -\mathbf{1}_{(0,\infty)}(t)\frac{\sin t|\xi|}{|\xi|}. \end{aligned}$$

The remaining space Fourier transform, in term, can be inverted following the procedure outlined in the preceding part.

Without going into the details, let us point out other fundamental solutions to the wave equation (which are all tempered distributions whose Fourier transform agrees with $\frac{1}{\tau^2 - |\xi|^2}$ outside the cone $\{(\tau, \xi) : \tau^2 = |\xi|^2\}$) that naturally arise in applications. For instance, the *backward fundamental solution* takes the form

$$\mathcal{F}_{t,x}[E_-] = \lim_{\epsilon \to 0+} \mathcal{F}_{t,x}[e^{\epsilon t} E_-] = \lim_{\epsilon \to 0+} \frac{1}{(\tau + i\epsilon)^2 - |\xi|^2}.$$

Another example is the following fundamental solution (called the *Feynman propagator*), which is of importance in quantum field theory:

$$\mathcal{F}_{t,x}[E_F] = \lim_{\epsilon \to 0+} \frac{1}{\tau^2 + i\epsilon - |\xi|^2}.$$

10. Energy method, Part I: A-priori estimates

The *energy method*, at the rudimentary level, is a way of proving *a-priori estimates* by multiplying the PDE by a suitable function (multiplier) and then integrating by parts. Here, an *a-priori estimate* refers to an estimate (which is synonymous with "inequality" in analysis) for a solution to the PDE that is *a-priori* assumed to exist.

Thanks to the simplicity and concreteness of the procedure, the energy method tends to be *robust*, i.e., the method often goes through even when the PDE has *variable coefficients*, or when it is *nonlinear*. *This point is a decisive advantage over the previous methods* (fundamental solution, Fourier transform)! On the other hand, a drawback of the energy method is that it is less clear what the a-priori estimates tell you about the features of the solution; nor is it immediately clear whether a-priori estimate have anything to do with the important question of the existence of a solution. However, these points will be remedied to a large extent by studying the *Sobolev spaces* in the next part of the course. Another difficulty with the energy method, which in practice is the more serious one, is that there is no general recipe for finding a good multiplier that leads to nice a-priori estimates for a given PDE.

Because of the last point, it is challenging (and probably counter-productive) to give a systematic and general description of this method. Instead, we will content ourselves here by seeing the method in action for model constant-coefficient second-order PDEs that we considered so far.

10.1. **Laplace equation.** Consider a solution u to the Dirichlet problem

$$\begin{cases} -\Delta u = f \text{ in } U, \\ \quad\ u = g \text{ on } \partial U. \end{cases} \tag{10.1}$$

The uniqueness of the solution to (10.1) (under suitable regularity conditions) was proved in Theorem 4.12 using the maximum principle. As an instance of the energy method, we will give an alternative proof of the uniqueness result (with minor differences in the regularity assumptions).

Proposition 10.1. *Let U be a bounded C^1 domain, $f \in C^0(\overline{U})$ and $g \in C^0(\partial U)$. The solution u to* (10.1) *with $u \in C^2(\overline{U})$ is unique.*

Proof. Let $u_1, u_2 \in C^2(\overline{U})$ be solutions to (10.1); then $v = u_2 - u_1$ belongs to $C^2(\overline{U})$ and solves

$$\begin{cases} -\Delta v = 0 & \text{in } U, \\ \quad\ v = 0 & \text{on } \partial U. \end{cases}$$

It remains to show that $v = 0$ in U.

Let us multiply the equation by v and integrate by parts (i.e., apply the divergence theorem) over U. We have

$$\begin{aligned} 0 &= \int_U -\Delta v v \,\mathrm{d}x \\ &= \int_U Dv \cdot Dv \,\mathrm{d}x - \int_{\partial U} \nu_{\partial U} \cdot Dv\, v \,\mathrm{d}S_{\partial U}. \end{aligned}$$

The boundary integral is zero thanks to the boundary condition $v = 0$ on ∂U. Thus, $\int_U |Dv|^2 \,\mathrm{d}x = 0$, which implies that v is a constant in U. Invoking $v = 0$ on ∂U again, it follows that $v = 0$ in U as desired. $\square$

10.2. **Heat equation.** Consider a solution u to the initial value problem

$$\begin{cases} \partial_t u - \Delta u = f & \text{in } (0,\infty)_t \times \mathbb{R}^d \\ u = g & \text{on } \{t = 0\}. \end{cases} \tag{10.2}$$

To simplify the notation, in this subsection we adopt the convention that D^α **only contains space derivative**.

As in the case of the Laplace equation, it turns out that it is again a good idea to multiply the equation by u and integrate by parts.

Proposition 10.2. *Let $f \in L^1_t((0,T); L^2(\mathbb{R}^d))$ and $g \in L^2(\mathbb{R}^d)$. The solution u to (10.2) with $u \in C_t([0,T]; L^2(\mathbb{R}^d))$ and $Du \in L^2((0,T) \times \mathbb{R}^d)$ is unique. Moreover, there exist $C > 0$ such that*

$$\begin{aligned} &\sup_{t \in [0,T]} \|u(t)\|_{L^2(\mathbb{R}^d)} + \|Du\|_{L^2((0,T)\times\mathbb{R}^d)} \\ &\leq C \left(\|g\|_{L^2(\mathbb{R}^d)} + \|f\|_{L^1((0,T);L^2(\mathbb{R}^d))} \right). \end{aligned} \tag{10.3}$$

Proof. The key part of the proof is nothing but multiplication of the equation by u and integrating by parts for a "sufficiently nice" solution u. Since this is the first time we see an argument of this sorts, we will provide more details on the approximation procedure, which allows us to deduce the general case from the computation for "sufficiently nice" solutions.

Let us begin by proving (10.3) under the additional assumption that $u \in C^\infty([0,T] \times \mathbb{R}^d)$ and for some $R > 0$, $\operatorname{supp} u(t) \subset B(0,R)$ for every $0 \leq t \leq T$. Multiplying the equation by u and integrating by parts on $(0,t) \times \mathbb{R}^d$, we obtain

$$\begin{aligned} \int_{t_0}^{t_1} \int f u \, \mathrm{d}x\mathrm{d}t &= \int_{t_0}^{t_1} \int (\partial_t u) u - (\Delta u) u \, \mathrm{d}x\mathrm{d}t \\ &= \frac{1}{2} \int_{t_0}^{t_1} \int \partial_t |u|^2 \, \mathrm{d}x\mathrm{d}t + \int_{t_0}^{t_1} \int |Du|^2 \, \mathrm{d}x\mathrm{d}t \\ &= \frac{1}{2} \|u(t_1)\|^2_{L^2(\mathbb{R}^d)} - \frac{1}{2} \|u(t_0)\|^2_{L^2(\mathbb{R}^d)} + \|Du\|^2_{L^2((t_0,t_1)\times\mathbb{R}^d)}, \end{aligned}$$

where we used the extra regularity assumption to justify all the manipulations, and no boundary terms arose in the second equality thanks to the extra support assumption on u. Rearranging terms, taking $t_0 \to 0+$ and taking the supremum in $t_1 \in [0,T]$, we obtain

$$\frac{1}{2} \sup_{t \in (0,T)} \|u(t)\|^2_{L^2(\mathbb{R}^d)} + \|Du\|^2_{L^2((0,T)\times\mathbb{R}^d)} \leq \frac{1}{2} \|u(0)\|^2_{L^2(\mathbb{R}^d)} + \int_0^T |fu| \, \mathrm{d}x\mathrm{d}t.$$

Applying Hölder's inequality and Young's inequality, the last term can be estimated as

$$\begin{aligned} \int_0^T |fu| \, \mathrm{d}x\mathrm{d}t &\leq \|f\|_{L^1((0,T);L^2(\mathbb{R}^d))} \sup_{t \in (0,T)} \|u(t)\|_{L^2(\mathbb{R}^d)} \\ &\leq \|f\|^2_{L^1((0,T);L^2(\mathbb{R}^d))} + \frac{1}{4} \sup_{t \in (0,T)} \|u(t)\|^2_{L^2(\mathbb{R}^d)}. \end{aligned}$$

Absorbing the last term into the term $\frac{1}{2} \sup_{t \in (0,T)} \|u(t)\|^2_{L^2(\mathbb{R}^d)}$ on the LHS, we arrive at (10.3).

In the general case, we approximate u by u_ϵ's satisfying the extra assumptions. For this purpose, let us introduce a mollifier $\varphi \in C^\infty_c(\mathbb{R}^{1+d})$ such that

$\int \varphi = 1$ and $\operatorname{supp} \varphi \subset (-1,1) \times B(0,1)$. For each $\epsilon > 0$, we define[19] $\varphi_\epsilon(t,x) = \epsilon^{-d-2}\varphi(\epsilon^{-2}t, \epsilon^{-1}x)$. Let us also introduce a smooth cutoff $\chi \in C^\infty(\mathbb{R}^d)$ such that $\chi = 1$ on $B(0, \frac{1}{4})$ and $\operatorname{supp} \chi \subset B(0, \frac{1}{2})$. On the subset $[\epsilon^2, T - \epsilon^2] \times \mathbb{R}^d$, we define

$$u_\epsilon = \varphi_\epsilon * (\chi(\epsilon x) u)|_{[\epsilon^2, T-\epsilon^2]\times\mathbb{R}^d} .$$

It is not difficult to see that, for $\epsilon \ll 1$, u_ϵ obeys the additional conditions on $[\epsilon^2, T - \epsilon^2] \times \mathbb{R}^d$ with $R = \epsilon^{-1}$. Moreover, for any compact interval $J \subseteq (0,T)$, as $\epsilon \to 0$, using the original assumptions $u \in C_t([0,T]; L^2)$ and $Du \in L^2((0,T) \times \mathbb{R}^d)$, it is possible to check that

$$\begin{aligned} u_\epsilon \to u \quad & \text{in } C_t(J; L^2(\mathbb{R}^d)), \\ Du_\epsilon \to Du \quad & \text{in } L^2(J \times \mathbb{R}^d), \\ u_\epsilon(\epsilon^2) \to g \quad & \text{in } L^2, \\ (\partial_t - \Delta)u_\epsilon \to f \quad & \text{in } L^1(J; L^2(\mathbb{R}^d)). \end{aligned}$$

Combined with (10.3) for u_ϵ on $[\epsilon^2, T - \epsilon^2] \times \mathbb{R}^d$, the desired estimate (10.3) in the general case follows.

Finally, the uniqueness assertion follows from the application of (10.3) to the difference of two solutions (in which case $f = g = 0$, so $u = 0$). □

Another idea that works well in conjunction with the energy method is to commute the equation with an operator Y, and apply the energy method to Yu to derive new a-priori estimates. In the case of a constant-coefficient operator $\mathcal{P}$, one good choice is $Y = D^\alpha$, since such an operator commutes with $\mathcal{P}$.

More concretely, in the case of the heat equation, note that $D^\alpha(\partial_t - \Delta) = (\partial_t - \Delta)D^\alpha$. If u solves (10.2), then $D^\alpha u$ solves

$$\begin{cases} (\partial_t - \Delta)D^\alpha u = D^\alpha f & \text{in } \mathbb{R}^{1+d}_+, \\ D^\alpha u = g & \text{on } \mathbb{R}^{1+d}_+ = \{t = 0\}. \end{cases}$$

(Recall our convention that D^α only consists of space derivatives!)

Applying Proposition 10.2 to $D^\alpha u$, we obtain a-priori estimates for higher-order derivatives of u.

Proposition 10.3. *Let $D^\alpha f \in L^1_t((0,T); L^2(\mathbb{R}^d))$ and $D^\alpha g \in L^2(\mathbb{R}^d)$ for all $|\alpha| \leq k$. Then the unique solution u to (10.2) with $u \in C_t([0,T]; L^2(\mathbb{R}^d))$ and $Du \in L^2_t((0,T) \times \mathbb{R}^d)$ also obeys $D^\alpha u \in C_t([0,T]; L^2(\mathbb{R}^d))$ and $DD^\alpha u \in L^2((0,T) \times \mathbb{R}^d)$. Moreover, there exist $C > 0$ depending only on k such that*

$$\begin{aligned} &\sum_{\alpha : |\alpha| \leq k} \left(\sup_{t \in [0,T]} \|D^\alpha u(t)\|_{L^2(\mathbb{R}^d)} + \|DD^\alpha u\|_{L^2((0,T)\times\mathbb{R}^d)} \right) \\ &\leq C \sum_{\alpha : |\alpha| \leq k} \left(\|D^\alpha g\|_{L^2(\mathbb{R}^d)} + \|D^\alpha f\|_{L^1((0,T); L^2(\mathbb{R}^d))} \right) . \end{aligned} \tag{10.4}$$

We omit the straightforward proof.

[19] Although not strictly necessary, we adopted the natural scaling for the heat equation, which is $(t,x) \mapsto (\lambda^{-2}t, \lambda^{-1}x)$.

10.3. **Schrödinger equation (optional).** Consider a solution u to the initial value problem

$$\begin{cases} i\partial_t u - \Delta u = f & \text{in } (0,\infty)_t \times \mathbb{R}^d \\ u = g & \text{on } \{t=0\}. \end{cases} \tag{10.5}$$

As in Section 10.3, in this subsection we again adopt the convention that D^α **only contains space derivative**.

Multiplying by $-i\overline{u}$, taking the real part and integrating by parts, we obtain the following result (which is analogous to Proposition 10.2).

Proposition 10.4. *Let $f \in L^1_t((0,T); L^2(\mathbb{R}^d))$ and $g \in L^2(\mathbb{R}^d)$. The solution u to (10.5) with $u \in C_t([0,T]; L^2(\mathbb{R}^d))$ is unique. Moreover, there exist $C > 0$ such that*

$$\sup_{t\in[0,T]} \|u(t)\|_{L^2(\mathbb{R}^d)} \leq C \left(\|g\|_{L^2(\mathbb{R}^d)} + \|f\|_{L^1((0,T);L^2(\mathbb{R}^d))} \right). \tag{10.6}$$

Proof. Since the proof is very similar to Proposition 10.2, we will only present the formal integration by parts argument (i.e., assuming that u is sufficiently regular to make sense of each manipulation, and also that $u(t)$ is compactly supported for each t so that no boundary terms arise). Multiplying the equation by $-i\overline{u}$, taking the real part and integrating by parts on $(t_0, t_1) \times \mathbb{R}^d$, we obtain

$$\begin{aligned} \int_{t_0}^{t_1} \int \mathrm{Re}(-if\overline{u}) \,\mathrm{d}x\mathrm{d}t &= \int_{t_0}^{t_1} \int \mathrm{Re}(\partial_t u \overline{u}) + \mathrm{Re}(-\Delta u(-i)\overline{u}) \,\mathrm{d}x\mathrm{d}t \\ &= \int_{t_0}^{t_1} \int \frac{1}{2}\partial_t |u|^2 - \mathrm{Im}(\Delta u \overline{u}) \,\mathrm{d}x\mathrm{d}t \\ &= \int_{t_0}^{t_1} \int \frac{1}{2}\partial_t |u|^2 + \mathrm{Im}(Du \cdot \overline{Du}) \,\mathrm{d}x\mathrm{d}t \\ &= \frac{1}{2}\|u(t_1)\|^2_{L^2(\mathbb{R}^d)} - \frac{1}{2}\|u(t_0)\|^2_{L^2(\mathbb{R}^d)}. \end{aligned}$$

Using this identity and proceeding similarly as in Proposition 10.2, the present proposition follows. □

Commuting the equation with D^α, and then applying Proposition 10.4, we obtain the following a-priori estimates for higher-order derivatives of u.

Proposition 10.5. *Let $D^\alpha f \in L^1_t((0,T); L^2(\mathbb{R}^d))$ and $D^\alpha g \in L^2(\mathbb{R}^d)$ for all $|\alpha| \leq k$. Then the unique solution u to (10.5) with $u \in C_t([0,T]; L^2(\mathbb{R}^d))$ also obeys $D^\alpha u \in C_t([0,T]; L^2(\mathbb{R}^d))$. Moreover, there exist $C > 0$ depending only on k such that*

$$\sum_{\alpha:|\alpha|\leq k} \sup_{t\in[0,T]} \|D^\alpha u(t)\|_{L^2(\mathbb{R}^d)} \leq C \sum_{\alpha:|\alpha|\leq k} \left(\|D^\alpha g\|_{L^2(\mathbb{R}^d)} + \|D^\alpha f\|_{L^1((0,T);L^2(\mathbb{R}^d))} \right). \tag{10.7}$$

10.4. **Wave equation.** Consider a solution ϕ to the initial value problem

$$\begin{cases} \Box\phi = f & \text{in } (0,\infty)_t \times \mathbb{R}^d \\ (\phi, \partial_t\phi) = (g,h) & \text{on } \{t=0\}, \end{cases} \tag{10.8}$$

where we remind the reader that $\Box\phi = (-\partial_t^2 + \Delta)\phi$.

