Lecture Notes 2010

Sergiu Klainerman

Department of Mathematics, Princeton University, Princeton NJ 08544

E-mail address: seri@math.princeton.edu

1. INTRODUCTION

The world of Partial Differential Equations

To start with partial differential equations, just like ordinary differential or integral equations, are *functional equations*. That means that the unknown, or unknowns, we are trying to determine are functions. In the case of partial differential equations (PDE) these functions are to be determined from equations which involve, in addition to the usual operations of addition and multiplication, partial derivatives of the functions. Below are the most basic examples,

- (Laplace equation)

$$\Delta u = 0 \tag{1}$$

where $\Delta u = \frac{\partial^2}{\partial x^2}u + \frac{\partial^2}{\partial y^2}u + \frac{\partial^2}{\partial z^2}u$. The other two examples described in the section of fundamental mathematical definitions are

- (Heat Equation)

$$-\partial_t u + k\Delta u = 0, \tag{2}$$

- (Wave equation)

$$-\partial_t^2 u + c^2\Delta u = 0. \tag{3}$$

In both cases one is asked to find a function u, depending on the variables t, x, y, z, which verifies the corresponding equations. Observe that both (2) and (3) involve the symbol Δ which has the same meaning as in the first equation, that is $\Delta u = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})u = \frac{\partial^2}{\partial x^2}u + \frac{\partial^2}{\partial y^2}u + \frac{\partial^2}{\partial z^2}u$. In both (2) and (3) $k > 0$ and c are fixed constants (representing the rate of diffusion for the first and the speed of light in the second). It suffices to study to solve the equations for the special cases $k = 1$ and $c = 1$. Indeed if $u(t, x, y, z)$ is a solution of (3), for example, then $v(t, x, y, z) = u(t, x/c, y/c, z/c)$ verifies the same equation with $c = 1$. Both equations are called *evolution equations*, simply because they are supposed to describe the change relative to the time parameter t of a particular physical object. Observe that (1) can be interpreted as a particular case of both (3) and (2). Indeed solutions $u = u(t, x, y, z)$ of either (3) or (2) which are independent of t, i.e. $\partial_t u = 0$, verify (1).

Here are some further examples of important PDEs:

- (Schrödinger equation)

$$i\partial_t u + k\Delta u = 0 \tag{4}$$

with $u : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{C}$. The equation describes the quantum evolution of a massive particle, $k = \frac{\hbar}{2m}$, where $\hbar > 0$ is Planck's constant and m is the mass of the particle. As with the heat equation, one can normalize $k = 1$ by a simple change of variables. Though the equation is formally very similar to the heat equation, it has very different qualitative behavior. It is important to keep in mind when studying PDE's that small changes in the form of an equation can lead to very different properties of solutions.

- (Klein-Gordon equation)

$$-\partial_t^2 u + c^2 \Delta u - \left(\frac{mc^2}{\hbar}\right)^2 u = 0 \tag{5}$$

This is the relativistic counterpart to the Schrödinger equation, the parameter m has the physical interpretation of mass and mc^2 has the physical interpretation of rest energy (reflecting Einstein's famous equation $E = mc^2$). One can normalize the constants c and $\frac{mc^2}{\hbar}$ to make them both equal 1 by applying a suitable change of variables to both time and space.

Observe that all three PDE mentioned above satisfy the following simple property called the *principle of superposition*: If u_1, u_2 are solutions of an equation so is any linear combination of them $\lambda_1 u_1 + \lambda_2 u_2$ where λ_1 and λ_2 are arbitrary real numbers. Such equations are called linear. The following equation in the unknown $u = u(x, y)$, is manifestly not linear:

- (Minimal surfaces

$$\partial_x\Big(\frac{\partial_x u}{(1 + |\partial_x u|^2 + |\partial_y u|^2)^{\frac{1}{2}}}\Big) + \partial_y\Big(\frac{\partial_y u}{(1 + |\partial_x u|^2 + |\partial_y u|^2)^{\frac{1}{2}}}\Big) = 0. \tag{6}$$

Here ∂_x and ∂_y are short hand notations for the partial derivatives $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$.

The equations we have encountered so far can be written in the form $\mathcal{P}[u] = 0$, where $\mathcal{P}$ is a *differential operator* applied to u. A differential operator is simply a rule which takes functions u, defined in $\mathbb{R}^n$ or an open subset of it, into functions $\mathcal{P}[u]$ by performing the following operations:

- We can take partial derivatives $\partial_i u = \frac{\partial u}{\partial x^i}$ relative to the variables $x = (x^1, x^2, \ldots x^n)$ of $\mathbb{R}^n$. One allows also higher partial derivatives of u such as the mixed second partials $\partial_i \partial_j u = \frac{\partial^2 u}{\partial x^i \partial x^j}$ or $\partial_i^2 = \frac{\partial^2}{\partial x_i^2}$.

  The associated differential operators for (2) is $\mathcal{P} = -\partial_t + \Delta$ and that of (3) is $-\partial_t^2 + \Delta$
- Can add and multiply u and its partial derivatives between themselves as well as with given functions of the variables x. Composition with given functions may also appear.

In the case of the equation (1) the associated differential operator is $\mathcal{P} = \Delta = \partial_1^2 + \partial_2^2 + \partial_3^2 = \sum_{i,j=1}^3 e^{ij} \partial_i \partial_j$ where e^{ij} is the diagonal 3×3 matrix with entries

$(1,1,1)$ corresponding to the euclidean scalar product of vectors X, Y in $\mathbb{R}^3$,

$$< X, Y >= X_1Y_1 + X_2Y_2 + X_3Y_3 = \sum_{i,j=1}^{3} e^{ij} X_i X_j. \tag{7}$$

The associated differential operators for (2), (3) and (4) are, resp. $\mathcal{P} = -\partial_t + \Delta$, $\mathcal{P} = -\partial_t^2 + \Delta$ and $\mathcal{P} = i\partial_t + \Delta$ with variables are $t, x^1, x^2, x^3 \in \mathbb{R}^{1+3}$. In the particular case of the wave equation (3) it pays to denote the variable t by x^0. The wave operator can then be written in the form,

$$\square = -\partial_0^2 + \partial_1^2 + \partial_2^2 + \partial_3^2 = \sum_{\alpha,\beta=0}^{3} m^{\alpha\beta}\partial_\alpha\partial_\beta \tag{8}$$

where $m^{\alpha\beta}$ is the diagonal 4×4 matrix with entries $(-1,1,1,1)$, corresponding to the Minkowski scalar product in $\mathbb{R}^{1+3}$. This latter scalar product is defined, for 4 vectors $X = (X_0, X_1, X_2, X_3)$ and $Y = (Y_0, Y_1, Y_2, Y_3)$ by,

$$m(X, Y) = \sum_{\alpha,\beta=0}^{3} m^{\alpha\beta} X_\alpha Y_\beta = -X_0Y_0 + X_1Y_1 + X_2Y_2 + X_4Y_4 \tag{9}$$

The differential operator $\square$ is called D'Alembertian after the name of the French mathematician who has first introduced it in connection to the equation of a vibrating string.