In this case, it will be useful to start with the following divergence identity, which is called the *local conservation of energy*.

Proposition 10.6. *Let ϕ be a smooth solution to $\Box\phi = f$. We have*

$$\partial_t \left(\frac{1}{2}(\partial_t \phi)^2 + \frac{1}{2}\sum_j (\partial_j \phi)^2 \right) + \sum_j \partial_j(\partial_j \phi \partial_t \phi) = -f\partial_t \phi. \tag{10.9}$$

Proof. The idea is to multiply the equation by $\partial_t\phi$ and try to mimic the proof of integration by parts (some people call this "differentiation by parts"). The goal is to ensure that there are no terms with two derivatives, except those in the divergence of some expression.

We begin by writing

$$\begin{aligned} f\partial_t\phi &= (-\partial_t^2\phi + \Delta\phi)\partial_t\phi \\ &= -\partial_t^2\phi\partial_t\phi + \sum_j \partial_j^2\phi\partial_t\phi \end{aligned}$$

The first term can be rewritten as a divergence, i.e., $-\partial_t^2\phi\partial_t\phi = -\partial_t(\frac{1}{2}(\partial_t\phi)^2)$. For the second term, we first move ∂_j to $\partial_t\phi$ up to a divergence, and then proceed like the first term:

$$\begin{aligned} \sum_j \partial_j^2\phi\partial_t\phi &= \sum_j \partial_j(\partial_j\phi\partial_t\phi) - \partial_j\phi\partial_t\partial_j\phi \\ &= \sum_j \partial_j(\partial_j\phi\partial_t\phi) - \partial_t\left(\frac{1}{2}\sum_j(\partial_j\phi)^2\right). \end{aligned}$$

In conclusion,

$$-f\partial_t\phi = \partial_t\left(\frac{1}{2}(\partial_t\phi)^2 + \frac{1}{2}\sum_j(\partial_j\phi)^2\right) + \sum_j \partial_j(\partial_j\phi\partial_t\phi),$$

which is precisely the desired divergence identity. □

By integrating (10.9) on spacetime slabs of the form $(t_0, t_1) \times \mathbb{R}^d$ and applying the divergence theorem, we easily obtain the following analogue of Propositions 10.2 and 10.4.

Proposition 10.7. *Let $f \in L_t^1((0,T); L^2(\mathbb{R}^d))$ and $Dg, h \in L^2(\mathbb{R}^d)$. The solution ϕ to* (10.8) *with $D\phi, \partial_t\phi \in C_t([0,T]; L^2(\mathbb{R}^d))$ is unique. Moreover, there exist $C > 0$ such that*

$$\begin{aligned} &\sup_{t\in[0,T]} \left(\|Du(t)\|_{L^2(\mathbb{R}^d)} + \|\partial_t u(t)\|_{L^2(\mathbb{R}^d)}\right) \\ &\leq C\left(\|Dg\|_{L^2(\mathbb{R}^d)} + \|h\|_{L^2(\mathbb{R}^d)} + \|f\|_{L^1((0,T);L^2(\mathbb{R}^d))}\right). \end{aligned} \tag{10.10}$$

Proof. Since the proof is very similar to Propositions 10.2 and 10.4, we will only present the formal argument. Assume that ϕ is smooth and $u(t,\cdot)$ is compactly supported for each t. Starting from (10.9), integrating over $(0,t)\times\mathbb{R}^d$ and applying the divergence theorem, we obtain

$$\int_{\{t\}\times\mathbb{R}^d} \left(\frac{1}{2}|\partial_t\phi|^2 + \frac{1}{2}\sum_j |\partial_j\phi|^2\right) \mathrm{d}x$$

$$= \int_{\{0\}\times\mathbb{R}^d} \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2}\sum_j |\partial_j \phi|^2 \right) \mathrm{d}x + \int_0^t \int_{\mathbb{R}^d} (-f)\partial_t \phi \, \mathrm{d}x\mathrm{d}t'.$$

where no spatial boundary terms arise thanks to the support assumption. The desired inequality now follows by a Cauchy–Schwarz argument as in Proposition 10.2. Moreover, again as in Proposition 10.2, the general case follows by an approximation argument. □

Commuting the equation with D^α, and then applying Proposition 10.4, we obtain the following a-priori estimates for higher-order derivatives of u.

Proposition 10.8. *Let $D^\alpha f \in L^1_t((0,T); L^2(\mathbb{R}^d))$ and $DD^\alpha g, D^\alpha h \in L^2(\mathbb{R}^d)$ for all $|\alpha| \leq k$. Then the unique solution ϕ to* (10.8) *with $D\phi, \partial_t\phi \in C_t([0,T]; L^2(\mathbb{R}^d))$ also obeys $DD^\alpha\phi, D^\alpha\partial_t\phi \in C_t([0,T]; L^2(\mathbb{R}^d))$. Moreover, there exist $C > 0$ depending only on k such that*

$$\begin{aligned} &\sum_{\alpha:|\alpha|\leq k} \sup_{t\in[0,T]} \left(\|DD^\alpha u(t)\|_{L^2(\mathbb{R}^d)} + \|\partial_t D^\alpha u(t)\|_{L^2(\mathbb{R}^d)} \right) \\ &\leq C \sum_{\alpha:|\alpha|\leq k} \left(\|DD^\alpha g\|_{L^2(\mathbb{R}^d)} + \|D^\alpha h\|_{L^2(\mathbb{R}^d)} + \|D^\alpha f\|_{L^1((0,T);L^2(\mathbb{R}^d))} \right). \end{aligned} \tag{10.11}$$

Finally, the local conservation of energy (10.9) leads to a simple proof the *finite speed of propagation* for the wave equation. To formulate this result, given $r > 0$ and $x_0 \in \mathbb{R}^d$, we introduce the notation

$$\mathcal{D}(B_r(x)) = \{(t,x) \in \mathbb{R}^{1+d} : t > 0, \, |x - x_0| < r - t\}.$$

The region $\mathcal{D}(B_r(x))$ is called the domain of dependence of the initial ball $\{0\} \times B_r(x_0)$. Geometrically, it is the cone with aperture 90° (or half angle 45°) and base $B_r(x_0)$. Finite speed of propagation says, roughly, that a solution ϕ to (10.8) is uniquely determined in $\mathcal{D}(B_r(x_0))$ by the initial data on $B_r(x_0)$ and f in $\mathcal{D}(B_r(x_0))$; see also Proposition 10.9 below. This is a rigorous formulation of the property that the data at (s,y) cannot affect $\phi(t,x)$ unless (t,x) can be reached from (s,y) by a ray of speed ≤ 1.

For simplicity, we take ϕ to be very regular in the following result, but we remark that it can be easily improved if desired.

Proposition 10.9 (Finite speed of propagation)**.** *Consider ϕ in $C^2(\mathcal{D}(B_r(x_0))) \cap C^1(\overline{\mathcal{D}(B_r(x_0))})$. Let ϕ solve the homogeneous wave equation $\Box\phi = 0$ in $\mathcal{D}(B_r(x_0))$ and obey $(\phi, \partial_t\phi) = (0,0)$ on $\{0\} \times B_r(x_0)$. Then $\phi = 0$ in $\mathcal{D}(B_r(x_0))$.*

Proof. Without loss of generality, we may set $x_0 = 0$. For this proof, we introduce the following pieces of notation:

$$\begin{aligned} S_{t_0} &= \mathcal{D}(B_r(0)) \cap \{t = t_0\} = \{t_0\} \times B_{r-t_0}(0), \\ C_{t_0,t_1} &= \mathcal{D}(B_r(0)) \cap \{t_0 < t < t_1\} = \{(t,x) : t_0 < t < t_1, \, |x| < r - t\}, \\ \partial_{\mathrm{lat}} C_{t_0,t_1} &= \{(t,x) : t_0 < t < t_1, \, |x| = r - t\}. \end{aligned}$$

Here, ∂_{lat} refers to the lateral boundary of the cone. Fix $t_1 < r$. Integrating (10.9) over C_{0,t_1} and using the divergence theorem, we obtain

$$\int_{S_{t_1}} \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2}\sum_j |\partial_j \phi|^2 \right) \mathrm{d}x = \int_{S_0} \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2}\sum_j |\partial_j \phi|^2 \right) \mathrm{d}x$$

$$-\int_{\partial_{\mathrm{lat}} C_{0,t_1}} \boldsymbol{\nu} \cdot \mathbf{e} \, \mathrm{d}\sigma,$$

where $\boldsymbol{\nu}$ is the outward unit normal to C_{0,t_1} and

$$\mathbf{e} = \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2} \sum_j |\partial_j \phi|^2, \partial_1 \phi \partial_t \phi, \dots, \partial_d \phi \partial_t \phi \right),$$

so that (10.9) is equivalent to $\mathrm{div}\mathbf{e} = 0$. By elementary geometry, we have

$$\boldsymbol{\nu} = \frac{1}{\sqrt{2}} \left(1, \frac{x^1}{|x|}, \dots, \frac{x^d}{|x|} \right).$$

Hence, the integral on $\partial_{\mathrm{lat}} C_{0,t_1}$ takes the form

$$\int_{\partial_{\mathrm{lat}} C_{0,t_1}} \boldsymbol{\nu} \cdot \mathbf{e} \, \mathrm{d}\sigma = \frac{1}{\sqrt{2}} \int_{\partial_{\mathrm{lat}} C_{0,t_1}} \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2} \sum_j |\partial_j \phi|^2 \right) + \frac{x}{|x|} \cdot D\phi \partial_t \phi \, \mathrm{d}\sigma.$$

By the inequality $2ab \leq a^2 + b^2$ and Cauchy–Schwarz, we have

$$\left| \frac{x}{|x|} \cdot D\phi \partial_t \phi \right| \leq \frac{1}{2} \left(|\frac{x}{|x|} \cdot D\phi|^2 + |\partial_t \phi|^2 \right) \leq \frac{1}{2} \left(|D\phi|^2 + |\partial_t \phi|^2 \right).$$

Therefore, it follows that

$$\int_{\partial_{\mathrm{lat}} C_{0,t_1}} \boldsymbol{\nu} \cdot \mathbf{e} \, \mathrm{d}\sigma \geq 0,$$

and by the integrated energy identity and our assumptions on $S_0 = \{t = 0\} \times B_r(0)$,

$$\int_{S_{t_1}} \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2} \sum_j |\partial_j \phi|^2 \right) \mathrm{d}x \leq \int_{S_0} \left(\frac{1}{2}|\partial_t \phi|^2 + \frac{1}{2} \sum_j |\partial_j \phi|^2 \right) \mathrm{d}x = 0.$$

Since this inequality holds for every $0 < t_1 < r$, it follows that ϕ is constant in $\mathcal{D}(B_r(0))$, which must be zero by the assumption $\phi = 0$ on S_0. □

11. Sobolev spaces

Recall that the *energy method* typically gives an a-priori estimate for the solution of a PDE of the form

$$\sum_{\alpha:|\alpha|\leq k} \|D^\alpha u\|_{L^p(U)} \leq C(\text{Data}). \tag{11.1}$$

Given such estimates, a natural question is: *Can we convert the control of such a form into the control of other norms of* u? For instance, *can we control the pointwise values of* u *through such a control?*

The answer is clearly yes in the one-dimensional case. For instance, if u is a smooth function supported in a compact interval I, by the fundamental theorem of calculus and Hölder's inequality

$$\sup_{x\in I}|u(x)| \leq \sup_{x\in I}|\int_{-\infty}^{x} u'(x')\,\mathrm{d}x'| \leq \|u'(x)\|_{L^1(I)} \leq |I|^{\frac{p-1}{p}}\|u'\|_{L^p(I)}.$$

By the same method, we may even obtain a control of the modulus of continuity: for $x > y$, we have

$$|u(x)-u(y)| \leq \int_{x}^{y}|u'(x')|\,\mathrm{d}x' \leq |x-y|^{\frac{p-1}{p}}\|u'\|_{L^p(I)}.$$

The multi-dimensional generalization of the above inequalities are called *Sobolev inequalities*. These will be one of the main topics that we cover in this part (Section 11.5). The most natural setting for such inequalities is the (vector) space of functions u such that the LHS of (11.1) is finite, equipped with the norm given (essentially) by the LHS of (11.1); this space is called the *Sobolev space* with regularity index k and integrability index p.

Another motivation for studying the Sobolev spaces that *they provide a nice infinite-dimensional vector space (i.e., functional-analytic) framework that allows us to convert a-priori estimates for (hypothetical) solutions*, of which energy estimates are key examples, *to statements about the existence of such solutions*. Roughly speaking, the story is as follows. In finite-dimensional linear algebra, we know that the image of a linear operator P (i.e., existence of u such that $Pu = f$) is closely related to kernel of the adjoint operator P' (i.e., the degree of failure of uniqueness of $P'\phi = 0$), i.e., $\mathrm{im}P = (\ker P')^{\perp}$. If we are able to extend this idea to the setting of a linear partial differential operator $\mathcal{P}$ between suitable function spaces (i.e., infinite-dimensional vector spaces), then we would be able to characterize the f's for which there exist a u such that for $\mathcal{P}u = f$ by characterizing $\ker\mathcal{P}'$, which is the set of all solutions to $\mathcal{P}'u = 0$. This is roughly how a-priori estimates for solutions to a PDE problem (formalized as $\mathcal{P}'u = 0$) leads to existence results[20]. The relevant tools and concepts are *Rellich–Kondrachov compactness theorem* (Section 11.6) and *characterization of dual Sobolev spaces* (Section 11.8).

11.1. **Definitions and basic properties.**

Definition 11.1 (Sobolev spaces)**.** Let k be a nonnegative integer and $1 \leq p \leq \infty$. We define the *Sobolev space with regularity index* k *and integrability index* p by

$$W^{k,p}(U) = \{u \in \mathcal{D}'(U) : D^\alpha u \in L^p(U) \text{ for all } \alpha,\ |\alpha| \leq k\}.$$

[20]Note that we used the other side of this idea to motivate the derivation of representation formulae (which are expressions of uniqueness) from a fundamental solution (which is, at first, motivated by the existence problem)!

We equip the space $W^{k,p}(U)$ with the norm[21]

$$\|u\|_{W^{k,p}(U)} := \begin{cases} \left(\sum_{\alpha:|\alpha|\leq k} \|D^\alpha u\|_{L^p}^p\right)^{\frac{1}{p}} & \text{when } 1 \leq p < \infty, \\ \sum_{\alpha:|\alpha|\leq k} \|D^\alpha u\|_{L^\infty} & \text{when } p = \infty. \end{cases}$$

(Indeed, it is easy to check that the RHS defines a norm.)

As usual for a normed vector space, we will say that $u_j \to u$ in $W^{k,p}(U)$ as $j \to \infty$ if

$$\|u_j - u\|_{W^{k,p}(U)} \to 0 \quad \text{as } j \to \infty.$$

Note that any $u \in C_c^\infty(U)$ is clearly an element of $W^{k,p}(U)$. We define $W_0^{k,p}(U)$ to be the closure of $C_c^\infty(U)$ in $W^{k,p}(U)$, i.e.,

$$W_0^{k,p}(U) := \{u \in W^{k,p}(U) : \exists u_j \in C_c^\infty(U) \text{ s.t. } u_j \to u \text{ in } W^{k,p}(U) \text{ as } j \to \infty\}.$$

The space $W_0^{k,p}(U)$ should be understood as a closed subspace of $W^{k,p}(U)$ that consist of functions whose "boundary values on ∂U vanish up to all relevant orders".

When $p = 2$, $\|\cdot\|_{W^{k,2}(U)}$ is derived from an inner product in the sense that

$$\|u\|_{W^{k,2}(U)}^2 = \langle u, u\rangle_{W^{k,2}(U)}, \quad \langle u, v\rangle_{W^{k,2}} := \sum_{\alpha:|\alpha|\leq k} \int_U D^\alpha u \cdot D^\alpha v \,\mathrm{d}x.$$

As we will see soon, $W^{k,2}(U)$ will be a Hilbert space with respect to $\langle\cdot,\cdot\rangle_{W^{k,2}(U)}$. Accordingly, we will use the notation

$$H^k(U) := W^{k,2}(U), \quad H_0^k(U) := W_0^{k,2}(U), \quad \langle\cdot,\cdot\rangle_{H^k(U)} := \langle\cdot,\cdot\rangle_{W^{k,2}(U)}.$$

Some basic properties of the Sobolev spaces are as follows.

Proposition 11.2. *Let k be a nonnegative integer, $1 \leq p \leq \infty$ and U a domain in $\mathbb{R}^d$.*

(1) The normed space $(W^{k,p}(U), \|\cdot\|_{W^{k,p}(U)})$ is complete, i.e., it is a Banach space.

(2) The inner product space $(H^k(U), \langle\cdot,\cdot\rangle_{H^k(U)})$ is complete, i.e., it a Hilbert space.

(3) A function u belongs to $H^k(\mathbb{R}^d)$ if and only if $\|(1+|\xi|^2)^{\frac{k}{2}}\widehat{u}(\xi)\|_{L^2} \in L^2(\mathbb{R}^d)$. Moreover, there exists a constant C, depending only on the dimension d and k, such that for any $u \in H^k(\mathbb{R}^d)$,

$$C^{-1}\|u\|_{H^k(\mathbb{R}^d)} \leq \|(1+|\xi|^2)^{\frac{k}{2}}\widehat{u}(\xi)\|_{L^2(\mathbb{R}^d)} \leq C\|u\|_{H^k(\mathbb{R}^d)}.$$

Parts (1) and (2) are easy consequences of the completeness of L^p; see [Eva10, Section 5.2, Theorem 2] for a detailed proof. Part (3) follows from basic properties of the Fourier transform; see [Eva10, Section 5.8, Theorem 8].

11.2. **Approximation results.** A general element u of $W^{k,p}(U)$ is a fairly abstract object, which is cumbersome to work with directly. In this subsection, we will discuss a number of results that allows us to approximate u by smooth functions.

One basic tool for proving approximation results is the idea of *mollifiers*, which we already encountered in the context of distribution theory. The gist of the mollifier method was as follows: Let φ be a smooth compactly supported function on $\mathbb{R}^d$ such that $\int \varphi = 1$. For each $\epsilon > 0$, define $\varphi_\epsilon(x) := \epsilon^{-d}\varphi(\epsilon^{-1}x)$ (which are called *mollifiers*). Then for any $u \in \mathcal{D}'(\mathbb{R}^d)$, the family $\{\varphi_\epsilon * u\}$ provides an approximation

[21]As usual, the sum $\sum_{\alpha:|\alpha|\leq k}$ includes $\alpha = 0$, where $|0| = 0$ and $D^0 u = u$.

of u by smooth functions, in the sense that $\varphi_\epsilon * u \in C^\infty(\mathbb{R}^d)$ for each $\epsilon > 0$ and $\varphi_\epsilon * u \to u$ in $\mathcal{D}'(\mathbb{R}^d)$ as $\epsilon \to 0$.

Let us now prove that $\varphi_\epsilon * u$ is also a good approximation of u for $u \in W^{k,p}(\mathbb{R}^d)$. The essential analytic fact we need is as follows. For $y \in \mathbb{R}^d$ and any $u \in L^1_{loc}(\mathbb{R}^d)$, define the *translation (by y) operator*[22]

$$\tau_y u(x) := u(x-y).$$

Lemma 11.3. *For any $1 \leq p < \infty$, the mapping $y \mapsto \tau_y$ is continuous as a linear map on $L^p(\mathbb{R}^d)$; equivalently, for any $u \in L^p(\mathbb{R}^d)$,*

$$\|\tau_y u - u\|_{L^p(\mathbb{R}^d)} \to 0 \quad \text{as } y \to 0. \tag{11.2}$$

Proof. To show the continuity of the mapping $y \mapsto \tau_y$, it suffices to verify its continuity at 0 thanks to the property $\tau_y \tau_z = \tau_z \tau_y = \tau_{y+z}$. In other words, for each fixed $\epsilon > 0$, we wish to show the existence of $\delta > 0$ such that $\|\tau_y u - u\|_{L^p(\mathbb{R}^d)} < \epsilon$ for $|y| < \delta$. Using the basic fact that $C_0(\mathbb{R}^d)$ (i.e., the space of continuous, compactly supported functions) is dense in $L^p(\mathbb{R}^d)$ for $1 \leq p < \infty$, we can find $v \in C_0(\mathbb{R}^d)$ such that $\|v - u\|_{L^p(\mathbb{R}^d)} < \frac{\epsilon}{3}$. Moreover, since v is uniformly continuous (since $\operatorname{supp} v$ is compact), we can find $\delta > 0$ such that $|v(x-y) - v(x)| < \frac{\epsilon}{3}(1 + |\operatorname{supp} v|)^{-\frac{1}{p}}$ for all $|y| < \delta$, which implies

$$\|\tau_y v - v\|_{L^p(\mathbb{R}^d)} < \frac{\epsilon}{3} \quad \text{for all } |y| < \delta.$$

In conclusion, for $|y| < \delta$,

$$\begin{aligned}
\|\tau_y u - u\|_{L^p(\mathbb{R}^d)} &\leq \|\tau_y u - \tau_y v\|_{L^p(\mathbb{R}^d)} + \|\tau_y v - v\|_{L^p(\mathbb{R}^d)} + \|v - u\|_{L^p(\mathbb{R}^d)} \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,
\end{aligned}$$

where on the last line, we used the preceding bounds as well as the simple fact that $\|\tau_y f\|_{L^p(\mathbb{R}^d)} = \|f\|_{L^p(\mathbb{R}^d)}$. □

We have the following quick corollary.

Corollary 11.4. *If $u \in L^p(\mathbb{R}^d)$, then $\varphi_\epsilon * u \to u$ in $L^p(\mathbb{R}^d)$.*

Proof. We write

$$\begin{aligned}
\|\varphi_\epsilon * u(x) - u(x)\|_{L^p(\mathbb{R}^d)} &= \|\int \varphi_\epsilon(y) u(x-y) \, \mathrm{d}y - u(x)\|_{L^p(\mathbb{R}^d)} \\
&= \|\int \varphi(z) u(x - \epsilon z) \, \mathrm{d}z - \int \varphi(z) u(x) \, \mathrm{d}z\|_{L^p(\mathbb{R}^d)} \\
&\leq \int |\varphi(z)| \|\tau_{\epsilon z} u(x) - u(x)\|_{L^p(\mathbb{R}^d)} \, \mathrm{d}z,
\end{aligned}$$

where on the second line, we used the property that $\int \varphi(z) \, \mathrm{d}z = 1$, and on the last line, we used the Minkowski inequality. Then by Lemma 11.3 and the dominated convergence theorem, the last line goes to zero as $\epsilon \to 0$, as desired. □

Moreover, using the property $D^\alpha(\varphi_\epsilon * u) = \varphi_\epsilon * D^\alpha u$, we obtain the following smooth approximation result for $u \in W^{k,p}(\mathbb{R}^d)$; we omit the obvious proof.

---

[22]In our previous discussion of distribution theory, we already implicitly used this operator. It is possible to formally define this operator on $\mathcal{D}'(\mathbb{R}^d)$ via the adjoint method, i.e., $\langle \tau_y u, \phi \rangle := \langle u, \tau_{-y} \phi \rangle$.

Proposition 11.5. *Let k be a nonnegative integer and $1 \leq p < \infty$. If $u \in W^{k,p}(\mathbb{R}^d)$, then $\varphi_\epsilon * u \to u$ in $W^{k,p}(\mathbb{R}^d)$. In particular, $C^\infty(\mathbb{R}^d)$ is dense in $W^{k,p}(\mathbb{R}^d)$.*

Our next goal is to prove a similar approximation result for $u \in W^{k,p}(U)$ when U is a general domain (see Proposition 11.9 below). To reduce this problem to the case $U = \mathbb{R}^d$ that we already handled, we use the idea of a *smooth partition of unity*:

Definition 11.6 (Smooth partition of unity)**.** Let U be a (not necessarily open) subspace of $\mathbb{R}^d$. A collection $\{V_j\}_{j \in J}$ (indexed by $j \in J$) of open sets $V_j \subseteq U$ (with respect to the subspace topology) is called an *open covering* if $\cup_{j \in J} V_j = U$. A collection $\{\chi_j\}_{j \in J}$ of functions is called a *smooth partition of unity subordinate to* $\{V_j\}$ if the following properties hold:

- each χ_j is smooth;
- $\operatorname{supp} \chi_j \subset V_j$;
- for each $x \in U$, $0 \leq \chi_j(x) \leq 1$;
- for each $x \in U$, $\sum_{j \in J} \chi_j(x) = 1$, where at most finitely many summands are non-zero.

The basic existence result for smooth partitions of unity is as follows:

Lemma 11.7. *Let U be a nonempty subspace in $\mathbb{R}^d$, and let $\{V_j\}_{j \in J}$ be an open covering of U. Then there exists a smooth partition of unity $\{\chi_j\}_{j \in J}$ subordinate to $\{V_j\}_{j \in J}$.*

The deepest part of the proof is a result from point-set topology that, since U is a subspace of a metric space $\mathbb{R}^d$, there always exists a continuous partition of unity $\{\chi_j\}_{j \in J}$ subordinate to any open covering $\{V_j\}_{j \in J}$ (in general, it is a consequence of the fact that any metric space is Hausdorff and paracompact, although more direct constructions exist in the case of $U \subseteq \mathbb{R}^d$). Afterwards, it is a matter of performing a tedious but straightforward mollification procedure to construct a *smooth* partition of unity subordinate to $\{V_j\}_{j \in J}$; we will not go into the details.

Remark 11.8. (1) By the chain rule, it is not difficult to show that

$$\|\chi_j u\|_{W^{k,p}(V_j)} \leq C \|u\|_{W^{k,p}(U)},$$

where C depends on d, k, p and $\sup_{x \in V_j} |D^\alpha \chi_j|$ for $|\alpha| \leq k$. On the other hand, by the triangle inequality,

$$\|u\|_{W^{k,p}(U)} \leq \sum_{j \in J} \|\chi_j u\|_{W^{k,p}(V_j)}.$$

(2) When V_j is an open set in $\mathbb{R}^d$, then $\chi_j u$ extends in an obvious way to a function on the whole space $\mathbb{R}^d$ by defining $\chi_j u(x) = 0$ for $x \notin V_j$. Moreover, $\|\chi_j u\|_{W^{k,p}(\mathbb{R}^d)} = \|\chi_j u\|_{W^{k,p}(U)}$.

Proposition 11.9. *Let k be a nonnegative integer and $1 \leq p < \infty$. Let U be any domain in $\mathbb{R}^d$. If $u \in W^{k,p}(U)$, then there exists a sequence $u_j \in C^\infty(U)$ such that $u_j \to u$ in $W^{k,p}(\mathbb{R}^d)$. In other words, $C^\infty(U)$ is dense in $W^{k,p}(U)$.*

Proof. Let $u \in W^{k,p}(U)$, and let $\epsilon > 0$. We want to find $u_\epsilon \in C^\infty(U)$ such that $\|u_\epsilon - u\|_{W^{k,p}(U)} < \epsilon$.

Consider an open cover $\{V_j\}_{j=1,2,\dots}$ of U defined by

$$V_j = U_{j+3} \setminus \overline{U_{j+1}}, \quad U_j := \{x \in U : \mathrm{dist}(x, \partial U) > 1/j\}.$$

Let χ_j be a smooth partition of unity subordinate to $\{V_j\}$. We write

$$u = \sum_{j=1}^{\infty} \chi_j u.$$

By Remark 11.8.(2), we may view each $\chi_j u$ as an element in $W^{k,p}(U)$. Fix $\varphi_0 \in C^\infty(\mathbb{R}^d)$ such that $\operatorname{supp} \varphi_0 \subset B(0,1)$ and $\int \varphi_0 = 1$. For each j, choose $\epsilon_j > 0$ small enough so that

$$\|\varphi_{\epsilon_j} * \chi_j u - \chi_j u\|_{W^{k,p}} < 2^{-j}\epsilon, \quad \operatorname{supp}(\varphi_{\epsilon_j} * \chi_j u) \subseteq W_j := U_{j+4} \setminus \overline{U_j}.$$

For the first property, we used Proposition 11.5.

The second property implies that

$$u_\epsilon := \sum_{j=1}^{\infty} \varphi_{\epsilon_j} * \chi_j u$$

belongs to $C^\infty(U)$, since in each ball $B(x,r) \subset \overline{B(x,r)} \subset U$, there are at most finitely many nonzero terms in the sum $\sum_{j=1}^{\infty} \varphi_{\epsilon_j} * \chi_j u$. Finally, by the first property,

$$\|u_\epsilon - u\|_{W^{k,p}(U)} \leq \sum_{j=1}^{\infty} \|\varphi_{\epsilon_j} * \chi_j u - \chi_j u\|_{W^{k,p}(U)} \leq \sum_{j=1}^{\infty} 2^{-j}\epsilon = \epsilon,$$

as desired. □

Next, we ask the question of whether a general element $u \in W^{k,p}(U)$ can be approximated by functions that are smooth *up to the boundary* of U (or equivalently, which are restrictions to U of smooth functions on $\mathbb{R}^d$). For this purpose, we need to require some regularity on the boundary of U in order to rule out pathological behaviors.

Proposition 11.10. *Let k be a nonnegative integer and $1 \leq p < \infty$. Let U be a C^1 domain in $\mathbb{R}^d$. If $u \in W^{k,p}(U)$, then there exists a sequence $u_j \in C^\infty(\overline{U})$ such that $u_j \to u$ in $W^{k,p}(\mathbb{R}^d)$. In other words, $C^\infty(\overline{U})$ is dense in $W^{k,p}(U)$.*

The basic tools for proving this result are again mollifiers and a smooth partition of unity. For the proof, see [Eva10, Section 5.3, Theorem 3].

So far we were concerned with approximation of an element of $u \in W^{k,p}(U)$ by smooth functions. Our last approximation result concerns approximation of an element $u \in W^{k,p}(\mathbb{R}^d)$ by compactly supported functions. Let $\chi(x)$ be a smooth compactly supported function on $\mathbb{R}^d$ such that $\chi(0) = 1$.

Proposition 11.11. *Let k be a nonnegative integer and $1 \leq p < \infty$. If $u \in W^{k,p}(\mathbb{R}^d)$, then $\chi(R^{-1}x)u \to u$ in $W^{k,p}(\mathbb{R}^d)$ as $R \to \infty$.*

Proof. Let $u \in W^{k,p}(\mathbb{R}^d)$. For each α such that $|\alpha| \leq k$, we have

$$\|D^\alpha \left(\chi(R^{-1}x)u - u\right)\|_{L^p(\mathbb{R}^d)} \leq \|(\chi(R^{-1}x) - 1)D^\alpha u\|_{L^p(\mathbb{R}^d)} + C \sum_{\beta,\gamma:\beta+\gamma=\alpha, |\beta|\geq 1} R^{-|\beta|} \|D^\beta \chi\|_{L^\infty(\mathbb{R}^d)} \|D^\gamma u\|_{L^p(\mathbb{R}^d)}$$

As $R \to \infty$, the first term goes to zero by the dominated convergence theorem; the last term vanishes since $|\beta| \geq 1$. □

Combining Propositions 11.5 and 11.11, we immediately obtain the following result:

Corollary 11.12. *Let k be a nonnegative integer and $1 \leq p < \infty$. Then $C^\infty_c(\mathbb{R}^d)$ is dense in $W^{k,p}(\mathbb{R}^d)$. In short, $W^{k,p}_0(\mathbb{R}^d) = W^{k,p}(\mathbb{R}^d)$.*

We remark that this result necessarily fails for any C^1 domain U other than $\mathbb{R}^d$.

11.3. **Extensions.** Next, we seek for ways to extend an element in $W^{k,p}(U)$ to a function in $W^{k,p}(\mathbb{R}^d)$.

Proposition 11.13. *Let k be a nonnegative integer and $1 \leq p < \infty$. Let U be a C^k domain in $\mathbb{R}^d$ and let V be a domain in $\mathbb{R}^d$ such that $\overline{U} \subset V$. Then there exists a linear mapping $\mathcal{E} : W^{k,p}(U) \to W^{k,p}(\mathbb{R}^d)$ with the following properties:*

(1) $\mathcal{E}$ is bounded, i.e., there exists $C > 0$ that depends only on k, p, U and V such that $\|\mathcal{E}[u]\|_{W^{k,p}(\mathbb{R}^d)} \leq C\|u\|_{W^{k,p}(U)}$;
(2) $\mathcal{E}[u]|_U = u$;
(3) $\operatorname{supp}\mathcal{E}[u] \subset V$.