Observe that the differential operators associated to the equations (1)–(4) are all linear i.e.

$$\mathcal{P}[\lambda u + \mu v] = \lambda\mathcal{P}[u] + \mu\mathcal{P}[v],$$

for any functions u, v and real, or complex, numbers λ, μ. The following is another simple example of a linear differential operator

$$\mathcal{P}[u] = a_1(x)\partial_1 u + a_2(x)\partial_2 u \tag{10}$$

where $x = (x_1, x_2)$ and a_1, a_2 are given functions of x. They are called the *coefficients* of the linear operator. An equation of the form

$$\mathcal{P}[u] = f, \tag{11}$$

corresponding to a linear differential operator $\mathcal{P}$ and a given function $f = f(x)$, is called *linear-inhomogeneous* . Any solution u of such an equation can be expressed in the form $u = u_0 + v$ where u_0 is a special solution of (11) and v solution to the homogeneous equation

$$\mathcal{P}[v] = 0. \tag{12}$$

In the case of the equation (6) the differential operator $\mathcal{P}$ can be written, relative to the variables x^1 and x^2, in the form,

$$\mathcal{P}[u] = \sum_{i=1}^{2} \partial_i\Big(\frac{1}{(1 + |\partial u|^2)^{\frac{1}{2}}}\partial_i u\Big),$$

where $|\partial u|^2 = (\partial_1 u)^2 + (\partial_2 u)^2$. Clearly $\mathcal{P}[u]$ is not linear in this case. We call it a nonlinear operator; the corresponding equation (6) is said to be a *nonlinear equation.* An important property of both linear and nonlinear differential operators is *locality*. This means that whenever we apply $\mathcal{P}$ to a function u, which vanishes in some open set D, the resulting function $\mathcal{P}[u]$ also vanish in D.

Observe also that our equations (1)-(4) are also translation invariant. This means, in the case (1) for example, that whenever the function $u = u(x)$ is a solution so is the function $u_c(x) := u(T_c x)$ where T_c is the translation $T_c(x) = x + c$. On the other hand the equation $\mathcal{P}[u] = 0$, corresponding to the operator $\mathcal{P}$ defined by (10) is not, unless the coefficients a_1, a_2 are constant. Clearly the set of invertible transformations[1] $T : \mathbb{R}^n \to \mathbb{R}^n$ which map any solution $u = u(x)$, of $\mathcal{P}[u] = 0$, to another solution $u_T(x) := u(Tx)$ form a group, called the invariance group of the equation. The composition of two symmetries is again a symmetry, as is the inverse of a symmetry, and so it is natural to view a collection of symmetries as forming a GROUP (which is typically a finite or infinite-dimensional LIE GROUP).

The Laplace equation (1) is invariant not only with respect to translations but also rotations, i.e linear transformations $O : \mathbb{R}^3 \to \mathbb{R}^3$ which preserve the euclidean scalar product (7) in the sense that $< OX, OY >=< X, Y >$ for all vectors $X, Y \in \mathbb{R}^3$. Similarly the wave equation (3) and Klein-Gordon equation (5) are invariant under Lorentz transformations, i.e. linear transformations $L : \mathbb{R}^{1+3} \to \mathbb{R}^{1+3}$ which preserve the Minkowski scalar product (9), i.e. $m(LX, LY) = m(X, Y)$. Our other evolution equations (2) and (4) are clearly invariant under rotations of the space variables $x = (x^1, x^2, x^3) \in \mathbb{R}^3$, keeping t fixed. They are also Galilean invariant, which means, in the particular case of the Schrödinger equation (180), that whenever $u = u(t, x)$ is a solution so is $u_v(t, x) = e^{i(x\cdot v)} e^{it|v|^2} u(t, x - vt)$ for any vector $v \in \mathbb{R}^3$.

So far we have tacitly assumed that our equations take place in the whole space – $\mathbb{R}^3$ for the Laplace equation, $\mathbb{R}^4$ for the Heat, Wave and Schrödinger equations and $\mathbb{R}^2$ for the minimal surface equation. In reality, one is often restricted to a domain of the corresponding space. Thus, for example, the equation (1) is usually studied on a bounded open domain of $\mathbb{R}^3$ subject to a specified *boundary condition*. Here is a typical example.

Example. The Dirichlet problem on an open domain of $D \subset \mathbb{R}^3$ consists of finding a continuous functions u defined on the closure $\bar{D}$ of D, twice continuously differentiable in D, such that $\Delta u = 0$ in D and the restriction of u to ∂D, the boundary of D, is prescribed to be a continuous function u_0. More precisely we require that,

$$u|_{\partial D} = u_0 \tag{13}$$

One can impose the same boundary condition for solutions of (6), with D a bounded open domain of $\mathbb{R}^2$. A solution $u = u(x, y)$ of (6) in D, verifying the boundary condition (13), solves the Plateau problem of finding minimal surfaces in $\mathbb{R}^3$ which pass through a given curve. One can show that the surface given by the graph $\Gamma_u = \{(x, y, u(x, y))/(x, y) \in D \subset \mathbb{R}^2\}$ has minimum area among all other graph surfaces Γ_v verifying the same boundary condition, $v|_{\partial D} = u_0$.

[1] The transformations are often linear maps.

Natural boundary conditions can also be imposed for the evolution equations (2)–(4). The simplest one is to prescribe the values of u on the hyperplane $t = 0$. In the case of the heat and Schrödinger equation we set,

$$u|_{t=0} = u_0$$

while in the case of the wave equation, which involves a second derivative in t, we impose two conditions

$$u|_{t=0} = u_0 \text{ and } \partial_t u|_{t=0} = u_1 \tag{14}$$

where u_0, u_1 are functions of the coordinates (x, y, z), called *initial conditions*. To solve the *initial value problem* in both cases means to find solutions of the equations for $t > 0$ which verify the corresponding initial conditions at $t = 0$. In addition one may restrict the variables (x, y, z) to an open domain of $D \subset \mathbb{R}^3$. More to the point one may try to solve a boundary value problem in a domain $[0, \infty) \times D$ with a boundary condition, such as (13), on $[0, \infty) \times \partial D$ and an initial condition at $t = 0$.

The choice of boundary condition and initial conditions, for a given PDE, is very important. Finding which are the good boundary and initial conditions is an important aspect of the general theory of PDE which we shall address in section 2. For equations of physical interest these appear naturally from the context in which they are derived. For example, in the case of a vibrating string, which is described by solutions of the one dimensional wave equation $\partial_t^2 u - \partial_x^2 u = 0$ in the domain $(a, b) \times \mathbb{R}$, the initial conditions $u = u_0$, $\partial_t u = u_1$ at $t = t_0$, amount to specifying the original position and velocity of the string. On the other hand the boundary condition $u(a) = u(b) = 0$ simply mean that the two ends of the of the string are fixed.

So far we have only considered equations in *one unknown*. In reality many of the equations of interest appear as systems of partial differential equations. The following important example, contains two unknown functions $u_1 = u_1(x^1, x^2), u_2 = u_2(x^1, x^2)$ which verify,

- (Cauchy-Riemann)

$$\partial_1 u_2 - \partial_2 u_1 = 0, \qquad \partial_1 u_1 + \partial_2 u_2 = 0 \tag{15}$$

It was first observed by Cauchy that $u = u_1 + iu_2$, as a function of $z = x^1 + ix^2$, is a complex analytic function if and only if (15) is satisfied. Setting also $\bar{\partial} = \partial_1 + i\partial_2$, observe that (15) is equivalent to

$$\bar{\partial} u = 0. \tag{16}$$

Equation (15) can also be written in the form $\mathcal{P}[u] = 0$ by introducing $u = (u_1, u_2)$ as a column vector and $\mathcal{P}[u]$ the differential operator,

$$\mathcal{P}[u] = \begin{pmatrix} -\partial_2 & \partial_1 \\ \partial_1 & \partial_2 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

The system of equations (15) contains two equations and two unknowns. This is the standard situation of a *determined system*. A system is called *over-determined* if it contains more equations than unknowns and *underdetermined* if it contains fewer equations than unknowns. For example the system of two equations and one unknown $\partial_x u(x,y) = f, \partial_y u(x,y) = g$ is clearly overdetermined. A necessary condition for a solution to exist is $\partial_y f = \partial_x g$, condition which can be interpreted as requiring that the one-form $w = f(x,y)dx + g(x,y)dy$ is exact, i.e. its exterior derivative $d\omega$ is identically zero. Overdetermined systems, such as De Rham complexes, play a very important role in geometry.