Proof. We begin by noting that it suffices construct the operator $\mathcal{E}$ for $u \in C^\infty(\overline{U})$. Indeed, by Proposition 11.10, $C^\infty(\overline{U})$ is dense in $W^{k,p}(U)$, so once we construct a bounded linear operator $\mathcal{E}$ on $C^\infty(\overline{U})$, we can extend $\mathcal{E}$ to $W^{k,p}(U)$ by continuity. The remainder of the proof splits into two steps:

Step 1. The first step is to reduce the proof of Proposition 11.13 to verifying the following statement:

$$
\begin{gathered}
\text{There exists a linear operator that maps } u \in C^k(B(0,1) \cap \overline{\mathbb{R}^d_+}) \\
\text{with } \operatorname{supp} u \subset B(0,1) \cap \overline{\mathbb{R}^d_+} \text{ to an element } [u] \in C^k(B(0,1)) \\
\text{such that } \operatorname{supp}[u] \subset B(0,1),\ [u]|_{B(0,1)\cap\overline{\mathbb{R}^d_+}} = u \text{ and} \\
\|\tilde{\mathcal{E}}[u]\|_{W^{k,p}(B(0,1))} \leq C\|u\|_{W^{k,p}(B(0,1)\cap\mathbb{R}^d_+)} \\
\text{for some constant } C \text{ that only depends on } d,\ k \text{ and } p.
\end{gathered}
\tag{11.3}
$$

Indeed, by the assumption that ∂U is C^k, for each $x \in \partial U$, there exists $r(x) > 0$ and a C^k-diffeomorphism $\Psi_x : B(x, r(x)) \to B(0,1)$ such that $\Psi_x(B(x, r(x)) \cap U) = B(0,1) \cap \mathbb{R}^d_+$ and $\Psi_x(B(x, r(x)) \cap \partial U) = B(0,1) \cap \partial\mathbb{R}^d_+$. Shrinking $r(x)$ if necessary, we may assume further that $\overline{B(x, r(x))} \subset V$. By compactness, we can find finitely many such balls $W_1, \dots, W_N$ that cover ∂U. In addition, let W_0 be an open set such that

$$U \setminus (W_1 \cup \cdots \cup W_N) \subset W_0 \subset \overline{W_0} \subset U.$$

Then $\{W_0, W_1 \cap \overline{U}, \dots, W_N \cap \overline{U}\}$ is an open covering of $\overline{U}$. Let $\chi_0, \dots, \chi_N$ be a smooth partition of unity subordinate to this covering.

We now split

$$u = \chi_0 u + \sum_{j=1}^N \chi_j u$$

and define $\mathcal{E}[u]$ by extending each piece separately. Note that $\chi_0 u$ is easily extended to $\mathbb{R}^d$ by zero outside W_0. For each $\chi_j u$, we define the extension by

$$\tilde{\mathcal{E}}[\chi_j u \circ \Psi_j^{-1}] \circ \Psi_j$$

where Ψ_j is the C^k-diffemorphism from V_j to $B(0,1)$. At this point, by the chain rule, it is easy to verify that if $\tilde{\mathcal{E}}$ has the properties listed in (11.3), then the resulting operator $\mathcal{E}[u]$ has the desired properties.

Step 2. It remains to prove (11.3). We will use a high-order reflection technique. We introduce $k+1$ real numbers $\alpha_0, \dots, \alpha_k$ and $k+1$ positive numbers $\beta_0, \dots, \beta_k$, which will be chosen later, and define the extension $\tilde{u} = \tilde{\mathcal{E}}[u]$ on $B(0,1)$ by

$$\tilde{u}(x) = \begin{cases} u(x) & \text{when } x^d \geq 0 \\ \sum_{j=0}^k \alpha_j u(x', -\beta_j x^d) & \text{when } x^d < 0 \end{cases}$$

where $x' = (x^1, \dots, x^{d-1})$. By construction, $\tilde{u}|_{B(0,1)\cap\overline{\mathbb{R}^d_+}} = u$. Now, our goal is to choose the parameters so that $\tilde{u}$ belongs to $C^k(B(0,1))$; the remaining properties will then easily follow.

To show that all derivatives up to the k-th order of $\tilde{u}$ are continuous in $B(0,1)$, we need to show that

$$\lim_{z\to 0+} \partial_{x^d}^\ell u(x', z) = \lim_{z\to 0+} \partial_{x^d}^\ell \tilde{u}(x', -z). \tag{11.4}$$

Note that

$$\lim_{z\to 0+} \partial_{x^d}^\ell \tilde{u}(x', -z) = \sum_{j=0}^k \alpha_j (-\beta_j)^\ell \lim_{z\to 0+} \partial_{x^d}^\ell u(x', z).$$

Thus, we need to find α_j's and β_j's so that

$$1 = \sum_{j=0}^k (-\beta_j)^\ell \alpha_j \quad \text{for } \ell = 0, 1, \dots, k,$$

or in the matrix notation,

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} (-\beta_0)^0 & \cdots & (-\beta_k)^0 \\ (-\beta_0)^1 & & (-\beta_k)^1 \\ \vdots & & \vdots \\ (-\beta_0)^k & \cdots & (-\beta_k)^k \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_k. \end{pmatrix}$$

The $(k+1)\times(k+1)$ matrix is the *Vandermonde matrix* in $-\beta_0, \dots, -\beta_k$. In particular, if we choose $-\beta_j$'s to be pairwise distinct, then this matrix is invertible[23], so there exists a choice of α's so that (11.4) holds. If we further restrict $\beta_j \leq 1$, then $\operatorname{supp} \tilde{u} \subset B(0,1)$. Finally, the inequality $\|\tilde{u}\|_{W^{k,p}(B(0,1))} \leq C\|u\|_{W^{k,p}(B(0,1)\cap\mathbb{R}^d_+)}$ is easy to check. □

---

[23]In general,

$$\det \begin{pmatrix} x_0^0 & \cdots & x_k^0 \\ x_0^1 & & x_k^1 \\ \vdots & & \vdots \\ x_0^k & \cdots & x_k^k \end{pmatrix} = \prod_{0\leq i<j\leq k} (x_i - x_j).$$

To see this, note that both sides define polynomials of degree $0+1+\ldots+k$ that vanish whenever $x_i = x_j$ for some $i \neq j$; it follows that the two polynomials are proportional. To show that the proportionality constant is 1, note that the coefficient of in front of the monomial $x_0^0 x_1^1 \cdots x_k^k$ is 1 on both sides.

Remark 11.14. Our construction of $\mathcal{E}$ clearly depends on k, and we need ∂U to be C^k in order to perform the k-th order reflection procedure. Amazingly, it turns out that there exists a *universal* extension operator $\mathcal{E}$ that works for all $W^{k,p}(U)$ with $k \geq 0$, $1 \leq p < \infty$, which moreover only requires ∂U to be C^1 (even a bit weaker). This result is due to E. Stein; see [Ste70, Chapter VI].

11.4. **Traces (optional).** Let U be a bounded C^1 domain. Then by Proposition 11.10, $C^\infty(\overline{U})$ is a dense subset of $W^{1,p}(U)$. Each element in $u \in C^\infty(\overline{U})$ can be meaningfully restricted to a smooth function $u|_{\partial U}$ on $C^\infty(\partial U)$; we will call $u|_{\partial U}$ the *trace* of u on ∂U, and will write $\mathrm{tr}_{\partial U}\, u = u|_{\partial U}$. The following result allows us to extend this notion to a general element of $W^{1,p}(U)$.

Proposition 11.15. *Let $1 < p < \infty$, and let U be a bounded C^1 domain.*

(1) There exists a constant $C > 0$ that depends only on p and ∂U, such that for all $u \in C^\infty(\overline{U})$

$$\|\mathrm{tr}_{\partial U}\, u\|_{L^p(\partial U)} \leq C\|u\|_{W^{1,p}(U)}.$$

Hence, $\mathrm{tr}_{\partial U}$ extends to a bounded linear map $W^{1,p}(U) \to L^p(\partial U)$.

(2) An element $u \in W^{1,p}(U)$ belongs to $W^{1,p}_0(U)$ if and only if $\mathrm{tr}_{\partial U}\, u = 0$.

See [Eva10, Section 5.5] for a proof.

Proposition 11.15 leaves open the question of precisely identifying the image of the trace map. It turns out that answering this question necessitates the introduction of *fractional regularity spaces*. Here, we will only discuss the model case $p = 2$, $U = \mathbb{R}^d_+$ and $\partial U = \mathbb{R}^{d-1} \times \{0\}$, and leave the details to other references (see Remark 11.19 below).

The advantage of this case is that it is easy to extend the definition of the Sobolev space to general regularity indices $s \in \mathbb{R}$ via the Fourier transform.

Definition 11.16. Let $s \in \mathbb{R}$. For $u \in \mathcal{S}'(\mathbb{R}^d)$, we define the L^2-Sobolev norm with regularity index s by

$$\|u\|_{H^s(\mathbb{R}^d)} := \|(1 + |\xi|^2)^{\frac{s}{2}} \widehat{u}\|_{L^2_{(2\pi)^{-d} \mathrm{d}\xi}}.$$

The L^2-Sobolev space with regularity index s on $\mathbb{R}^d$ is defined as

$$H^s(\mathbb{R}^d) = \{u \in \mathcal{S}'(\mathbb{R}^d) : \|u\|_{H^s(\mathbb{R}^d)} < \infty\}.$$

We equip $H^s(\mathbb{R}^d)$ with the norm $\|\cdot\|_{H^s(\mathbb{R}^d)}$.

By Proposition 11.2.(3), $H^s(\mathbb{R}^d)$ agrees with our previous definition when k is a nonnegative integer.

The sharp trace theorem in this context is as follows.

Proposition 11.17. *For $u \in H^1(\mathbb{R}^d_+) \cap C^\infty(\mathbb{R}^d_+)$, we have*

$$\|\mathrm{tr}_{\partial \mathbb{R}^d_+}\, u\|_{H^{\frac{1}{2}}(\mathbb{R}^{d-1})} \leq C\|u\|_{H^1(\mathbb{R}^d_+)}.$$

Proof. In what follows, we will denote the first $d-1$ variables by x' and the corresponding Fourier variables by ξ'. We will use $\widehat{\cdot}$ for the Fourier transform in the first $d-1$ variables, and $\widetilde{\cdot}$ for all d variables.

We begin by noting that, by a reflection argument as in the proof of Proposition 11.13, we may find an extension $\mathcal{E}u \in H^1(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$ with $\|\mathcal{E}u\|_{H^1(\mathbb{R}^d)} \leq$

$C\|u\|_{H^1(\mathbb{R}^d_+)}$. Let us write $g = \mathrm{tr}_{\partial\mathbb{R}^d_+} u = \mathcal{E}u|_{\partial\mathbb{R}^d_+}$. By the Fourier inversion theorem in x^d, we have

$$\widehat{g}(\xi') = \int \widetilde{\mathcal{E}u}(\xi', \xi_d) e^{i\xi_d x^d} \frac{\mathrm{d}\xi_d}{2\pi}.$$

Hence we may estimate

$$\begin{aligned}
\|g\|_{H^{\frac{1}{2}}(\mathbb{R}^{d-1})} &= \|(1+|\xi'|^2)^{\frac{1}{4}} \widehat{g}(\xi')\|_{L^2_{(2\pi)^{-(d-1)}\mathrm{d}\xi'}} \\
&\leq \Big\| \int \frac{(1+|\xi'|^2)^{\frac{1}{4}}}{(1+|\xi'|^2+\xi_d^2)^{\frac{1}{2}}} (1+|\xi'|^2+\xi_d^2)^{\frac{1}{2}} |\widetilde{\mathcal{E}u}(\xi',\xi_d)| \frac{\mathrm{d}\xi_d}{2\pi} \Big\|_{L^2_{(2\pi)^{-(d-1)}\mathrm{d}\xi'}} \\
&\leq \sup_{\xi' \in \mathbb{R}^{d-1}} \left(\int \frac{(1+|\xi'|^2)^{\frac{1}{2}}}{1+|\xi'|^2+\xi_d^2} \frac{\mathrm{d}\xi_d}{2\pi} \right)^{\frac{1}{2}} \|(1+|\xi'|^2+\xi_d^2)^{\frac{1}{2}} \widetilde{\mathcal{E}u}(\xi',\xi_d)\|_{L^2_{(2\pi)^{-d}\,\mathrm{d}\xi}} \\
&\leq C\|\mathcal{E}u\|_{H^1(\mathbb{R}^d)} \\
&\leq C\|u\|_{H^1(\mathbb{R}^d_+)},
\end{aligned}$$

where the second to last line follows by bounding the ξ_d-integral using the change of variables $s = (1+|\xi'|^2)^{-\frac{1}{2}}\xi_d$. $\square$

That $H^{\frac{1}{2}}(\mathbb{R}^{d-1})$ is precisely the image of $\mathrm{tr}_{\partial\mathbb{R}^d_+}$ follows from the existence of a left inverse.

Proposition 11.18. *There exists a bounded linear map* $\mathrm{ext} : H^{\frac{1}{2}}(\mathbb{R}^{d-1}) \to H^1(\mathbb{R}^d_+)$ *such that* $\mathrm{tr}_{\partial\mathbb{R}^d_+} \circ \mathrm{ext} = Id$.

Proof. There are many possible ways to define ext; we will take ext to be the Poisson integral of $g \in H^{\frac{1}{2}}(\mathbb{R}^{d-1})$ and smoothly cut off in x^d. As in the previous proof, let us denote by $\widehat{\cdot}$ the Fourier transform in the first $d-1$ variables x'. For now, let $g \in \mathcal{S}(\mathbb{R}^{d-1})$. We define $u = \mathrm{ext} g$ by

$$\widehat{u}(\xi', x^d) = \eta(x^d) e^{-x^d|\xi'|} \widehat{g}(\xi'),$$

where $\eta \in C^\infty(\mathbb{R})$ is such that $\eta(s) = 1$ for $s < 1$ and $\eta(s) = 0$ for $s > 2$. It is not difficult to see that $\widehat{u} \in C^\infty(\mathbb{R}^d_+)$, and that $\mathrm{tr}_{\partial\mathbb{R}^d_+} u = g$. Moreover, by Plancherel's theorem, u and the tangential derivatives $\partial_j u$ ($j = 1, \ldots, d-1$) obey

$$\begin{aligned}
\|u\|^2_{L^2(\mathbb{R}^d_+)} + \sum_{j=1}^{d-1} \|\partial_j u\|^2_{L^2(\mathbb{R}^d_+)} &= \left\| \left\| (1+|\xi'|^2)^{\frac{1}{2}} \widehat{u}(\xi', x^d) \right\|_{L^2_{(2\pi)^{-(d-1)}\mathrm{d}\xi'}(\mathbb{R}^{d-1})} \right\|^2_{L^2_{x^d}(\mathbb{R}_+)} \\
&= \left\| \left\| \eta(x^d) e^{-x^d|\xi'|} \right\|_{L^2_{x^d}(\mathbb{R}_+)} (1+|\xi'|^2)^{\frac{1}{2}} \widehat{g}(\xi') \right\|^2_{L^2_{(2\pi)^{-(d-1)}\mathrm{d}\xi'(\mathbb{R}^{d-1})}}.
\end{aligned}$$

On the one hand, thanks to the support property of η, it is not difficult to show that $\left\| \eta(x^d) e^{-x^d|\xi'|} \right\|_{L^2_{x^d}(\mathbb{R}_+)} \leq 1$. On the other hand,

$$\left\| \eta(x^d) e^{-x^d|\xi'|} \right\|_{L^2_{x^d}(\mathbb{R}^d)} \leq \left(\int_0^\infty e^{-2x^d|\xi'|} \, \mathrm{d}x^d \int \right)^{\frac{1}{2}} \leq \frac{1}{(2|\xi'|)^{\frac{1}{2}}}.$$

It follows that

$$\left\| \eta(x^d) e^{-x^d|\xi'|} \right\|_{L^2_{x^d}(\mathbb{R}_+)} \leq C \min\{1, |\xi'|^{-\frac{1}{2}}\} \leq C(1+|\xi'|^2)^{-\frac{1}{4}},$$

so

$$\|u\|_{L^2(\mathbb{R}^d_+)}^2 + \sum_{j=1}^{d-1} \|\partial_j u\|_{L^2(\mathbb{R}^d_+)} \leq C\|(1+|\xi'|^2)^{\frac{1}{4}} \widehat{g}(\xi')\|_{L^2_{(2\pi)^{-(d-1)}\mathrm{d}\xi'}(\mathbb{R}^{d-1})}$$
$$= C\|g\|_{H^{\frac{1}{2}}(\mathbb{R}^{d-1})}.$$

Next, by the identity

$$\partial_d \widehat{u} = \eta'(x^d) e^{-x^d|x'|} \widehat{g}(\xi') - |\xi'| \eta(x^d) e^{-x^d|\xi'|} \widehat{g}(\xi'),$$

as well as the preceding bound, it follows that the normal derivative obeys

$$\|\partial_d \widehat{u}\|_{L^2(\mathbb{R}^d_+)} \leq C\|g\|_{H^{\frac{1}{2}}(\mathbb{R}^{d-1})}.$$

Hence, $\|u\|_{H^1(\mathbb{R}^d_+)} \leq C\|g\|_{H^{\frac{1}{2}}(\mathbb{R}^{d-1})}$, as desired. Now the general case follows by the density of $\mathcal{S}(\mathbb{R}^{d-1})$ in $H^{\frac{1}{2}}(\mathbb{R}^{d-1})$. □

The results that we discussed so far can be generalized to the case when U is a general bounded C^1 domain in $\mathbb{R}^d$. However, to define $H^s(\partial U)$ for $s \in \mathbb{R}$, we need additional tools that we do not currently have (e.g., interpolation theory). We refer to [Ste70, Chapter VI].