Observe that (15) is a linear system. Observe also that the operator $\mathcal{P}$ has the following remarkable property.

$$\mathcal{P}^2[u] = \mathcal{P}[\mathcal{P}[u]] = \begin{pmatrix} \Delta u_1 \\ \Delta u_2 \end{pmatrix}$$

In other words $\mathcal{P}^2 = \Delta \cdot I$, with I the identity operator $I[u] = u$, and therefore $\mathcal{P}$ can be viewed as a a square root of Δ. One can define a similar type of square root for the D'Alembertian $\Box$. To achieve this we need 4×4 complex matrices $\gamma^0, \gamma^1, \gamma^3, \gamma^4$ which satisfy the property

$$\gamma^\alpha \gamma^\beta + \gamma^\beta \gamma^\alpha = -2m^{\alpha\beta} I \tag{17}$$

with I the unit 4×4 matrix and $m^{\alpha\beta}$ as in (8). Using the γ matrices we can introduce the Dirac operator acting on $u = (u_1, u_2, u_3, u_4)$ defined from $\mathbb{R}^{1+3}$ with values in $\mathbb{C}^4$ by,

$$\mathcal{D}u = i\gamma^\alpha \partial_\alpha u \tag{18}$$

Using (17) we easily check that, $\mathcal{D}^2 u = \Box u$. Thus the Dirac operator $\mathcal{D}$ can be viewed as a square root of the D'Alembertian $\Box$. It leads to the following fundamental equation introduced by Dirac as the equation of free, massive, relativistic, particle such as the electron:

- (Dirac Equation)

$$\mathcal{D}u = ku \tag{19}$$

Partial differential equations are ubiquitous throughout Mathematics and Science. They provide the basic mathematical framework for some of the most important physical theories, such as Elasticity, Hydrodynamics, Electromagnetism, General Relativity and Non-relativistic Quantum Mechanics. The more modern relativistic quantum field theories lead, in principle, to equations in infinite number of unknowns, which lie beyond the scope of partial differential equations. Yet, even in that case, the basic equations preserve the locality property of PDE. Moreover the starting point of a quantum field theory is always a classical field theory, described by systems of PDE's. This is the case, for example, of the Standard Model of weak and strong interactions, based on a Yang -Mills-Higgs field theory. If we also include the ordinary differential equations of Classical Mechanics, which can be viewed as one dimensional PDE, we see that, essentially, all of Physics is described by differential equations. Other examples of partial differential equations underlining some of our most basic physical theories, are the Maxwell, Einstein, Euler and Navier Stokes equations. Note that each equation are at the heart of an entire field of Physics, i.e. Electrodynamics, General Relativity and Hydrodynamics,

An important feature of the main PDEs appearing in Physics is their apparent universality. Thus, for example, the wave equation, first introduced by D'alembert to describe the motion of a vibrating string was later found to be connected to the propagation of sound and electromagnetic waves. The heat equation, first introduced by Fourier to describe heat propagation, also comes in many other situations in which dissipative effects play an important role. The same thing can be said about the Laplace, Schrödinger and many other basic equations.

It is even more surprising that equations, originally introduced to describe specific physical phenomena, also play a fundamental role in areas of mathematics, which are considered pure, such as Complex Analysis, Differential Geometry, Topology and Algebraic Geometry. Complex Analysis, for example, which studies the properties of holomorphic functions, can be regarded as the study of solutions to the Cauchy-Riemann equations (15) in a domain of $\mathbb{R}^2$. Hodge theory, which plays a fundamental role in topology and algebraic geometry, is based on studying the space of solutions to a class of linear systems of partial differential equations on manifolds which generalize the Cauchy-Riemann equations. The Atiyah-Singer index theorem is formulated in terms of a special classes of linear PDE on manifolds, related to the euclidean version of the Dirac operator (18).[2]

Important problems in geometry can be reduced to finding solutions to specific partial differential equations, typically nonlinear. We have already seen such an example in the case of the Plateau problem of finding surfaces of minimal total area which pass through a given curve. The well known uniformization theorem provides another excellent example.

To state the uniformization theorem, we need to recall the definition of a compact Riemann surface S. This is a 2 -dimensional, compact manifold endowed with a smooth, positive definite metric g. The Gauss curvature $K = K(g)$ is an important invariant of the surface which can be calculated explicitely at every point $p \in S$ in terms of the components g_{ab} relative to a local system of coordinates $x = (x^1, x^2)$ near p. The calculation involves first and second partial derivatives of the components g_{ab} relative to x^1 and x^2. The remarkable fact is that the final value of K does not depend on the particular system of coordinates in which one makes the calculation. Moreover in the particular case when S is the standard sphere in $\mathbb{R}^3$, given by the equation $|x|^2 = a^2$, the Gauss curvature is equal to the expected value, corresponding to our intuition of curvature, that is $K = a^{-2}$. Another remarkable property of the Gauss curvature is that its total integral along S does not depend on the metric g but only on the topological properties of S. More precisely, according to the Gauss-Bonnet formula, we have

$$\chi(S) = (2\pi)^{-1} \int_S K da_g,$$

with da_g denoting the area element of the metric g. In coordinates x^1, x^2 we have $da_g = \sqrt{|g|} dx^1 dx^2$ with $|g|$ the determinant of the matrix $(g_{ab})_{a,b=1,2}$. The number $\chi(S)$ is one of the integers $2, 0, -2, \ldots - 2k \ldots$, called the Euler characteristic of S, and has simple topological interpretation. Thus any surface which can be continuously deformed to the standard sphere has $\chi(S) = 2$ while any surface which

[2]This is the operator obtained when we change the minkowski metric m to the euclidean one e in (17).

can be continuously deformed to a torus has $\chi(S) = 0$. We can now state the uniformization theorem:

Theorem 1.1. *Let S be a 2-dimensional, compact, Riemann surface with metric g, Gauss curvature $K = K(g)$ and Euler characteristic $\chi(S)$. There exists a conformal transformation of the metric g, i.e. $\tilde{g} = \Omega^2 g$, for some smooth non-vanishing function Ω, such that the Gauss curvature $\tilde{K}$ of the new metric $\tilde{g}$ is identical equal to 1, 0 or -1 according to whether $\chi(S) > 0$, $\chi(S) = 0$ or $\chi(S) < 0$.*

To prove this very important geometric result, which leads to the complete classification of all compact surfaces according to their Euler characteristic, we are led to a nonlinear partial differential equation on S. Indeed assume that $\chi(S) = 2$ and therefore we want the Gauss curvature $\tilde{K}$ of the metric $\tilde{g} = e^{2u} g$ to be exactly 1. It is easy to relate $\tilde{K}$, by a simple calculation, to the Gauss curvature K of the original metric g . This leads to the following equation in u,

$$\Delta_S u + e^{2u} = K \tag{20}$$

where Δ_S, called the Laplace-Beltrami operator of S, is a straightforward adaptation of the Laplace operator, see (1), to the surface S. Thus the proof of the uniformization theorem reduces to solve equation (20), i.e. for a given surface S with Gauss curvature K, find a real valued function u which verifies (20).

We give below a precise definition of the operator Δ_S relative to a system of local coordinates $x = (x^1, x^2)$ on an open coordinate chart $D \subset S$. Denote by $G(x) = (g_{ab}(x))_{a,b=1,2}$ the 2×2 matrix whose entries are the components of our Riemannian metric on D. Let $G^{-1}(x)$ denote the matrix inverse to $G(x)$ and denote its components by $(g^{ab}(x))_{a,b=1,2}$. Thus, for all $x \in D$,

$$\sum_c g_{ac}(x) g^{cb}(x) = \delta_{ab}$$

with δ_{ab} the usual Kronecker symbol. We also set, as before, $|g(x)| = \det(G(x))$ and define,

$$\Delta_S u(x) = \frac{1}{\sqrt{|g(x)|}} \sum_{a,b=1,2} \partial_b(\sqrt{|g(x)|}\, g^{ab}(x)\, \partial_a u(x))$$

Typically we suppress the explicit dependence on x in the above formula. It is also very convenient to use Einstein's summation convention over repeated indices, and thus write,

$$\Delta_S u = \frac{1}{\sqrt{|g|}} \partial_b(\sqrt{|g|}\, g^{ab}\, \partial_a u) \tag{21}$$

As a third example we consider the Ricci flow equation on a compact n dimensional manifold M, which is described in one of the articles of the Compendium. In the particular case of three dimensions the equation has been recently used, decisively, to provide the first proof of Thurston's geometrization conjecture, including the well known Poincaré conjecture. The geometrization conjecture, described in the topology section of the Compendium, is the precise analogous, in three space dimensions, of the 2-dimensional uniformization theorem mentioned above. The

Ricci flow is defined, in arbitrary local coordinates $x = (x^1, x^2, x^3)$ on M, by the equation:

- (Ricci Flow)

$$\partial_t g_{ij} = R_{ij}(g) \tag{22}$$

Here $g_{ij} = g_{ij}(t)$ is a family of Riemannian metrics depending smoothly on the parameter t and $R_{ij}(g)$ denotes the Ricci curvature of the metric g_{ij}. This is simply a three dimensional generalization of the Gauss curvature we have encountered in the uniformization theorem. In a given system of coordinates $R_{ij}(g)$ can be calculated in terms of the metric coefficients g_{ij} and their first and second partial derivatives. Since both g_{ij} and R_{ij} are symmetric relative to $i, j = 1, 2, 3$ we can interpret (22) as a non-linear system of six equations with six unknowns. On a closer look it turns out that (22) is related to the heat equation (2). Indeed, by a straightforward calculation relative to a particular system of coordinates $x = (x^1, x^2, x^2)$ called *harmonic*, it can be shown that the Ricci flow (22) takes the form

$$\partial_t g_{ij} - \Delta_g g_{ij} = N_{ij}(g, \partial g) \tag{23}$$

where each N_{ij}, $i, j = 1, 2, 3$, are functions of the components g_{ij} and their first partial derivatives with respect to the coordinates x and Δ_g is, again, a differential operator very similar to the Laplacian Δ in $\mathbb{R}^3$, see (**??**). More precisely, if $G^{-1} = (g^{ab})_{a,b=1,2,3}$ denotes the matrix inverse to $G = (g_{ab})_{a,b=1,2,3}$ we can write, using the summation convention,

$$\Delta_g = g^{ab}\partial_a\partial_b = \sum_{a,b=1}^{3} g^{ab}\partial_a\partial_b.$$

In a small neighborhood of a point $p \in M$ we can choose the harmonic coordinate x^a such that $g^{ab}(p) = \delta^{ab}$ with δ^{ab} denoting the usual Kronecker symbol. Thus, near p, Δ_g looks indeed like $\Delta = \delta^{ab}\partial_a\partial_b$.

The Ricci flow[3] allows one to deform an arbitrary Riemannian metric on M to a a simple metric of constant sectional curvature. The idea is to start with a metric g and look for solutions $g(t)$ of (22) which verify the *initial condition* $g(0) = g$. One hopes that as $t \to \infty$ the metric $g(t)$ will converge to a metric of constant curvature. Intuitively one can see this working out the same way heat gets evenly distributed in space, as t increases, according to the heat equation (2). Indeed since (22) is similar to (2) we expect the variations in the curvature of $g(t)$ to become smaller and smaller as the metric evolves according to (22). The type of metric we get in the limit as $t \to \infty$ will allow us to determine the topological character of M. The flow, however, can *develop singularities* before we achieve that goal. To overcome this major technical difficulty one needs to make a detailed qualitative analysis of the behavior of solutions to (22), task which requires just about all the advances made in geometric PDE in the last hundred years.

[3]In reality one needs to change the equation (22) slightly to make sure that the total volume of of M, calculated with respect to the metric $g(t)$, stays constant.

As we have seen above the choice of harmonic coordinates allows us to write the Ricci flow as a system of *nonlinear heat equations* (23). This fact is quite typical to geometric equations. It is useful at this point to discuss another important example, that of the Einstein equations in vacuum. An introduction to this equation and short discussion of its importance in General Relativity can be found (see compendium article). Solutions to the Einstein vacuum equations are given by Ricci flat *spacetimes*, that is Lorentzian manifolds (M, g) with M a four dimensional manifold and g a Lorentz metric on it, for which the corresponding Ricci curvature vanishes identically.

- (EINSTEIN-VACUUM)

$$\mathrm{Ric}(g) = 0. \tag{24}$$

The Ricci curvature of a Lorentz metric, $\mathrm{Ric}(g)$, can be defined in exactly the same way as in the Riemannian case. Thus relative to a coordinate system x^α, with $\alpha = 0, 1, 2, 3$, the Ricci curvature, denoted by $R_{\alpha\beta}$, can be expressed in terms of the first and second partial derivatives of the metric coefficients $g_{\alpha\beta}$. As before, we denote by $g^{\alpha\beta}$ the components of the inverse metric. Moreover, by picking a specified system of coordinates, called wave coordinates[4], we can express the Einstein-vacuum equations (24) in the form of a system of equations related to the wave equation (3), in the same way the Ricci flow system (23) was related to the heat equation (2). More precisely,

$$\Box_g \, g_{\alpha\beta} = N_{\alpha\beta}(g, \partial g) \tag{25}$$

where, as in the case of the Ricci flow, the terms $N_{\alpha\beta}(g, \partial g)$ are expressions, which can be calculated explicitely, depending on the metric $g_{\alpha\beta}$, its inverse $g^{\alpha\beta}$ and the first derivatives of $g_{\alpha\beta}$ relative to the coordinates x^α. This is a system of 10 equations with respect to the ten unknown components of the metric $(g_{\alpha\beta})_{\alpha,\beta=0,1,2,3}$. The differential operator,

$$\Box_g = \sum_{\mu,\nu} g^{\mu\nu} \partial_\mu \partial_\nu$$

appearing on the left hand side is very similar to the wave operator $\Box = m^{\mu\nu}\partial_\mu\partial_\nu = -\partial_0^2 + \Delta$ which we have encountered before in (8). Indeed, in a neighborhood of a point $p \in M$ we can pick our wave coordinates x^α in such a way that $g^{\mu\nu}(p) = m^{\mu\nu}$. Thus, locally, $\Box_g$ looks like $\Box = \Box_m$ and we can thus interpret (25) as a nonlinear system of *wave equations*.

The two last examples illustrate the importance of choosing *good coordinates* for equations which are defined in terms of geometric quantities, such as the Ricci curvature. To solve such equations and find interesting properties of the solutions, it is often very important to pick up a well adapted system of coordinates. In the case of gauge field theories, such as Yang-Mills equations, the role of coordinates is replaced by *gauge transformations*.

Finally we need to note that PDE arise not only in Physics and Geometry but also in many fields of applied science. In engineering, for example, one often wants to

[4]they are the exact analogue of the harmonic coordinates discussed above.

impose auxiliary conditions on solutions of a PDE, corresponding to a part of a physical system which we can directly influence, such as the portion of the string of a violin in direct contact with the bow, in order to control their behavior, i.e. obtain a beautiful sound. The mathematical theory dealing with this issue is called *Control Theory.*

Often, when dealing with complex physical systems, when we cannot possible have complete information about the state of the system at any given time, one makes various randomness assumptions about various factors which influence it. This leads to a very important class of equations called stochastic differential equations. To give a simple example consider the $N \times N$ system of the ordinary differential equation,

$$\frac{dx}{dt} = f(x(t)) \tag{26}$$

Here f is a given function $f : \mathbb{R}^N \to \mathbb{R}^N$. A solution $x(t)$ is a vector valued function $x : [0, \infty) \to \mathcal{R}^N$. Given an initial data $x(0) = x_0$ we can precisely determine the position $x(t)$ and velocity $\frac{dx}{dt}$ of the solution at any given time t. In applied situations, because of various factors which are hard to take into account, the state of the solution may not be so neatly determined. It is thus reasonable to modify the equation to take into account random effects which influence the system. One then looks at en equation of the form,

$$\frac{dx}{dt} = f(x(t)) + B(x(t))\frac{dW}{dt}(t) \tag{27}$$

where $B(x)$ is a $N \times M$ dimensional matrix and $W(t)$ denotes the brownian motion in $\mathbb{R}^M$. Similar modifications, which take randomness into account, can be made for partial differential equations. A particularly interesting example of a PDE, which is derived from a stochastic process, related to the price of stock options in finance, is the well known Black- Scholes equation. The real price of a stock option $u(s, t)$ at time t and value s, verifies the PDE,

$$\partial_t u + rs\partial_s u + \frac{\sigma^2}{2} s^2 \partial_s^2 u - ru = 0, \qquad s > 0, \quad t \in [0, T], \tag{28}$$

subject to the terminal condition at expiration time T, $u = \max(0, (s - p))$ and boundary condition $u(0, t) = 0, t \in [0, T]$. Here p is the strike price of the option. Observe that this equation is in fact a (time-reversed) variant of the heat equation (2), thus illustrating the point made above that a single class of mathematical equations can arise in several completely different applications (in this case, thermodynamics and mathematical finance).