Remark 11.19. When $p \neq 2$, the image of $W^{1,p}(U)$ under the trace map turns out to be slightly different from the space $W^{1-\frac{1}{p},p}(\partial U)$; in fact, it is equal to what is called the *Besov space* $B_p^{1-\frac{1}{p},p}(\partial U)$. We will not go into any details, but note that as in the case $p = 2$, there exists an extension map $\mathrm{ext} : B_p^{1-\frac{1}{p},p}(\partial U) \to W^{1,p}(U)$ such that $\mathrm{tr}_{\partial U}\, \mathrm{ext} = Id$. Moreover, $B_2^{s,2} = H^s$. See [Ste70, Chapter VI] for more details.

11.5. **Sobolev inequalities.** *Sobolev inequalities* relate a Sobolev norm of a function with other norms (such as Sobolev, C^k or Hölder norms, where the latter will be defined later).

Gagliardo–Nirenberg–Sobolev inequality. We start with an inequality for smooth and compactly supported functions, which will be one of the basic building blocks for obtaining general Sobolev inequalities later.

Theorem 11.20 ($W^{1,1}$-Gagliardo–Nirenberg–Sobolev for $C_c^\infty(\mathbb{R}^d)$)**.** *Let $d \geq 2$ and $u \in C_c^\infty(\mathbb{R}^d)$. Then*

$$\|u\|_{L^{\frac{d}{d-1}}(\mathbb{R}^d)} \leq \|Du\|_{L^1(\mathbb{R}^d)}.$$

Remark 11.21 (Dimensional analysis & scaling exponent). The exponent $\frac{d}{d-1}$ on the LHS need not be memorized; it can be quickly computed through a process called *dimensional analysis*, which is also called *scaling analysis*. The idea is to note that both sides of the inequality behaves in a simple way under the scaling transformation $u(t,x) \to u_\lambda(t,x) := u(\lambda^{-1}t, \lambda^{-1}x)$. Then by requiring that the inequality to hold for u_λ for all $\lambda > 0$ and a nonzero function u, we will be able to read off the exponent $\frac{d}{d-1}$.

The ideas are as follows. We will say that a semi-norm (or more generally, a nonnegative function) $u \mapsto \|u\|_X$ is *homogeneous* if there exists $a \in \mathbb{R}$ such that

$$\|u_\lambda\|_X = \lambda^a \|u\|_X \quad \text{for all } \lambda > 0.$$

The exponent a is called the *degree of homogeneity* of $\|\cdot\|_X$. An example of a homogeneous norm is $\|D^\alpha(\cdot)\|_{L^p(\mathbb{R}^d)}$. By a quick computation, we see that

$$\|D^\alpha u_\lambda\|_{L^p(\mathbb{R}^d)} = \lambda^{\frac{d}{p}-|\alpha|}\|D^\alpha u\|_{L^p(\mathbb{R}^d)},$$

i.e., $\|D^\alpha(\cdot)\|_{L^p(\mathbb{R}^d)}$ is homogeneous of degree $\frac{d}{p} - |\alpha|$. A quick way to read off the degree of homogeneity is to note that:

- each derivative gives a factor of λ^{-1};
- the L^p-integral in each variable gives a factor of $\lambda^{\frac{1}{p}}$.

Let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be homogeneous semi-norms of degrees a and b, respectively. If an inequality of the form

$$\|u\|_X \leq C\|u\|_Y$$

were to hold for all u_λ, where $\|u\|_X \neq 0$, then both sides must have the same degree of homogeneity. Indeed, we would have

$$\lambda^a\|u\|_X = \|u_\lambda\|_X \leq C\|u_\lambda\|_Y = C\lambda^b\|u\|_Y$$

so unless $a = b$, we can take $\lambda \to 0$ or ∞ to conclude that $\|u\|_Y = 0$, which is a contradiction.

Applying the above procedure to the inequality of the form

$$\|u\|_{L^p} \leq C\|Du\|_{L^1},$$

we see that in order for such an inequality to hold for $u \neq 0$, the value of p must be exactly $\frac{d}{d-1}$, as in Theorem 11.20.

In the proof of Theorem 11.20, we will use the following inequality, which is of independent interest:

Lemma 11.22 (Loomis–Whitney inequality)**.** *For each $j = 1, \ldots, d$, let f_j be a nonnegative measurable function of all of $x^1, \ldots, x^d$* except x^j *(we will write this as $f = f(x^1, \ldots, \widehat{x^j}, \ldots, x^d)$). Then*

$$\int \cdots \int f_1 \cdots f_d \, \mathrm{d}x^1 \cdots \mathrm{d}x^d \leq \|f_1\|_{L^{d-1}(\mathbb{R}^{d-1})} \cdots \|f_d\|_{L^{d-1}(\mathbb{R}^{d-1})},$$

where

$$\|f_j\|_{L^{d-1}(\mathbb{R}^{d-1})} = \left(\int f_1^{d-1} \, \mathrm{d}x^1 \cdots \widehat{\mathrm{d}x^j} \cdots \mathrm{d}x^d\right)^{\frac{1}{d-1}}.$$

Proof of Lemma 11.22. The proof is integrating one variable at a time, and repeatedly applying of Hölder's inequality. In what follows, we use the notation $L^p_{x^{j_1}\cdots x^{j_k}}$ to denote the L^p norm in the variables $x^{j_1}, \ldots, x^{j_k}$.

We start by integrating $f_1 \cdots f_d$ in x^1. Using the independence of f_1 on x^1 to pull it out of the integral, and applying Hölder's inequality for the rest, we obtain

$$\begin{aligned}\int f_1 \cdots f_d \, \mathrm{d}x^1 &= f_1 \int f_2 \cdots f_d \, \mathrm{d}x^1 \\ &\leq f_1 \|f_2\|_{L^{d-1}_{x^1}} \cdots \|f_d\|_{L^{d-1}}.\end{aligned}$$

Next, we integrate in x^2. Using the independence of $\|f_2\|_{L^{d-1}_{x^1}}$ on x^2 to pull it out of the integral, and applying Hölder's inequality for the rest, we obtain

$$\iint f_1 \cdots f_d \, \mathrm{d}x^1 \mathrm{d}x^2 \leq \|f_1\|_{L^{d-1}_{x^2}} \|f_2\|_{L^{d-1}_{x^1}} \|f_3\|_{L^{d-1}_{x^1x^2}} \cdots \|f_d\|_{L^{d-1}_{x^1x^2}}.$$

If we carry out this procedure for each variable, all the way up to x^d, then

$$\int \cdots \int f_1 \cdots f_d \, \mathrm{d}x^1 \cdots \mathrm{d}x^d \leq \|f_1\|_{L^{d-1}_{x^2 \cdots x^{d-1}}} \cdots \|f_d\|_{L^{d-1}_{x^1 \cdots x^{d-1}}},$$

which prove the lemma. □

Remark 11.23 (Application to geometry). Lemma 11.22 has the following amusing geometric application. Consider a measurable subset E of $\mathbb{R}^d$. For each $j = 1, \dots, d$, let $\pi_j : \mathbb{R}^d \to \mathbb{R}^{d-1}$ be the j-th projection map $(x^1, \dots, x^d) \mapsto (x^1, \dots, \widehat{x^j}, \dots x^d)$, where $\widehat{x^j}$ indicates that the x^j-th coordinate is taken out. The question under consideration is this: *If we know the measure of each projection $|\pi_j(E)|$ of E, do we have an upper bound on the measure of the original set E?* As we will see, the answer is yes; in fact, we have

$$|E| \leq \prod_{j=1}^{d} |\pi_j(E)|^{\frac{1}{d-1}}.$$

The constant 1 in this inequality is sharp, as we can easily check by taking E to be the unit cube.

Indeed, applying Lemma 11.22 to $f_j = \mathbf{1}_{\pi_j(E)}$, it follows that

$$\begin{aligned} |E| &= \int \cdots \int \mathbf{1}_E \mathrm{d}x^1 \cdots \mathrm{d}x^d \\ &= \int \cdots \int \mathbf{1}_{\pi_1(E)} \cdots \mathbf{1}_{\pi_d(E)} \mathrm{d}x^1 \cdots \mathrm{d}x^d \\ &\leq \prod_{j=1}^{d} \|\mathbf{1}_{\pi_j(E)}\|_{L^{d-1}} = \prod_{j=1}^{d} |\pi_j(E)|^{\frac{1}{d-1}}, \end{aligned}$$

as desired.

We are now ready to prove Theorem 11.20.

Proof of Theorem 11.20. By the fundamental theorem of calculus, as well as the assumption that u is compactly supported, for $j = 1, \dots, d$, we have

$$\begin{aligned} |u(x)| &= \left| \int_{-\infty}^{x^j} \partial_{x^j} u(x + y e_j) \, \mathrm{d}y \right| \\ &\leq \int_{-\infty}^{\infty} |Du|(x + y e_j) \, \mathrm{d}y, \end{aligned}$$

where e_j is the unit vector in x^j-direction. Note furthermore that expression on the last line is independent of x^j. Thus, introducing the notation

$$g_j(x) = \left(\int_{-\infty}^{\infty} |Du|(x + y e_j) \, \mathrm{d}y \right)^{\frac{1}{d-1}},$$

we have

$$|u(x)|^{\frac{d}{d-1}} \leq \prod_{j=1}^{d} g_j(x), \quad g_j(x) = g_j(x^1, \dots, \widehat{x^j}, \dots, x^d).$$

Thus, by Lemma 11.22,

$$\int |u(x)|^{\frac{d}{d-1}} \, \mathrm{d}x \leq \int \prod_{j=1}^{d} g_j(x), \quad g_j(x) = g_j(x^1, \dots, \widehat{x^j}, \dots, x^d)$$
$$\leq \prod_{j=1}^{d} \|g_j\|_{L^{d-1}_{x^1 \cdots \widehat{x^j} \cdots x^d}}.$$

But for each j,

$$\|g_j\|_{L^{d-1}_{x^1 \cdots \widehat{x^j} \cdots x^d}} = \left(\int \cdots \int \int |Du| \, \mathrm{d}x^1 \cdots \mathrm{d}x^d \right)^{\frac{1}{d-1}} = \|Du\|_{L^1}^{\frac{1}{d-1}}.$$

Now the desired inequality follows. □

Remark 11.24 (Relationship with the isoperimetric inequality). Theorem 11.20 implies the *isoperimetric inequality*: If U is a sufficiently regular domain (say C^1), then

$$|U|^{\frac{d-1}{d}} \leq C|\partial U|. \tag{11.5}$$

Applying Theorem 11.20 to $\varphi_\epsilon * \mathbf{1}_U$ and taking $\epsilon \to 0$, it is not difficult to show that (11.5) holds with $C = 1$.

Remarkably, it turns out that the (11.5) also implies the Gagliardo–Nirenberg–Sobolev inequality, with the same constant. The proof involves approximating a general smooth compactly supported function by a linear combination of the characteristic functions of suitably regular domains, to each of which we apply (11.5). This connection is often used in geometric analysis to control the constant in the Sobolev inequality on a Riemannian manifold in terms of geometric information.

Sharp Sobolev inequalities for $W^{1,p}(U)$ for $1 \leq p < d$. From Theorem 11.20, we can deduce analogous inequalities for $W^{1,p}(\mathbb{R}^d)$ when $1 < p < d$.

Theorem 11.25 ($W^{1,p}$-Gagliardo–Nirenberg–Sobolev inequalities for $u \in C_c^\infty(\mathbb{R}^d)$)**.** *Let $d \geq 2$ and $u \in C_c^\infty(\mathbb{R}^d)$. Suppose that $1 < p < d$ and let $p^* = \frac{dp}{d-p}$. Then there exists a constant C, which depends only on d and p, such that*

$$\|u\|_{L^{p^*}(\mathbb{R}^d)} \leq C\|Du\|_{L^p}.$$

As in Theorem 11.20, there is no need to remember the exponent p^*; it can be read off by a dimensional analysis, which leads to

$$\frac{d}{p^*} = \frac{d}{p} - 1.$$

Proof of Theorem 11.25. We apply Theorem 11.20 to $|u|^\gamma$, where $\gamma = \frac{d-1}{d}p^*$. Then

$$\int |u|^{p^*} \, \mathrm{d}x = \int |u|^{\gamma \frac{d}{d-1}} \, \mathrm{d}x$$
$$\leq \left(\int |D|u|^\gamma| \, \mathrm{d}x \right)^{\frac{d}{d-1}}$$
$$\leq \gamma \left(\int |u|^{\gamma-1} |Du| \, \mathrm{d}x \right)^{\frac{d}{d-1}}.$$

To justify the inequality on the third line, we may approximate $|u|^\gamma$ by $(\epsilon^2 + |u|)^{\frac{\gamma}{2}}$, for which we can apply the usual chain rule, and take $\epsilon \to 0$ via the dominated convergence theorem. By Hölder's inequality, the last line is bounded by

$$\begin{aligned} &\leq \gamma \left(\| |u|^{\gamma-1} \|_{L^{\frac{p}{p-1}}(\mathbb{R}^d)} \|Du\|_{L^p} \right)^{\frac{d}{d-1}} \\ &= \gamma \|u\|_{L^{\frac{p}{p-1}(\gamma-1)}(\mathbb{R}^d)}^{\frac{d}{d-1}(\gamma-1)} \|Du\|_{L^p}^{\frac{d}{d-1}}. \end{aligned}$$

At this point, we note that

$$\frac{p}{p-1}(\gamma - 1) = p^*, \quad \frac{d}{d-1}(\gamma - 1) = p^* - \frac{d}{d-1}.$$

(The algebra may seem miraculous, but they are supposed to work out due to homogeneity!) Then after rearranging factors, the desired inequality follows. □

We now discuss the extension of the above results to elements in Sobolev spaces.

Theorem 11.26 (Sobolev inequalities for $W^{1,p}(U)$, $1 \leq p < d$)**.** *Let U be a domain in $\mathbb{R}^d$ and let $1 \leq p < d$. Let p^* be defined as in Theorem 11.25.*

(1) Then any $u \in W_0^{1,p}(U)$ belongs to $L^{p^*}(U)$, and there exists $C > 0$, that depends only on d, and p, such that

$$\|u\|_{L^{p^*}(U)} \leq C \|Du\|_{L^p(U)}.$$

(2) Assume, in addition, that U is a bounded C^1 domain. Then any $u \in W^{1,p}(U)$ belongs to $L^{p^*}(U)$, and there exists $C > 0$, that depends only on d, p and U, such that

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

Proof. The statement for $u \in W_0^{1,p}(U)$ is obvious, since by definition u can be approximated by smooth and compactly supported functions on $\mathbb{R}^d$. Next, when U is a bounded C^1 domain, Proposition 11.13 allows us to extend a general element $u \in W^{1,p}(U)$ to $E[u]$ in $W^{1,p}(\mathbb{R}^d)$ with a compact support. By Proposition 11.5, $E[u]$ can be approximated by smooth and compactly supported functions. By these observations, the second part follows. □

Failure of the Sobolev inequality from $W^{1,d}$ into L^∞. The borderline case $(p, p^*) = (d, \infty)$ turns out to be exceptional, and the Sobolev inequality *fails* in this case unless $d = 1$. For instance, the function

$$u(x) = \log \log \left(1 + \frac{1}{|x|} \right),$$

turns out to belong to $W^{1,d}(B(0,1))$ for $d \geq 2$, but it is unbounded near $x = 0$ (although ever so slowly, at a double-logarithmic rage!). By applying a smooth cutoff and mollifying this example, we can also produce a family of counterexamples to the inequality $\|u\|_{L^\infty(\mathbb{R}^d)} \leq C \|Du\|_{L^d(\mathbb{R}^d)}$ for $u \in C_c^\infty(\mathbb{R}^d)$.

Later, we will discuss a substitute for the false L^∞-Sobolev inequality that turns to be useful in many applications; see Proposition 11.35 below.

A potential estimate. As a preparation for the discussion of the case $p > d$, we state and prove an inequality for smooth functions on $\mathbb{R}^d$, which will be our basic building block.

Lemma 11.27 (A potential estimate)**.** *Let $u \in C^1(\overline{B(x,r)})$. Then*

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |u(y) - u(x)| \,\mathrm{d}y \leq \frac{1}{d\alpha(d)}\int_{B(x,r)} \frac{|Du(y)|}{|x-y|^{d-1}}\,\mathrm{d}y. \tag{11.6}$$

We call (11.6) a potential estimate, since the RHS resembles the gradient of the Newtonian potential.