Part 1

Basic Tools of Analysis

CHAPTER 1

Distribution Theory

This is a very short summary of distribution theory, for more exposure to the subject I suggest F.G. Friedlander and M. Joshi's excellent book *Introduction to the Theory of Distributions*, [**?**]. Hörmander's first volume of *The Analysis of Linear Partial Differential Operators*, [**?**], in Springer can also be useful.

Notation. Throughout these notes we use the notation $A \lesssim B$ to mean $a \leq cB$ where c is a numerical constant, independent of A, B. When $\Omega \subseteq \mathbb{R}^n$ is a set, we may write $(x \in \Omega)$ to denote the indicator function of the set Ω. For instance, $(5 \leq |x| < 7)$ is a function equal to 1 for $5 \leq x < 7$ and 0 otherwise.

1. Introduction to Distribution Theory

A short description of the theory of distributions contains an unavoidable oxymoron: *It is an enabling theory which allows us to differentiate functions which are in no way differentiable and manipulate them as if there were no problems whatsoever.* Its main application is to the theory of partial differential equations.

We begin by recounting how the notion of a "fundamental solution" in partial differential equations was born through classical electromagnetism. When charge does not move, any charge distribution $\rho : \mathbb{R}^3 \to \mathbb{R}$ gives rise to an electric field which (up to a conventional sign and physical constant) is the gradient of a "potential function" $V : \mathbb{R}^3 \to \mathbb{R}$. The classical physical law relating V to the charge density ρ is Poisson's equation

$$\Delta V = \rho \tag{29}$$

where $\Delta = \sum_{i=1}^{3} \partial_i^2$ denotes the Laplacian, and on physical grounds we may require V (or at least its derivative) to vanish at infinity so that distant interactions are weak.

As with any other field theory, the physical theory cannot be valid and complete unless there exists a unique solution to the equation (29) (for reasonable data ρ) which depends continuously, in some sense, on the data. In addition to resolving these issues, we seek at least a qualitative understanding of the behavior of the solution. In the present case, thanks to a huge amount of symmetry, we will even be able to derive an explicit formula, but for the heuristic analysis involved, it will only be important that the operator $\Delta = \sum_{i=1}^{3} \partial_i^2$ is linear and commutes with translations (i.e. it is a linear differential operator with constant coefficients). In fact, the Laplacian is invariant under both translations and rotations in the sense that $\Delta(f \circ T) = (\Delta f) \circ T$ for all smooth functions whenever T is a rigid motion of the Euclidean space.

The idea is to solve (29) first in the special case where the charge density $\rho(x)$ is a unit charge completely concentrated at the point $y \in \mathbb{R}^3$ (we formally write $\rho(x) = \delta_y(x)$ where the Dirac delta function corresponds to the density function of a unit point mass at y). We will discuss the meaning of the Dirac delta function later, but for the moment let us accept the formal definition of it as an operator whose action on continuous functions f is to produce the value of f at the point y i.e. δ_y is the measure,

$$\delta_y(f) = f(y).$$

Thus, we look for a solution $V_y(x)$ to the equation $\Delta V_y(x) = \delta_y(x)$ (which is currently meaningless since δ_y is not a function). By linearity of Δ, we can then obtain the general solution as a superposition of solutions from the point contributions

$$V(x) = \int V_y(x)\rho(y)dy \tag{30}$$

Formally, we can even manage to solve the equation $\Delta V_y(x) = \delta_y(x)$ for any fixed $y \in \mathbb{R}^3$. In view of the translation invariance of Δ, we may assume that $y = 0$. Since Δ is rotationally invariant (see Exercise 1) and so is δ_0 , then any solution $V_0(x) = V_0(|x|)$ should also be rotationally invariant if solutions are to be unique. We call $V_0(x)$ a "fundamental solution" for the Laplace operator. Then, postulating the existence and spherical symmetry of $V_0(x)$, we obtain (using the divergence theorem)

$$\begin{aligned}
1 &= \int_{|x|\le R} \delta_0(x)dx \\
&= \int_{|x|\le R} \Delta V_0(x)dx \\
&= \int_{|x|=R} \frac{\mathrm{d}V}{\mathrm{d}r}(|x|)d\sigma(x) \\
&= 4\pi R^2 \frac{\mathrm{d}V}{\mathrm{d}r}(R)
\end{aligned}$$

We choose the only fundamental solution decaying at infinity, namely $V_0(x) = \frac{-1}{4\pi}\frac{1}{R}$. Therefore, translating back to δ_y, we find $V_y(x) = \frac{-1}{4\pi}\frac{1}{|x-y|}$. One can see by direct computation that $\Delta V_y(x) = 0$ away from y, and one can even prove that (30) does indeed solve (29) for (say) smooth, compactly supported densities ρ. Furthermore, by taking the gradient of (30), one obtains the experimentally refutable conclusion that the electric field decays asymptotically as

$$\frac{\text{total charge}}{(\text{distance})^2}$$

far away from the charge source.

Definition. *We call $V_y(x) = \frac{-1}{4\pi}\frac{1}{|x-y|}$ a fundamental solution solution for Δ in $\mathbb{R}^3$. More generally, we define a fundamental solution for a linear operator L in $\mathbb{R}^n$ (i.e. acting on functions in $\mathbb{R}^n$) is a " generalized" function $V_y(x)$ such that $L(V_y) = \delta_y$.*

Given a fundamental solution for L we can find solutions for the equation $Lu = f$, for any smooth, compactly supported f by setting, formally (never mind, for the moment, that the integration may make no sense),

$$u(x) = \int V_y(x) f(y) dy$$

Exercise 1. Show, informally, that if L commutes with translations in the sense that $(Lf)(\cdot + y) = L(f(\cdot + y))$ for all translations $x \mapsto x + y$ then the fundamental solution also commutes with translations, in the sense that $V_y(x) = V(x - y)$ with V verifying $L(V) = \delta_0$.

Once a fundamental solution V_y of an operator L has been found, we need to make sense of it as a generalized function as well as of the formal integration above. This is precisely what the theory of distributions accomplishes. Distribution theory allows us to make heuristic calculations rigorous and, even more importantly, enables us to deal with singular objects as if they were regular functions. There are, of course, limits to this new freedom which a good theory should spell out.

Exercise 2. It is not difficult to show that, for $\rho \in \mathcal{C}_0^\infty(\mathbb{R}^3)$, the potential $V(x) = \frac{-1}{4\pi}\int_{\mathbb{R}^3}\frac{1}{|x-y|}\rho(y)\mathrm{d}y$ behaves near infinity like $\frac{-1}{4\pi}\frac{\int_{\mathbb{R}^3}\rho(y)\mathrm{d}y}{|x|} + o(|x|^{-2})$ away from the support of ρ. One way to prove this asymptotic and understand the error is to Taylor expand $\frac{1}{|x-sy|} = \frac{1}{|x|} + \int_0^1 \frac{d}{ds}\frac{1}{|x-sy|}ds$ (the idea being that the parameter y is relatively small).

When the charge distribution is centered at the origin (that is, the vector-valued integral $\int_{\mathbb{R}^3} y\rho(y)dy = 0$), show the more precise result (with explicit remainder) that, as $|x| \to \infty$,

$$V(x) = \frac{-1}{4\pi}\frac{\int_{\mathbb{R}^3}\rho(y)\mathrm{d}y}{|x|} + O(|x|^{-3})$$

It may help keep computations simple to apply the precise, first order Taylor expansion $\phi(1) = \phi(0) + \phi'(0) + \int_0^1 (1-s)\phi''(s)\mathrm{d}s$ to the auxiliary function $\phi(s) = \frac{1}{|x-sy|}$. Also, a convenient way to differentiate the absolute value function is to observe that $|x - sy|^2 = < x - sy, x - sy >$ where $<, >$ denotes the Euclidean inner product.

Remark: If the total charge $\int \rho(y)dy$ is not 0, then one can find a "center of charge" $y_c = \int y\rho dy / \int \rho dy$ so that $\int (y - y_c)\rho(y)dy = 0$. In this situation, we could Taylor expand about $y = y_c$ to see that the associated potential behaves asymptotically as though it were centered at y_c:

$$V(x) \approx \frac{-1}{4\pi} \frac{\int_{\mathbb{R}^3} \rho(y)\mathrm{d}y}{|x - y_c|}.$$

Notice, however, that when the charge "cancels out" in the sense that $\int \rho(y)dy = 0$, the associated potential function V decays more rapidly at infinity as $\frac{C}{|x|^2}$. This phenomenon of increased decay for localized, oscillatory data is not only physically important for explaining why electric forces are weak over distances when charge cancels, but it is also important in analysis where a similar cancelation arises in many other naturally occurring situations. We will see this sort of cancelation being used in a critical way later in the notes.

Exercise 3. The reasoning in the previous section can be extended to "solve" for the potential inside of a bounded region whose boundary is grounded. That is, consider the problem $\Delta V(x) = \rho(x)$ for x in a bounded domain Ω with $V = 0$ on the boundary. In principle, how could you construct a general solution of the form $V(x) = \int K(x, y)\rho(y)dy$? Where does linearity come in?

Exercise 4. Suppose that a unit of negative charge has been distributed uniformly over the sphere of radius R_1 in $\mathbb{R}^3$, and that a unit of positive charge has been distributed uniformly on the sphere of radius R_2. Find the electrostatic potential function V associated to this charge configuration ρ.

Exercise 5. a. Use the informal argument from the introduction to find the fundamental solution $K_n(x)$ of Δ in $\mathbb{R}^n$ for every $n \geq 2$; i.e. solve $\Delta K_n(x) = \delta_0(x)$ with an explicit formula for K_n.

b. Discuss the behavior as $|x| \to \infty$ of the corresponding solution

$$V(x) = \int K_n(x - y)\rho(y)dy$$

for ρ compactly supported. Namely, as $|x| \to \infty$, what is the main term and how large is the error?

2. Test Functions. Distributions

We start with some standard notation. We denote vectors in $\mathbb{R}^n$ by $x = (x_1, \ldots, x_n)$ and set $\lambda x = (\lambda x_1, \ldots, \lambda x_n)$, $x + y = (x_1 + y_1, \ldots, x_n + y_n)$. We denote by $x \cdot y$ the standard scalar product and by $|x| = (x \cdot x)^{\frac{1}{2}}$ the Euclidean length of x. Given a function $f : \Omega \to \mathbb{C}$ we denote by $\operatorname{supp}(f)$ the closure in Ω of the set where $f(x) \neq 0$. We denote by $\mathcal{C}^k(\Omega)$ the set of complex valued functions on Ω which are k times continuously differentiable and by $\mathcal{C}_0^k(\Omega)$ the subset of those which are also compactly supported. We also denote by $\mathcal{C}^\infty(\Omega) = \cap_{k\in\mathbb{N}}\mathcal{C}^k(\Omega)$ the space of infinitely differentiable functions, and by $\mathcal{C}_0^\infty(\Omega)$ the subset of those which also have compact support. The latter plays a particularly important role in the theory of distributions; it is called the space of test functions on Ω.

Let $\Omega \subset \mathbb{R}^n$ and $f \in \mathcal{C}^\infty(\Omega)$. We denote by $\partial_i f$ the partial derivative $\frac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$. For derivatives of higher order we use the standard multi-index notation. A multi-index α is an n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of nonnegative integers with length $|\alpha| = \alpha_1 + \cdots + \alpha_n$. Set $\alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$. We denote by $\alpha!$ the product of factorials $\alpha_1! \cdots \alpha_n!$. Now set $\partial^\alpha f = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n} f$. Clearly $\partial^{\alpha+\beta} f = \partial^\alpha \partial^\beta f$. Given two smooth functions u, v we have the Leibniz formula,

$$\partial^\alpha (u \cdot v) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} \partial^\beta u \partial^\gamma v.$$

Taylor's formula, around the origin, for a smooth function $f : \mathbb{R}^n \to \mathbb{C}$ can be written as follows,

$$f(x) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} \partial^\alpha f(0) x^\alpha + O(|x|^{k+1}) \qquad \text{as} \quad x \to 0.$$

Here x^α denotes the monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$.

We start by explaining a general method (often called "mollification") which can be used to approximate rougher functions by smooth ones. Essentially, one takes the function f to be approximated, and replaces f by its average after randomly translating f according to some smooth probability measure ρ with small support. It is intuitively clear and easy to prove that the randomly perturbed function is a smooth approximation to the original (imagine a sharply formed sandpile after a small earthquake), and in order to get a better approximation one shrinks the support of ρ to 0. The technical implementation of this method appears as follows.

Proposition 2.1. *Let $f \in \mathcal{C}_0^k(\mathbb{R}^n)$, $0 \le k < \infty$. Let $\rho \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ be a smooth function with support supp(ρ) contained in the unit ball $B(0,1) = \{|x| \le 1\}$ and $\int \rho(x)dx = 1$. We set $\rho_\epsilon(x) = \epsilon^{-n}\rho(x/\epsilon)$ and let*

$$f_\epsilon(x) = f * \rho_\epsilon(x) = \epsilon^{-n} \int f(y) \rho(\frac{x-y}{\epsilon}) dy = \int f(x - \epsilon z) \rho(z) dz.$$

We have:

1. *The functions f_ϵ are in $\mathcal{C}_0^\infty(\mathbb{R}^n)$ and supp$(f_\epsilon) \subset$ supp$(f) + B(0,\epsilon)$.*
2. *We have $\partial^\alpha f_\epsilon \longrightarrow \partial^\alpha f$ uniformly as $\epsilon \to 0$.*

Proof : The first part of the proposition follows immediately from the definition since the statement about supports is immediate and, by integration by parts, we can transfer all derivatives of f_ϵ on the smooth part of the integrand ρ_ϵ. To prove the second statement we simply write,

$$\partial^\alpha f_\epsilon(x) - \partial^\alpha f(x) = \int \Big(\partial^\alpha f(x - \epsilon z) - \partial^\alpha f(x)\Big) \rho(z) dz.$$

Therefore, for $|\alpha| \leq k$,

$$\begin{aligned} |\partial^\alpha f_\epsilon(x) - \partial^\alpha f(x)| &\leq \int |\partial^\alpha f(x-\epsilon z) - \partial^\alpha f(x)||\rho(z)|\mathrm{d}z \\ &\leq \int |\rho(z)|\mathrm{d}z \sup_{|z|\leq 1} |\partial^\alpha f(x-\epsilon z) - \partial^\alpha f(x)| \\ &\lesssim \sup_{|z|\leq 1} |\partial^\alpha f(x-\epsilon z) - \partial^\alpha f(x)| \end{aligned}$$

The proof follows now easily in view of the uniform continuity of the functions $\partial^\alpha f$. ■

As a corollary of the Proposition, one can easily check that the space of test functions $\mathcal{C}_0^\infty(\Omega)$ is dense in the spaces $\mathcal{C}^k(\Omega)$ as well as $L^p(\Omega)$, $1 \leq p < \infty$. Of course, one must first exhibit at least one such $\rho \in C_0^\infty(\mathbb{R}^n)$ with $\int \rho dx = 1$. Some multiple of the bump function $\rho(x) = e^{\frac{-1}{1-|x|^2}} \cdot (|x| < 1)$ will do. Another way to construct an example is by starting with any $\mathcal{C}^1$ bump function and taking advantage of the smoothing effects of random translations (as in the above proposition) but keeping the support under control to obtain a smooth bump function as a limit of an iterative process.