Proof. We start by estimating the integral

$$\int_{\partial B(x,r')} |u(y) - u(x)|\,\mathrm{d}S(y).$$

By the fundamental theorem of calculus and the change of variables formula from polar coordinates to rectangular coordinates, we may estimate

$$\begin{aligned}\int_{\partial B(x,r')} |u(y) - u(x)|\,\mathrm{d}S(y) &= (r')^{d-1}\int_{\partial B(0,1)} |u(x + r'z) - u(x)|\,\mathrm{d}S(z)\\ &\leq (r')^{d-1}\int_{\partial B(0,1)}\int_0^{r'} |Du(x+sz)|\,\mathrm{d}s\,\mathrm{d}S(z)\\ &= (r')^{d-1}\int_{\partial B(0,1)}\int_0^{r'} s^{-d+1}|Du(x+sz)|s^{d-1}\,\mathrm{d}s\,\mathrm{d}S(z)\\ &= (r')^{d-1}\int_{B(x;r')} \frac{|Du(y)|}{|x-y|^{d-1}}\,\mathrm{d}y. \qquad (11.7)\end{aligned}$$

Taking $\int_0^r (\cdots)\,\mathrm{d}r'$ of both sides, we obtain

$$\begin{aligned}\int_{B(x,r)} |u(y) - u(x)|\,\mathrm{d}y &= \int_0^r \int_{\partial B(x,r')} |u(y) - u(x)|\,\mathrm{d}S(y)\,\mathrm{d}r'\\ &= \int_0^r (r')^{d-1}\int_{B(x;r')} \frac{|Du(y)|}{|x-y|^{d-1}}\,\mathrm{d}y\,\mathrm{d}r'\\ &\leq \frac{1}{d}r^d \int_{B(x;r)} \frac{|Du(y)|}{|x-y|^{d-1}}\,\mathrm{d}y.\end{aligned}$$

Recalling that $|B(x,r)| = \alpha(d)r^d$, the desired inequality follows. □

Remark 11.28. Let $u \in C_c^\infty(\mathbb{R}^d)$. Then if we take $r' \to \infty$ in (11.7), the integral $\frac{1}{|\partial B(x',r)|}\int_{\partial B(x,r')} u\,\mathrm{d}y$ vanishes, so we obtain

$$|u(x)| \leq \int_{\mathbb{R}^d} \frac{|Du(y)|}{|x-y|^{d-1}}\,\mathrm{d}y. \tag{11.8}$$

This estimate can be used as an alternative starting point for Theorem 11.25 for $p > 1$; see Remark 11.38 below.

Hölder spaces and Morrey's inequality. In the case $p > d$, it turns out that an element $u \in W^{1,p}(U)$ is not only continuous[24], but also it enjoys a bound on the modulus of continuity. To precisely state this property, we introduce the notion of *Hölder spaces.*

Definition 11.29 (Hölder space)**.** Let K be a closed subset of $\mathbb{R}^d$. For $0 < \alpha < 1$, the *Hölder semi-norm of regularity index* α of a continuous function $uinC(K)$ is defined as

$$[u]_{C^{0,\alpha}(K)} = \sup_{x,y \in K} \frac{|u(x) - u(y)|}{|x - y|^\alpha}.$$

The $C^{0,\alpha}$-norm of $u \in C(K)$ is defined as

$$\|u\|_{C^{0,\alpha}(K)} = \|u\|_{C^0(K)} + [u]_{C^{0,\alpha}(K)}.$$

The space $C^{0,\alpha}(K)$ is defined to be all continuous functions on K for which $\|u\|_{C^{0,\alpha}(K)} < \infty$, equipped with the norm $\|\cdot\|_{C^{0,\alpha}(K)}$.

More generally, for any C^k function u on K, we define

$$\|u\|_{C^{k,\alpha}(K)} = \|u\|_{C^k(K)} + \sum_{\alpha : |\alpha| = k} \|D^\alpha u\|_{C^{0,\alpha}(K)}, \quad \|u\|_{C^k(K)} = \sum_{\alpha : |\alpha| \leq k} \|D^\alpha u\|_{C^0}.$$

The space $C^{k,\alpha}(K)$ is defined to be all continuous functions on K for which $\|u\|_{C^{k,\alpha}(K)} < \infty$, equipped with the norm $\|\cdot\|_{C^{0,\alpha}(K)}$.

We state, without detailed proofs, some elementary properties of Hölder spaces:

Lemma 11.30. *Let K be a closed subset of $\mathbb{R}^d$. Let k be a nonnegative integer and $0 < \alpha < 1$.*
(1) $C^{k,\alpha}(K)$, equipped with the norm $\|\cdot\|_{C^{k,\alpha}(K)}$ is a Banach space.
(2) We have $\|u\|_{C^k(K)} \leq \|u\|_{C^{k,\alpha}(K)} \leq C\|u\|_{C^{k+1}(K)}$. Moreover, for $0 < \alpha' < \alpha$,

$$\|u\|_{C^{k,\alpha'}(K)} \leq C\|u\|_{C^{k,\alpha}(K)}.$$

(3) If $L \subseteq K$, then

$$\|u\|_{C^{k,\alpha}(L)} \leq \|u\|_{C^{k,\alpha}(K)}.$$

Using Lemma 11.27, we obtain the following inequality.

Theorem 11.31. *Let $u \in C^\infty(\mathbb{R}^d) \cap W^{1,p}(\mathbb{R}^d)$. Let $p > d$, and define α by*

$$\alpha = 1 - \frac{d}{p}.$$

Then there exists $C > 0$, which depends only on d and p, such that

$$\|u\|_{C^{0,\alpha}(\mathbb{R}^d)} \leq C\|u\|_{W^{1,p}(\mathbb{R}^d)}.$$

Although both sides are not homogeneous, the exponent α can still be read off by performing dimensional analysis of the top-order terms. Indeed, note that the degree of homogeneity of $[u]_{C^{0,\alpha}}$ is $-\alpha$, whereas that of $\|Du\|_{L^p(\mathbb{R}^d)}$ is $\frac{d}{p} - 1$; equating the two gives the above value of α.

[24]To be pedantic, we have to be careful since u is, at the outset, only a locally integrable function, it is defined only up to identity almost everywhere. See Theorem 11.32 for a precise statement.

Proof. We begin by bounding $\|u\|_{C^0(\mathbb{R}^d)} = \sup_{x\in\mathbb{R}^d} |u(x)|$. For any $x \in \mathbb{R}^d$ and $r > 0$ to be chosen, we may estimate

$$|u(x)| \leq \frac{1}{|B(x,r)|} \int_{B(x,r)} |u(x) - u(y)| \,\mathrm{d}y + \frac{1}{|B(x,r)|} \int_{B(x,r)} |u(y)| \,\mathrm{d}y.$$

For the second term, we simply use Hölder's inequality to estimate

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} |u(y)| \,\mathrm{d}y \leq C r^{-\frac{d}{p}} \|u\|_{L^p(B(x,r))}.$$

For the first term, we use Lemma 11.27 and Hölder's inequality to estimate

$$\begin{aligned}\frac{1}{|B(x,r)|} \int_{B(x,r)} |u(x) - u(y)| \,\mathrm{d}y &\leq C \int_{B(x,r)} \frac{Du(y)}{|x-y|^{d-1}} \,\mathrm{d}y \\ &\leq C \left(\int_{B(x,r)} \frac{\mathrm{d}y}{|x-y|^{(d-1)\frac{p}{p-1}}} \right)^{\frac{p-1}{p}} \|Du\|_{L^p(B(x,r))}.\end{aligned}$$

Since $p > d$, we have $(d-1)\frac{p}{p-1} < d$ so that the integral converges. Computing its value, we obtain

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} |u(x) - u(y)| \,\mathrm{d}y \leq C r^{\alpha} \|Du\|_{L^p(B(x,r))}. \tag{11.9}$$

Thus,

$$|u(x)| \leq C r^{\alpha} \|Du\|_{L^p(B(x,r))} + C r^{-\frac{d}{p}} \|u\|_{L^p(B(x,r))}.$$

Taking $r = 1$, we obtain a bound for $\|u\|_{C^0(\mathbb{R}^d)}$ in terms of $\|u\|_{W^{1,p}(\mathbb{R}^d)}$.

Next, we estimate the Hölder semi-norm. For $x, y \in \mathbb{R}^d$ such that $|x - y| = r$, we estimate

$$\begin{aligned}|u(x) - u(y)| &\leq \frac{1}{B(x,r) \cap B(y,r)} \int_{B(x,r)\cap B(y,r)} |u(x) - u(z)| + |u(z) - u(y)| \,\mathrm{d}z \\ &\leq C \frac{1}{B(x,r)} \int_{B(x,r)} |u(x) - u(z)| \,\mathrm{d}z + C \frac{1}{B(y,r)} \int_{B(y,r)} |u(y) - u(z)| \,\mathrm{d}z \\ &\leq C \int_{B(x,r)} \frac{|Du(z)|}{|x-z|^{d-1}} \,\mathrm{d}z + C \int_{B(y,r)} \frac{|Du(z)|}{|y-z|^{d-1}} \,\mathrm{d}z.\end{aligned}$$

On the second line, we used the simple geometric fact that all of $|B(x,r) \cap B(y,r)|$, $|B(x,r)|$ and $|B(y,r)|$ are proportional to r^d. On the third line, we used Lemma 11.27. By (11.9), it follows that

$$|u(x) - u(y)| \leq C r^{\alpha} \|Du\|_{L^p(\mathbb{R}^d)}.$$

Recalling that $r = |x - y|$, it follows that $[u]_{C^\alpha(\mathbb{R}^d)} \leq C \|Du\|_{L^p(\mathbb{R}^d)}$, as desired. □

Theorem 11.32 (Sobolev inequalities for $W^{1,p}(U)$, $p > d$)**.** *Let U be domain in $\mathbb{R}^d$ and let $p > d$. Let α be defined as in Theorem 11.31.*

(1) For any $u \in W^{1,p}_0(U)$, there exists a function $u^* \in C^{0,\alpha}(\overline{U})$ that agrees with u almost everywhere in U. Moreover, there exists $C > 0$, which depends only on d and p, such that

$$\|u\|_{C^{0,\alpha}(\overline{U})} \leq C \|u\|_{W^{1,p}(U)}.$$

(2) Assume, in addition, that U is a bounded C^1 domain. Then for any $u \in W^{1,p}(U)$, there exists a function $u^* \in C^{0,\alpha}(\overline{U})$ that agrees with u almost everywhere in U. Moreover, there exists $C > 0$, which depends only on d, p and U, such that

$$\|u\|_{C^{0,\alpha}(\overline{U})} \leq C\|u\|_{W^{1,p}(U)}.$$

Like Theorem 11.26, this result follows from Theorem 11.31 via approximation and extension.

The exceptional case: $W^{1,d} \not\hookrightarrow L^\infty$ and the space of bounded mean oscillation (Optional). Just like what happened for Sobolev inequalities, for many results in analysis concerning Lebesuge spaces, the space L^∞ often turns out be exceptional. In many cases, the following larger space serves as a good substitute:

Definition 11.33 (Functions of bounded mean oscillation)**.** For a locally integrable function u, the *bounded mean oscillation (BMO) semi-norm* is defined as

$$[u]_{BMO} = \sup_{x\in\mathbb{R}^d,\, r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |u(y) - \frac{1}{B(x,r)} \int_{B(x,r)} u(z)\,\mathrm{d}z|\,\mathrm{d}y.$$

If $[u]_{BMO} < \infty$, then we say that u has *bounded mean oscillation*, and the space of all functions of bounded mean oscillation on $\mathbb{R}^d$ is denoted by $BMO(\mathbb{R}^d)$.

Remark 11.34. Note that $[u]_{BMO} = 0$ if and only if $u = const$. Thus, it is natural to identity two elements in $BMO(\mathbb{R}^d)$ that differ by a constant function (i.e., quotient out by the subspace of constant functions). On the resulting quotient space, $[u]_{BMO}$ becomes a complete norm.

A proper discussion of the uses of the BMO space in analysis, and an explanation of why BMO often serves as a good substitute for L^∞, lies outside the scope of this course; we refer to [Ste93, Chapter IV] for those who are interested. Here, let us just show $W^{1,d}$ indeed embeds into BMO.

Proposition 11.35 (Sobolev inequality for $W^{1,d}$ into BMO)**.** *Let $u \in C^\infty(\mathbb{R}^d)$ for $d \geq 2$. Then*

$$[u]_{BMO(\mathbb{R}^d)} \leq C\|u\|_{W^{1,d}(\mathbb{R}^d)}.$$

In [Eva10, Section 5.8.1], you can find a proof that involves a contradiction argument. Here, we give an alternative direct proof, which instead relies on Lemma 11.27 and the following result from real analysis:

Theorem 11.36 (Hardy–Littlewood maximal function theorem)**.** *Given a locally integrable function f on $\mathbb{R}^d$, define the associated* maximal function *Mf as*

$$Mf(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\,\mathrm{d}y.$$

Then for any $1 < p \leq \infty$, there exists $C > 0$ that depends only on p, such that

$$\|Mf\|_{L^p(\mathbb{R}^d)} \leq C\|f\|_{L^p(\mathbb{R}^d)}.$$

For a proof, see [Fol99, Theorem 3.17].

Proof of Proposition 11.35. Let $x \in \mathbb{R}^d$ and $r > 0$. We first estimate

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} |u(y) - \frac{1}{B(x,r)} \int_{B(x,r)} u(z)\,\mathrm{d}z|\,\mathrm{d}y$$

$$
\begin{aligned}
&\leq \frac{1}{|B(x,r)|}\int_{B(x,r)} \frac{1}{B(x,r)}\int_{B(x,r)} |u(y)-u(z)|\,\mathrm{d}y\,\mathrm{d}z \\
&\leq 2^{-d}\frac{1}{|B(x,r)|}\int_{B(x,r)} \frac{1}{B(z,2r)}\int_{B(z,2r)} |u(y)-u(z)|\,\mathrm{d}y\,\mathrm{d}z \\
&\leq C\frac{1}{|B(x,r)|}\int_{B(x,r)}\int_{B(z,2r)} \frac{|Du(y)|}{|z-y|^{d-1}}\,\mathrm{d}y\,\mathrm{d}z.
\end{aligned}
$$

Here, it is tempting to apply Young's inequality, but it unfortunately fails since $|x-y|^{-d+1}$ (barely) fails to be in $L^{d-1}(\mathbb{R}^d)$. To get around this issue, we appeal to Theorem 11.36 as follows. For each z, we may estimate

$$
\begin{aligned}
\int_{B(z,2r)} \frac{|Du(y)|}{|z-y|^{d-1}}\,\mathrm{d}y &\leq \sum_{k=0}^{\infty}\int_{2^{-k-1}r<|z-y|\leq 2^{-k}r} \frac{|Du(y)|}{|z-y|^{d-1}}\,\mathrm{d}y \\
&\leq C\sum_{k=0}^{\infty} 2^{k(d-1)}r^{-(d-1)}\int_{2^{-k-1}r<|z-y|\leq 2^{-k}r} |Du(y)|\,\mathrm{d}y \\
&\leq C\sum_{k=0}^{\infty} 2^{-k} r M|Du|(z) \leq CrM|Du|(z).
\end{aligned}
$$

Therefore, by Theorem 11.36 with $p=d$,

$$
\begin{aligned}
&C\frac{1}{|B(x,r)|}\int_{B(x,r)}\int_{B(z,2r)} \frac{|Du(y)|}{|z-y|^{d-1}}\,\mathrm{d}y\,\mathrm{d}z \\
&\leq Cr^{-d+1}\int_{B(x,r)} M|Du|(z)\,\mathrm{d}z \\
&\leq C\|M|Du|\|_{L^d(\mathbb{R}^d)} \leq C\|Du\|_{L^d(\mathbb{R}^d)},
\end{aligned}
$$

as desired. □

Remark 11.37 (Hardy–Littlewood–Sobolev fractional integration)*.* By essentially same argument as in the previous proof, one obtains the *Hardy–Littlewood–Sobolev fractional integration theorem*: Let $0<\alpha<d$ and $1<p<q<\infty$ obey

$$
\frac{d}{p} = \frac{d}{q} + d - \alpha.
$$

Then for any $u \in C^\infty_c(\mathbb{R}^d)$,

$$
\left\| \int_{\mathbb{R}^d} \frac{|u(y)|}{|x-y|^\alpha}\,\mathrm{d}y \right\|_{L^q(\mathbb{R}^d)} \leq C\|u\|_{L^p(\mathbb{R}^d)},
$$

where $C>0$ depend only on d, p, q and α.

Remark 11.38 (Alternative proof of Gagliardo–Nirenberg–Sobolev for $p>1$)*.* Combining Remarks 11.28 and 11.37, we also obtain an alternative proof of the Gagliardo–Nirenberg–Sobolev inequality (Theorem 11.25) for $1<p<d$.

General Sobolev inequalities. From the inequalities proved so far, it is not difficult to deduce the following general Sobolev inequalities for $W^{k,p}$.