Definition 2.2. A distribution $u \in \mathcal{D}'(\Omega)$ is a linear functional $u : \mathcal{C}_0^\infty(\Omega) \to \mathbb{C}$ verifying the following property:

For any compact set $K \subset \Omega$ there exists an integer N and a constant $C = C_{K,N}$ such that for all $\phi \in \mathcal{C}_0^\infty(\Omega)$, with supp($\phi$) $\subset K$ we have

$$| < u, \phi > | \leq C \sum_{|\alpha|\leq N} \sup |\partial^\alpha \phi|.$$

If the same integer N can be used in the above definition for every K, then the smallest such N is called the *order* of the distribution. For example, the Riesz Representation theorem (characterizing the dual of $\mathcal{C}(X)$ for compact Hausdorff spaces) guarantess that distributions of order 0 are Borel measures.

Equivalently, a distribution u is a linear functional $u : \mathcal{C}_0^\infty(\Omega) \to \mathbb{C}$ which is continuous with respect to some topology defined on $\mathcal{C}_0^\infty(\Omega)$. This topology turns out to be a rather unorthodox one (non-metrizable[1], locally convex) but never mind all this; we can go quite far without worrying in the least about the precise definition. All we need to know is that in this topology a sequence ϕ_j converges to 0 in $\mathcal{C}_0^\infty(\Omega)$ if all the supports of ϕ_j are included in a compact subset of Ω and, for each multi-index α, $\partial^\alpha \phi_j \to 0$ in the uniform norm. With this definition in mind we have the following very useful characterization of distributions:

[1]This topology can be constructed as an inductive limit topology of Fréchet spaces $\mathcal{C}_K$, where $K \subseteq \Omega$ is compact and $\mathcal{C}_K$ is the space of all smooth functions supported in K, endowed with a Fréchet space structure by the seminorms $\phi \mapsto \sup_K |\partial^\alpha \phi|$ for all multi-indices α. We do not, however, need the precise definition.

Proposition 2.3. *A linear form* $u : \mathcal{C}_0^\infty(\Omega) \longrightarrow \mathbb{C}$ *is a distribution in* $\mathcal{D}'(\Omega)$ *iff* $\lim_{j\to\infty} u(\phi_j) = 0$ *for every sequence of test functions* ϕ_j *which converges to* 0*, in* $\mathcal{C}_0^\infty(\Omega)$*, as* $j \to \infty$*.*

Proof : This proof can be found in Friedlander, section 1.3, Theorem 1.3.2. ■

Example 1: Any locally integrable function $f \in L^1_{\text{loc}}(\Omega)$ defines a distribution,

$$< f, \phi >= \int f\phi, \qquad \forall \phi \in \mathcal{C}_0^\infty(\Omega).$$

We can thus identify $L^1_{\text{loc}}(\Omega)$ as a subspace of $\mathcal{D}'(\Omega)$. This is true in particular for the space $\mathcal{C}^\infty(\Omega) \subset L^1_{\text{loc}}(\Omega)$.

One often uses the formal notation $< u, \phi >= \int_\Omega u(x)\phi(x)dx$ even when $u \in \mathcal{D}'(\Omega)$ is not a locally integrable function, and even when ϕ is not technically a test function. This notation can be conceptually simpler, but keep in mind that this is in no way a genuine Lebesgue integral. One can, however, typically interpret this formal integration as a limit of classical integrals.

Example 2: The Dirac measure with mass 1 supported at $x_0 \in \mathbb{R}^n$ is defined by

$$< \delta_{x_0}, \phi >= \phi(x_0).$$

Remark: We shall also often denote the action of a distribution u on a test function by $u(\phi)$ instead of $< u, \phi >$. Thus $\delta_{x_0}(\phi) = \phi(x_0)$.

Definition 2.4. A sequence of distributions $u_j \in \mathcal{D}'(\Omega)$ is said to converge, *weakly*, to a distribution $u \in \mathcal{D}'(\Omega)$ if, $u_j(\phi) \to u(\phi)$ for all $\phi \in \mathcal{C}_0^\infty(\Omega)$.

For example the sequence $u_m = e^{imx}$ converges weakly to 0 in $\mathcal{D}'(\mathbb{R})$ as $m \to \infty$. Also if $f \in L^1(\mathbb{R}^n)$, with $\int_{\mathbb{R}^n} f(x)dx = 1$, the family of functions $f_\lambda(x) = \lambda^n f(\lambda x)$ converges weakly to δ_0 as $\lambda \to \infty$.

We will be able to show that any distribution is the weak limit of a sequence $u_n \in \mathcal{C}_0^\infty(\Omega)$. Due to this fact, many operations defined initially for functions extend by continuity in a unique, natural way to $\mathcal{D}'(\Omega)$.

Exercise. Given a compact set $K \subseteq \mathbb{R}^n$ and a positive distance $\delta > 0$, construct a smooth function $\eta : \mathbb{R}^n \to \mathbb{R}$ whose support is contained in $\{x \mid |y - x| \leq \delta \text{ for some } y \in K\}$ within a distance δ of K, and such that $\eta(y) = 1$ for y in some neighborhood of K. Hint: start with a rough cutoff and leave some wiggle room.

2.5. Operations with distributions. The advantage of working with the space of distributions is that while this space is much larger than the space of smooth functions most important operations on test functions can be carried over to distributions.

1. Multiplication by smooth functions: Given $u \in \mathcal{D}'(\Omega)$ and $f \in \mathcal{C}^\infty(\Omega)$ we define,

$$< fu, \phi >=< u, f\phi >, \qquad \forall \phi \in \mathcal{C}_0^\infty(\Omega)$$

in order to be consistent with the identity when u is a function

$$\int (fu)\phi dx = \int u(f\phi)dx.$$

It is easily verified that multiplication by a smooth function is a continuous endomorphism of the space of distributions.

2. Convolution with a test-function: Consider, $u \in \mathcal{D}'(\mathbb{R}^n), \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. Generalizing the convolution of two functions in a natural way, we define

$$u{*}\phi(x) =< u_y, \phi(x-y) >,$$

the subscript specifying that u is understood to be acting on functions of the variable y. Observe that the definition coincides with the usual one if u is a locally integrable function, $u \in L^1_{\text{loc}}(\mathbb{R}^n)$, for which

$$u * \phi(x) = \int u(y)\phi(x-y)dy.$$

Remark: The convolution of a distribution and a test function is not merely another distribution. Rather, observe that for every distribution $u \in \mathcal{D}'(\mathbb{R}^n)$ and $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ we have that $u{*}\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ is in fact a smooth function. For example, if e_k denotes a standard unit vector, then we can differentiate in the direction e_k as follows:

$$\begin{aligned} \frac{u{*}\phi(x+he_k) - u{*}\phi(x)}{h} &= h^{-1} < u_y, \phi(x+he_k-y) - \phi(x-y) > \\ &= < u_y, \int_0^1 \partial_k\phi(x+the_k-y)dt > . \end{aligned}$$

Now, since $x \in K$ is restricted to some compact set $K \subset \mathbb{R}^n$, then for every sequence $h_i \to 0$, the associated sequence of functions $y \mapsto \int_0^1 \partial_k\phi(x+th_ie_k-y)dt$, together with all its derivatives, converge uniformly toward $\partial_k\phi(x-y)$ and its corresponding derivatives. Moreover they are all compactly supported with supports contained in some compact set K'. Therefore,

$$\lim_{h\to 0} \frac{u{*}\phi(x+he_k) - u{*}\phi(x)}{h} = u{*}\partial_k\phi(x).$$

and thus $u{*}\phi$ has continuous partial derivatives. We can continue in this manner and conclude that in fact $u{*}\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$.