Theorem 11.39 (Sobolev inequalities for $W^{k,p}$)**.** *Let k be a nonnegative integer and let $1 \leq p < \infty$. Assume that either*

- *U is a domain in $\mathbb{R}^d$ and $u \in W^{k,p}_0(U)$; or*
- *U is a bounded C^k domain in $\mathbb{R}^d$ and $u \in W^{k,p}(U)$.*

Then the following statements hold.

(1) *Let ℓ be a nonnegative integer such that $\ell \leq k$ and let $1 \leq q < \infty$. If*

$$\frac{d}{q} - \ell \geq \frac{d}{p} - k,$$

then u belongs to $W^{\ell,q}(U)$. Moreover, there exists $C > 0$, which depends only on d, k, ℓ, p, q and U, such that

$$\|u\|_{W^{\ell,q}(U)} \leq C\|u\|_{W^{k,p}(U)}.$$

(2) *Let ℓ be a nonnegative integer such that $\ell \leq k$ and let $0 < \alpha < 1$. If*

$$-\ell - \alpha \geq \frac{d}{p} - k,$$

then there exists a function $u^* \in C^{k,\alpha}(U)$ such that $u^* = u$ almost everywhere in U. Moreover, there exists $C > 0$, which depends only on d, k, ℓ, p, α and U, such that

$$\|u^*\|_{C^{\ell,\alpha}(U)} \leq C\|u\|_{W^{k,p}(U)}.$$

The assumptions seem rather complicated, but actually they are not too difficult to remember. The key points are:

- The regularity exponent ℓ on the LHS *cannot* exceed the regularity exponent k on the RHS;
- The integrability exponent q on the LHS *cannot* be ∞;
- The degree of homogeneity of the top-order term on the LHS *cannot* be bigger than that of the RHS (to remember which direction this condition goes, just think about the trivial case $\|u\|_{L^p} \leq \|u\|_{W^{k,p}}$!)

Theorem 11.39 is a straightforward consequence of concatenating earlier results; we leave the details of the proof as an exercise.

11.6. **Compactness (optional).** We now study compactness properties of a sequence of functions that are bounded in $W^{k,p}(U)$ or $C^{k,\alpha}(K)$. Compactness is a key tool to show the existence of a solution to a PDE; roughly speaking, its typical use is to show that an appropriate sequence of "approximate solutions" to the equation converges (may be after passing to a subsequence) to an actual solution.

A bounded sequence in $W^{k,p}(U)$ or $C^{k,\alpha}(K)$ will *not* be compact in the same space (because they are infinite dimensional!), but it will be in appropriate larger spaces. The key notion is that of a *compactly embedded Banach space*:

Definition 11.40. Let X, Y be Banach spaces such that $X \subset Y$. We say that X is *compactly embedded in* Y, and write $X \subset\subset Y$ if

(1) $\|u\|_Y \leq C\|u\|_X$ for some constant $C > 0$ (independent of $u \in X$); and
(2) if $\{u_k\}$ is a bounded sequence in X (i.e., $\sup_k \|u_k\|_X < \infty$), then there exists a subsequence $\{u_{k_j}\}$ of $\{u_k\}$ that is convergent in Y.

Recall the *Arzela–Ascoli theorem*:

Theorem 11.41. *Let K be a compact subset of $\mathbb{R}^d$, and let $\{u_k\}$ be a sequence of continuous functions on K with the following properties:*

(1) *(uniform boundedness) $\sup_k \sup_{x \in K} |u_k(x)| < \infty$.*
(2) *(equicontinuity) for every $\epsilon > 0$, there exists $\delta > 0$ such that $|u_k(x) - u_k(y)| < \epsilon$ for every k and x, y such that $|x - y| < \delta$.*

Then there exists a subsequence $\{u_{k_j}\}$ of $\{u_k\}$ that is uniformly convergent on K.

By Theorem 11.41, it is not difficult to prove the following compact embedding property of Hölder spaces.

Proposition 11.42. *Let K be a compact subset of $\mathbb{R}^d$. Let $0 \leq \alpha' < \alpha < 1$. Then*

$$C^{0,\alpha}(K) \subset\subset C^{0,\alpha'}(K),$$

where $C^{0,0}(K)$ should be interpreted as $C(K)$ equipped with the uniform topology.

Proof. Let $\{u_k\}$ be a bounded sequence in $C^{0,\alpha}(K)$. Clearly, $\{u_k\}$ is uniformly bounded; moreover, since $|u_k(x) - u_k(y)| \leq C|x-y|^\alpha$ for a constant $C > 0$ independent of x, y and k, it follows that $\{u_k\}$ is equicontinuous. By Theorem 11.41, there exists a uniformly convergent subsequence $\{u_{k_j}\}$. Hence, the case $\alpha' = 0$ follows. When $0 < \alpha' < \alpha$, we note that

$$\begin{aligned}
[u_{k_j} - u_{k_{j'}}]_{C^{0,\alpha'}} &= \sup_{x,y\in K} |x-y|^{-\alpha'} |(u_{k_j} - u_{k_{j'}})(x) - (u_{k_j} - u_{k_{j'}})(y)| \\
&\leq \sup_{x,y\in K} |x-y|^{-\alpha'} |(u_{k_j} - u_{k_{j'}})(x) - (u_{k_j} - u_{k_{j'}})(y)|^{\frac{\alpha'}{\alpha}} \\
&\quad \times \sup_{x,y\in K} \Big(|(u_{k_j} - u_{k_{j'}})(x)| + |(u_{k_j} - u_{k_{j'}})(y)| \Big)^{1-\frac{\alpha'}{\alpha}} \\
&\leq \Big([u_{k_j}]_{C^{0,\alpha}(K)} + [u_{k_{j'}}]_{C^{0,\alpha}(K)} \Big)^{\frac{\alpha'}{\alpha}} \Big(2\|u_{k_j} - u_{k_{j'}}\|_{C^0(K)} \Big)^{1-\frac{\alpha'}{\alpha}}.
\end{aligned}$$

The first factor is uniformly bounded, where as the second factor goes to zero as $j, j' \to \infty$ by the uniform convergence. Hence, $\{u_{k_j}\}$ is convergent in $C^{0,\alpha'}(K)$ as well. □

The key idea was that the excess regularity $\alpha - \alpha'$ implies the equicontinuity property needed for compactness in the Arzela–Ascoli theorem. It turns out that a similar phenomenon holds for non-sharp Sobolev inequalities:

Theorem 11.43 (Rellich–Kondrachov)**.** *Let U be a bounded C^1 domain. Let $1 \leq p < d$ and $1 \leq q < p^*$. Then*

$$W^{1,p}(U) \subset\subset L^q(U). \tag{11.10}$$

As a preparation for the proof, we prove a property of convolutions that is of independent interest, namely, that if $u \in W^{k,p}(\mathbb{R}^d)$ for $k > 0$, then mollifications of u converge to u in $L^p(\mathbb{R}^d)$ at a controlled, accelerated rate.

Proposition 11.44 (Accelerated mollification)**.** *Let k be a positive integer and $1 \leq p < \infty$. Let $\varphi \in C_c^\infty(\mathbb{R}^d)$ obey*

$$\int \varphi = 1, \quad \int x^\alpha \varphi = 0 \quad \text{for } 1 \leq |\alpha| \leq k-1.$$

If $u \in W^{k,p}(\mathbb{R}^d)$, then

$$\|u - \varphi_\epsilon * u\|_{L^p} \leq C\epsilon^k \sum_{\alpha:|\alpha|=k} \|D^\alpha u\|_{L^p}$$

We note that the second condition for φ is vacuous when $k = 1$.

Proof. As in Proposition 11.5, we begin with the identity

$$u(x) - \varphi_\epsilon * u(x) = \int \varphi(z)\,(u(x) - u(x - \epsilon z))\,\mathrm{d}z.$$

By Taylor's formula, we have

$$\begin{aligned} u(x - \epsilon z) = & \sum_{\alpha:|\alpha|\leq k-1} \frac{(-\epsilon)^{|\alpha|}}{\alpha!} z^\alpha D^\alpha u(x) \\ & + k(-\epsilon)^k \sum_{\alpha:|\alpha|=k} \int_0^1 z^\alpha D^\alpha u(x - \epsilon t z)(1-t)^{k-1}\,\mathrm{d}t. \end{aligned}$$

Thus,

$$\begin{aligned} & \int \varphi(z)\,(u(x) - u(x - \epsilon z))\,\mathrm{d}z \\ & = \sum_{\alpha:1\leq|\alpha|\leq k-1} \frac{(-\epsilon)^{|\alpha|}}{\alpha!} \int z^\alpha \varphi(z)\,\mathrm{d}z\, D^\alpha u(x) \\ & \quad + k(-\epsilon)^k \sum_{\alpha:|\alpha|=k} \int z^\alpha \varphi(z) \int_0^1 D^\alpha u(x - \epsilon t z)(1-t)^{k-1}\,\mathrm{d}t\,\mathrm{d}z. \end{aligned}$$

By hypothesis, all terms but the last term on the RHS vanish. For the L^p norm of the last term, we use Minkowski's inequality to estimate

$$\left\| k(-\epsilon)^k \sum_{\alpha:|\alpha|=k} \int_0^1 z^\alpha D^\alpha u(x - \epsilon t z)(1-t)^{k-1}\,\mathrm{d}t \right\|_{L^p(\mathbb{R}^d)} \leq C\epsilon^k \sum_{\alpha:|\alpha|=k} \|D^\alpha u\|_{L^p(\mathbb{R}^d)},$$

as desired. □

Proof of Theorem 11.43. The first step of the proof is notice that it suffices prove

$$W^{1,p}(U) \subset\subset L^p(U). \tag{11.11}$$

Indeed, if $1 \leq q < p$, then (11.10) would follow from (11.11) and the embedding $L^p(U) \subset L^q(U)$ (Hölder inequality). In the case, $p < q < p^*$, by Hölder's inequality and Theorem 11.26, we have

$$\|u\|_{L^q(U)} \leq \|u\|_{L^p(U)}^\theta \|u\|_{L^{p^*}(U)}^{1-\theta} \leq C\|u\|_{L^p(U)}^\theta \|u\|_{W^{1,p}(U)}^{1-\theta},$$

where $0 < \theta < 1$ is characterized by $\frac{1}{q} = \theta\frac{1}{p} + (1-\theta)\frac{1}{p^*}$. It follows that if $\{u_k\}$ is a sequence that is bounded in $W^{1,p}(U)$ and convergent in $L^p(U)$, then it is convergent in $L^q(U)$. Using this observation, (11.10) follows from (11.11).

It remains to prove (11.11). Here, the idea is combine Proposition 11.44 with the Arzela–Ascoli theorem. We claim if $\{u_k\}$ is a bounded sequence in $W^{1,p}(U)$, then for every $n > 0$, there exists a subsequence u_{k_j} and J such that $\|u_{k_j} - u_{k_{j'}}\|_{L^p} < \frac{1}{n}$ for all $j, j' \geq J$. Then by a standard diagonal argument, we may extract a convergent subsequence of the original sequence.

Let us prove the claim. Let $\{u_k\}$ be a bounded sequence in $W^{1,p}(U)$ and fix $n > 1$. Choosing an open domain V such that $\overline{U} \subset V$, we may apply Proposition 11.13 to extend $\{u_k\}$ to a bounded sequence (which we will still denote by u_k) in $W^{1,p}(\mathbb{R}^d)$

with $\operatorname{supp} u_k \subset V$. Fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ such that $\int \varphi = 1$. If $\epsilon > 0$ is sufficiently small, then $\operatorname{supp} \varphi_\epsilon * u \subset V$. Moreover, by Proposition 11.44,

$$\|u_k - \varphi_\epsilon * u_k\|_{L^p} \leq C\epsilon \|Du_k\|_{L^p}, \tag{11.12}$$

so choosing ϵ small enough, we may also ensure that

$$\|u_k - \varphi_\epsilon * u_k\|_{L^p(\mathbb{R}^d)} < \frac{1}{3n} \tag{11.13}$$

for every k.

Next, for such an $\epsilon > 0$, note that $\{\varphi_\epsilon * u_k\}_k$ is a sequence of continuous function supported in V whose C^0 and C^1 norms are uniformly bounded. Hence Theorem 11.41 is applicable, so there exists a subsequence u_{k_j} so that $\{\varphi_\epsilon * u_{k_j}\}$ is uniformly convergent. In particular, there exists J such that for $j, j' \geq J$,

$$\|\varphi_\epsilon * u_{k_j} - \varphi_\epsilon * u_{k_j}\|_{L^p(\mathbb{R}^d)} \leq |V|^{\frac{1}{p}} \|\varphi_\epsilon * u_{k_j} - \varphi_\epsilon * u_{k_j}\|_{L^\infty(\mathbb{R}^d)} < \frac{1}{3n}.$$

Combined with (11.13), it follows that

$$\|u_{k_j} - u_{k_{j'}}\|_{L^p(\mathbb{R}^d)} < \frac{1}{n} \quad \text{for } j, j' \geq J$$

as desired. □

We note the following consequence of Theorems 11.43 (when $1 \leq p < \infty$) and 11.41 ($p = \infty$).

Corollary 11.45. *Let U be a bounded C^1 domain, then for any $1 \leq p \leq \infty$,*

$$W^{1,p}(U) \subset\subset L^p(U).$$

We omit the straightforward proof.

11.7. **Poincaré and Hardy inequalities.** We now discuss ways to obtain information about a function u from only the information $Du \in L^p(U)$. A principal example of such an inequality is *Poincaré's inequality*:

Proposition 11.46 (Poincaré's inequality)**.** *Let U be a bounded connected C^1 domain. For any $1 \leq p \leq \infty$ and $u \in W^{1,p}(U)$, we have*

$$\left\| u - \frac{1}{|U|} \int_U u(y) \, \mathrm{d}y \right\|_{L^p(U)} \leq C \|Du\|_{L^p(U)},$$

where C only depends on d, p and U.

The proof involves application of Theorem 11.43 and argues by contradiction.

Proof. For the purpose of contradiction, assume that Proposition 11.46 does not hold; then there exist a sequence $u_j \in W^{1,p}(U)$ of nonconstant functions such that

$$\left\| u_j - \frac{1}{|U|} \int_U u_j(y) \, \mathrm{d}y \right\|_{L^p(U)} \geq j \|Du_j\|_{L^p(U)}.$$

Define

$$v_j = \frac{1}{\|u_j - \frac{1}{|U|} \int u_j(y) \, \mathrm{d}y\|_{L^p(U)}} \left(u_j - \frac{1}{|U|} \int u_j(y) \, \mathrm{d}y \right).$$

Then v_j obeys the following properties:

$$\|v_j\|_{L^p(U)} = 1, \quad \|Dv_j\|_{L^p(U)} \leq \frac{1}{j}, \quad \frac{1}{|U|} \int_U v_j(y) \, \mathrm{d}y = 0.$$

In particular, $\{v_j\}$ is a bounded sequence in $W^{1,p}(U)$, so after passing to a subsequence, it converges strongly in L^p to some limit v (when $p = \infty$, this statement follows from Arzela–Ascoli). By the first property,

$$\|v\|_{L^p(U)} = \lim_{j\to\infty} \|v_j\|_{L^p(U)} = 1.$$

On the other hand, since $\|Dv_j\|_{L^p(U)} \to 0$, it follows that $Dv_j = 0$ in the sense of distributions; hence v must be a constant. But then, by the third property,

$$v = \frac{1}{|U|}\int_U v(y)\,\mathrm{d}y = \lim_{j\to\infty} \frac{1}{|U|}\int_U v_j(y)\,\mathrm{d}y = 0,$$

which contradicts $\|v\|_{L^p(U)} = 1$. □

It is interesting that the proof gives the existence of a constant $C > 0$, but no control whatsoever on its size!

Another example, which we already saw, is Theorem 11.26.(1); when $1 \leq p < d$, if $u \in W_0^{1,p}(U)$ (i.e., u is "vanishing on the boundary"), then

$$\|u\|_{L^{p^*}(U)} \leq C\|Du\|_{L^p(U)},$$

where C only depends on d and p. Note also that, as a consequence of Theorem 11.26.(1), for any $1 \leq p < \infty$ and $u \in W_0^{1,p}(U)$, we also have

$$\|u\|_{L^p(U)} \leq C\|Du\|_{L^p(U)}$$

where C only depends on d, p and $|U|$; this is sometimes called *Friedrich's inequality.* A useful strengthening of Friedrich's inequality near the boundary is *Hardy's inequality*:

Proposition 11.47 (Hardy's inequality near a boundary). *Let U be a bounded C^1 domain. For any $1 \leq p < \infty$ and $u \in W_0^{1,p}(U)$, we have*

$$\|\mathrm{dist}(\cdot, \partial U)^{-1} u\|_{L^p(\mathbb{R}^d)} \leq C\|Du\|_{L^p(\mathbb{R}^d)}. \tag{11.14}$$

where C depends only on d, p and U.

Proof. By density, we may assume that $u \in C_c^\infty(U)$. By a smooth partition of unity, boundary straightening and Friedrich's inequality (as in the proof of Proposition 11.13), it suffices to prove the following statement: For $u \in C_c^\infty(\mathbb{R}^d_+)$, we have

$$\|(x^d)^{-1}u\|_{L^p(\mathbb{R}^d_+)} \leq C\|Du\|_{L^p(\mathbb{R}^d_+)}.$$

where C depends only on p.

To prove this, we start with $\|(x^d)^{-1}u\|^p_{L^p(\mathbb{R}^d_+)}$ and compute as follows:

$$\begin{aligned}
&\int_{\mathbb{R}^{d-1}}\int_0^\infty \frac{1}{(x^d)^p}|u(x', x^d)|^p \,\mathrm{d}x^d\mathrm{d}x' \\
&= -\frac{1}{p-1}\int_{\mathbb{R}^{d-1}}\int_0^\infty \partial_{x^d}\frac{1}{(x^d)^{p-1}}|u(x', x^d)|^p \,\mathrm{d}x^d\mathrm{d}x' \\
&= \frac{1}{p-1}\int_{\mathbb{R}^{d-1}}\int_0^\infty \frac{1}{(x^d)^{p-1}}\partial_{x^d}|u(x', x^d)|^p \,\mathrm{d}x^d\mathrm{d}x',
\end{aligned}$$

where the boundary terms vanish by the support assumption on u. We have

$$\left|\frac{1}{p-1}\int_{\mathbb{R}^{d-1}}\int_0^\infty \frac{1}{(x^d)^{p-1}}\partial_{x^d}|u(x', x^d)|^p \,\mathrm{d}x^d\mathrm{d}x'\right|$$

$$
\begin{aligned}
&\leq \frac{p}{p-1} \int_{\mathbb{R}^{d-1}} \int_0^\infty \frac{1}{(x^d)^{p-1}} |u(x', x^d)|^{p-1} |\partial_{x^d} u| \, \mathrm{d}x^d \mathrm{d}x' \\
&\leq \frac{p}{p-1} \|(x^d)^{-(p-1)} u^{p-1}\|_{L^{\frac{p}{p-1}}(\mathbb{R}^d_+)} \|Du\|_{L^p(\mathbb{R}^d_+)} \\
&= \frac{p}{p-1} \|(x^d)^{-1} u\|^{p-1}_{L^p(\mathbb{R}^d_+)} \|Du\|_{L^p(\mathbb{R}^d_+)}.
\end{aligned}
$$

Dividing both sides by $\|(x^d)^{-1}u\|^{p-1}_{L^{\frac{p}{p-1}}(\mathbb{R}^d_+)}$, we obtain the desired inequality. □

Next, we discuss the case when $U = \mathbb{R}^d$. One useful inequality is the Gagliardo–Nirenberg–Sobolev inequality (Theorem 11.25), which states

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

when $1 \leq p < d$ and $u \in W^{1,p}(\mathbb{R}^d)$. Another useful inequality, which does *not* follow from Theorem 11.25, is *Hardy's inequality from infinity*:

Proposition 11.48 (Hardy's inequality from infinity)**.** *Let $1 \leq p < d$ and $u \in W^{1,p}(\mathbb{R}^d)$. Then $r^{-1}u \in L^p(\mathbb{R}^d)$ and*

$$\|r^{-1}u\|_{L^p(\mathbb{R}^d)} \leq C\|Du\|_{L^p(\mathbb{R}^d)}. \tag{11.15}$$

where C depends only on d and p.

Unlike Proposition 11.46, but like Theorem 11.25, note that Hardy's inequality is *homogeneous*. The proof is very similar to that of Proposition 11.47.

Proof. Since $C^\infty_c(\mathbb{R}^d)$ is dense in $W^{1,p}(\mathbb{R}^d)$ by Corollary 11.12, it suffices to prove (11.15) for $u \in C^\infty_c(\mathbb{R}^d)$. We work in the polar coordinates $x = ry$. We begin by performing an integration by parts in r as follows:

$$
\begin{aligned}
\int_{\partial B(0,1)} \int_0^\infty \frac{1}{r^p} |u(ry)|^p r^{d-1} \, \mathrm{d}r \mathrm{d}S(y) &= \frac{1}{d-p} \int_{\partial B(0,1)} \int_0^\infty |u(ry)|^p \partial_r r^{d-p} \, \mathrm{d}r \, \mathrm{d}S(y) \\
&= \frac{1}{d-p} \int_{\partial B(0,1)} |u(ry)|^p r^{d-p} \, \mathrm{d}S(y) \Big|_0^\infty \\
&\quad - \frac{1}{d-p} \int_{\partial B(0,1)} \int_0^\infty \partial_r |u(ry)|^p r^{d-p} \, \mathrm{d}r \, \mathrm{d}S(y).
\end{aligned}
$$

Since u is smooth and compactly supported, and $d - p > 0$ by hypothesis, the boundary term is zero. We estimate the rest as follows:

$$
\begin{aligned}
&\frac{1}{d-p} \left| \int_{\partial B(0,1)} \int_0^\infty \partial_r |u|^p r^{d-p} \, \mathrm{d}r \, \mathrm{d}y \right| \\
&\leq \frac{p}{d-p} \int_{\partial B(0,1)} \int_0^\infty |u|^{p-1} |\partial_r u| r^{d-p} \, \mathrm{d}r \, \mathrm{d}S(y) \\
&\leq \frac{p}{d-p} \int_{\partial B(0,1)} \int_0^\infty r^{-(p-1)} |u|^{p-1} |\partial_r u| r^{d-1} \, \mathrm{d}r \, \mathrm{d}S(y) \\
&\leq \frac{p}{d-p} \|r^{-1}u\|^{p-1}_{L^p(\mathbb{R}^d)} \|Du\|_{L^p(\mathbb{R}^d)}.
\end{aligned}
$$

Dividing both sides by $\|r^{-1}u\|^{p-1}_{L^p(\mathbb{R}^d)}$, we obtain the desired inequality. □

11.8. **Duality and negative regularity Sobolev spaces (optional).** We now turn to the question of identifying the dual spaces of $W^{k,p}_0(U)$ and $W^{k,p}(U)$. As we will soon see, the notion of *negative regularity Sobolev spaces* appears naturally in the process:

Definition 11.49 (Sobolev spaces with negative regularity index)**.** Let k be a nonnegative integer and $1 \leq p < \infty$. We define the *Sobolev space with regularity index $-k$ and integrability index p* by

$$W^{-k,p}(U) = \{u \in \mathcal{D}'(U) : \exists g_\alpha \in L^p(U) \text{ for } |\alpha| \leq k \text{ such that } u = \sum_{\alpha:|\alpha|\leq k} D^\alpha g_\alpha\}.$$

We equip this space with the norm

$$\|u\|_{W^{-k,p}(U)} = \inf_{g_\alpha \in L^p(U): u = \sum_{\alpha:|\alpha|\leq k} D^\alpha g_\alpha} \left(\sum_{\alpha:|\alpha|\leq k} \|g_\alpha\|_{L^p}^p \right)^{\frac{1}{p}}.$$

As usual, we adopt the convention of writing $p' = \frac{p}{p-1}$, so that $(L^p)' = L^{p'}$ for $1 < p < \infty$ by the Riesz representation theorem.

Identification of $(W^{k,p}_0(U))'$. When $1 < p < \infty$, the dual space of $W^{k,p}_0(U)$ turns out to be exactly $W^{-k,p'}(U)$:

Proposition 11.50. *Let k be a nonnegative integer and $1 < p < \infty$. For any domain U,*

$$(W^{k,p}_0(U))' = W^{-k,p'}(U),$$

where $u \in W^{-k,p'}(U)$ defines a linear functional on $W^{k,p}_0(U)$ by $C^\infty_c(U) \ni v \mapsto \langle u, v\rangle$, whose norm is equal to $\|u\|_{W^{-k,p'}(U)}$.

Proof. Let us start with the left inclusion $\supseteq$, which is easier (but this part breaks down for $W^{k,p}(U)$!). Let $u \in W^{-k,p'}(U)$, which by definition admits a decomposition of the form $u = \sum_{\alpha:|\alpha|\leq k} g_\alpha$ with $g_\alpha \in L^{p'}(U)$. Then for $v \in C^\infty_c(U)$,

$$\begin{aligned} \langle u, v\rangle &= \sum_{\alpha:|\alpha|\leq k} \int D^\alpha g_\alpha v \,\mathrm{d}x \\ &= \sum_{\alpha:|\alpha|\leq k} (-1)^{|\alpha|} \int g_\alpha D^\alpha v \,\mathrm{d}x, \end{aligned}$$

so

$$|\langle u, v\rangle| \leq \sum_{\alpha:|\alpha|\leq k} \|g_\alpha\|_{L^{p'}(U)} \|D^\alpha v\|_{L^p(U)} \leq C\|u\|_{W^{-k,p'}(U)} \|v\|_{W^{k,p}(U)},$$

where $C > 1$ can be taken to be arbitrarily close 1. Hence u defines a bounded linear functional on $W^{k,p}_0(U)$ (the closure of $C^\infty_c(U)$ with respect to $\|\cdot\|_{W^{k,p}(U)}$), whose norm does not exceed $\|u\|_{W^{-k,p'}(U)}$.

For the right inclusion $\subseteq$, we make an argument involving the Hahn–Banach theorem and the Riesz representation theorem. Let us enumerate all multi-indices α with $|\alpha| \leq k$ as $\alpha_0, \alpha_1, \ldots, \alpha_K$ (where $K = \sum_{j=0}^k \frac{d!}{j!(d-j)!}$). For $v \in C^\infty_c(U)$, consider the mapping

$$v \mapsto Tv := (D^{\alpha_0} v, D^{\alpha_1} v, \ldots, D^{\alpha_K} v) \in L^p(U)^{\oplus K}.$$

Note that T defines an injective map from $C_c^\infty(U)$ into the above direct sum of $L^p(U)$'s. Moreover, if $L^p(U)^{\oplus K}$ is equipped with the norm $\|(v_{\alpha_0}, \dots, v_{\alpha_K})\| = (\sum_{j=0}^K \|v_{\alpha_j}\|_{L^p(U)}^p)^{\frac{1}{p}}$, then T is an isometry. Hence any linear functional $u \in (W_0^{k,p}(U))'$ defines a bounded linear functional $\tilde{u}$ on $T(C_c^\infty(U))$ by $\tilde{u}(Tv) := u(v)$. By the Hahn–Banach theorem, $\tilde{u}$ extends to a bounded linear functional on $L^p(U)^{\oplus K}$ with the same bound, so by the Riesz representation theorem there exist $\tilde{g}_{\alpha_0}, \dots, \tilde{g}_{\alpha_K} \in L^{p'}(U)$ such that

$$\tilde{u}(Tv) = \sum_{j=0}^K \langle \tilde{g}_{\alpha_j}, D^{\alpha_j} v \rangle,$$

for every $v \in C_c^\infty(U)$, where

$$\left(\sum_{j=0}^K \|\tilde{g}_{\alpha_j}\|_{L^{p'}(U)}^{p'} \right)^{\frac{1}{p'}} \leq \sup_{v \in C_c^\infty(U) : \|v\|_{W^{1,p}(U)} \leq 1} |\tilde{u}(v)| = \|u\|_{(W^{k,p}(U))'}.$$

Defining $g_{\alpha_j} = (-1)^{|\alpha_j|} \tilde{g}_{\alpha_j}$, it follows that

$$u(v) = \tilde{u}(Tv) = \sum_{j=0}^K (-1)^{|\alpha_j|} \langle g_{\alpha_j}, D^{\alpha_j} v \rangle$$

for every $v \in C_c^\infty(U)$, i.e., $u = \sum_{j=0}^K D^{\alpha_j} g_{\alpha_j}$ as distributions. Hence, $u \in W^{-k,p'}(U)$ with $\|u\|_{W^{-k,p'}(U)} \leq \|u\|_{(W^{k,p}(U))'}$, as desired. □

Identification of $(W^{k,p}(U))'$ *(optional).* Next, we turn to the question of identifying $(W^{k,p}(U))'$. Note that, in general, it *cannot* be expressed as a subspace of $\mathcal{D}'(U)$, since $C_c^\infty(U)$ is not dense in $W^{k,p}(U)$. Instead, $(W^{k,p}(U))'$ turns out to be a closed subspace of $W^{-k,p}(\mathbb{R}^d)$ consisting of elements whose support is in $\overline{U}$.

To facilitate the statement of the result, let us introduce a notation. Let $k \in \mathbb{Z}$ and $1 \leq p < \infty$. Given a closed subset K of a domain U, we write

$$W_K^{k,p}(U) = \{u \in W^{k,p}(U) : \operatorname{supp} u \subseteq K\}.$$

Note that $W_K^{k,p}(U)$ is a closed subspace of $W^{k,p}(U)$.

Proposition 11.51. *Let k be a nonnegative integer and $1 < p < \infty$. Let U be a bounded C^k domain. Then*

$$(W^{k,p}(U))' = W_{\overline{U}}^{-k,p'}(\mathbb{R}^d).$$

We need the following result from functional analysis, whose straightforward proof we omit.

Lemma 11.52. *Let X be a Banach space, and let Y be a closed subspace of Y. Denote by $Y^\perp$ the subspace of X' consisting of bounded linear functionals whose kernel contains Y, i.e.,*

$$Y^\perp = \{u \in X' : u(v) = 0 \text{ for all } v \in Y\}.$$

Then the following relations hold:

$$Y' = X'/Y^\perp, \tag{11.16}$$

$$(X/Y)' = Y^\perp. \tag{11.17}$$

A sketch of the proof of Proposition 11.51 is as follows. By Proposition 11.13, it can be shown that

$$W^{k,p}(U) = W^{k,p}(\mathbb{R}^d)/W^{k,p}_{\mathbb{R}^d\setminus U}(\mathbb{R}^d).$$

By Lemma 11.52 with $X = W^{k,p}(\mathbb{R}^d)$ and $Y = W^{k,p}_{\mathbb{R}^d\setminus U}(\mathbb{R}^d)$, we have

$$(W^{k,p}(U))' = (W^{k,p}_{\mathbb{R}^d\setminus U}(\mathbb{R}^d))^{\perp}.$$

We claim that

$$(W^{k,p}_{\mathbb{R}^d\setminus U}(\mathbb{R}^d))^{\perp} = W^{-k,p'}_{\overline{U}}(\mathbb{R}^d).$$

The right inclusion $\subseteq$ is not difficult to show. For the left inclusion $\supseteq$, we need to show that if $u \in W^{-k,p'}(\mathbb{R}^d)$ with $\operatorname{supp} u \subseteq \overline{U}$ and $v \in W^{k,p}(\mathbb{R}^d)$ with $\operatorname{supp} v \subseteq \mathbb{R}^d \setminus U$, then $u(v) = 0$. To prove this, we need to find an approximating sequence v_ϵ such that $\operatorname{supp} v_\epsilon \subseteq \mathbb{R}^d \setminus \overline{U}$ and $v_\epsilon \to v$ in $W^{k,p}(U)$. If U is sufficiently regular, then such a sequence can be constructed by a boundary straightening and translating argument (cf. the proof of Proposition 11.10).

Remark 11.53. The two spaces $(W^{k,p}_0(U))'$ and $(W^{k,p}(U))'$ are related to each other as follows:

$$W^{-k,p'}(U) = W^{-k,p'}_{\overline{U}}(\mathbb{R}^d)/W^{-k,p'}_{\partial U}(\mathbb{R}^d)$$

This identity is a quick consequence of Lemma 11.52 with $X = W^{k,p}(U)$ and $Y = W^{k,p}_0(U)$.

References

[Dah79] B. E. J. Dahlberg. On the Poisson integral for Lipschitz and C^1-domains. *Studia Math.*, 66(1):13–24, 1979.

[EG15] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015.

[Eva10] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[Fol99] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[GT01] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[Ḧ03] L. Hörmander. *The analysis of linear partial differential operators. I*. Classics in Mathematics. Springer-Verlag, Berlin, 2003. Distribution theory and Fourier analysis, Reprint of the second (1990) edition [Springer, Berlin; MR1065993 (91m:35001a)].

[O'N83] B. O'Neill. *Semi-Riemannian geometry*, volume 103 of *Pure and Applied Mathematics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1983. With applications to relativity.

[Rud91] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw Hill, second edition, 1991.

[Sim15] L. Simon. Lectures on PDE, 2015.

[Ste70] E. M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[Ste93] E. M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.

[Tyc35] A. Tychonoff. Théorèmes d'unicité pour l'équation de la chaleur. *Mat. Sb.*, 42(2):199–216, 1935.

Department of Mathematics, UC Berkeley, Berkeley, CA 94720
Email address: sjoh@math.berkeley.edu