3. Differentiation of distributions: For every distribution $u \in \mathcal{D}'(\Omega)$ we define

$$< \partial^\alpha u, \phi >= (-1)^{|\alpha|} < u, \partial^\alpha \phi > .$$

We make this definition to be consistent with the integration by parts formula for functions

$$\int \partial_i u(x)\phi(x)dx = \int u(x)(-\partial_i\phi(x))dx, \quad \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n),$$

which may be proven, for example, be considering difference quotients.

Again, it is easily verified that we have thus defined a continuous endomorphism of the space of distributions. Of course, the operations above are the only possible extensions of the usual operations on smooth functions. The minus sign can be viewed dually as a differentiation of the measure $\mu = \phi(x)dx$. If we temporarily let "τ_v" denote the operation "translate by v", then

$$\partial_1 \mu = \lim_{h \to 0} \frac{\tau_{he_1}\mu - \mu}{h}$$

has a density function $-\partial_1 \phi(x)$ and the limit can be taken in the topology of $\mathcal{C}_0^\infty$. In this dual sense, we have $\partial_1 u = \lim_{h\to 0} \frac{u(x+he_1)-u(x)}{h}$ in the weak topology, which often enables us to "differentiate under the integral sign" provided we interpret all integrals in the distribution-theoretic sense.

We can now define the action of a general linear partial differential operator on distributions. Indeed let,

$$P(x,\partial) = \sum_{|\alpha| \le m} a_\alpha \partial^\alpha, \qquad a_\alpha \in \mathcal{C}^\infty(\Omega),$$

be such an operator. Then,

$$< P(x,\partial)u, \phi >=< u, P(x,\partial)^\dagger \phi >,$$

where $P(x,\partial)^\dagger$ is the *formal adjoint operator*,

$$P(x,\partial)^\dagger v = \sum_{|\alpha| \le m} (-1)^{|\alpha|} \partial^\alpha (a_\alpha v).$$

Observe that if $u_j \in \mathcal{D}'(\Omega)$ converges weakly to $u \in \mathcal{D}'(\Omega)$ then $P(x,\partial)u_j$ converges weakly to $P(x,\partial)u$.

Exercise. Show that for all $u \in \mathcal{D}'(\Omega)$ there exists a sequence $u_j \in \mathcal{C}_0^\infty(\Omega)$ such that $u_j \to u$ as $j \to \infty$ in the sense of distributions(weak convergence). Thus $\mathcal{C}_0^\infty(\Omega)$ is dense in $\mathcal{D}'(\Omega)$, with respect to the weak topology of the latter.

3. Examples of distributions on the real line

1.) The simplest nontrivial distribution is the Dirac delta function $\delta_0 = \delta_0(x)$, defined by $< \delta_0(x), \phi >= \phi(0)$. We will sometimes write $\delta(x)$ without a subscript to indicate the point mass at the origin on $\mathbb{R}$.
2.) Another simple example is the Heaviside function $H(x)$ equal to 1 for $x \ge 0$ and zero for $x < 0$. Or, using the standard identification between locally integrable functions and distributions,

$$< H(x), \phi >= \int_0^\infty \phi(x)dx.$$

Observe that $H'(x) = \delta(x)$ and that $H(x) = \int_{-\infty}^x \delta(t)dt$ is the cumulative distribution function of δ_0.

3.) A more elaborate example is $\text{pv}(\frac{1}{x})$, or simply $\frac{1}{x}$, called the principal value distribution,

$$< \frac{1}{x}, \phi >= \lim_{\epsilon \to 0} \big(\int_{-\infty}^{-\epsilon} \frac{1}{x}\phi(x)dx + \int_{\epsilon}^{\infty} \frac{1}{x}\phi(x)dx \big).$$

Observe that $\log |x|$ is locally integrable and thus a distribution by the standard identification. One can show easily that $\frac{d}{dx}\log|x| = \text{pv}(\frac{1}{x})$. Note that $\text{pv}(\frac{1}{x})$ is an odd distribution (it is orthogonal to even test functions), and is of order 1 even though it is of order 0 away from the origin. In fact, decomposing $\phi = \phi_{ev} + \phi_{odd}$ into even and odd parts, we have

$$< \text{pv}(\frac{1}{x}), \phi >= \int \frac{\phi_{odd}}{x} dx = \int \int_{-1}^{1} \phi'(tx)dt \; dx$$

. We also remark that the function $\frac{1}{|x|} \cdot (x \neq 0)$, in contrast, does not admit an extension as a distribution to the whole line.

Exercise 1. Show that the distribution $t\frac{\mathrm{d}}{\mathrm{d}t}\delta(t)$ on the line is equal to $-\delta(t)$, which is a nonzero distribution. This may seem counterintuitive since either t or $\delta'(t)$ seems to vanish at every point.

Exercise 2. Let, for $z \in \mathbb{C}$ with $0 < \arg(z) < \pi$, $\log z = \log|z| + i\arg(z)$. We can regard $x \to \log z = \log(x + iy)$ as a family of distributions depending on $y \in \mathbb{R}^+$. For $x \neq 0$ we have $\lim_{y \to 0^+} \log z = \log|x| + i\pi\big(1 - H(x)\big)$. Show that as $y \to 0$ in $\mathbb{R}^+$, $\partial_x \log z$ converges weakly to a distribution $\frac{1}{x+i0}$ and,

$$\frac{1}{x + i0} = x^{-1} - i\pi\delta_0(x). \tag{31}$$

Exercise 3. If Ω is open and connected, $u \in \mathcal{D}'(\Omega)$, and all derivatives $\partial_i u = 0$ in the sense of distributions, then u is a constant.

Exercise 4. Any non-negative distribution (i.e. $< u, \phi >\geq 0$ when $\phi \geq 0$) is in fact a Borel measure. By the Riesz representation theorem for measures, it suffices to prove that for every compact set $K \subseteq \mathbb{R}^n$ there is a constant $C = C_K$ such that

$$|u(\phi)| \leq C_K \max|\phi|$$

for all ϕ with support in K.

Exercise 5. (using the preceding exercises) If $u : (-\infty, b) \to \mathbb{R}$ is a nondecreasing function which (for simplicity) vanishes at $-\infty$, there exists a unique Borel measure $\mu \geq 0$ so that $u(x) = \int_{-\infty}^{x} 1 d\mu(t)$ for Lebesgue almost every $x \in (-\infty, b]$. In terms of μ, when is u continuous? Absolutely continuous?

Remark: The classical result that monotonic functions are almost everywhere differentiable can be derived from the above exercise in combination with some basic measure theory and the Lebesgue differentiation theorem of section (2.6).

Exercise 6. Characterize convex functions $f : (a, b) \to \mathbb{R}$. Namely, show that the following are equivalent:

(1) $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$ for all $\alpha \in [0, 1]$ and $x, y \in (a, b)$
(2) f is continuous; f' (in the distribution-theoretic sense) is a non-decreasing function, and is therefore locally bounded
(3) f is continuous; f'' is a non-negative distribution, and is therefore a finite measure when restricted to any bounded set.

Hint: The class of convex functions remains invariant under the operation of random translation, therefore mollification may help.

□

We now define an important family of distributions χ_+^z, with $z \in \mathbb{C}$, by analytic continuation. We will see this family again later while studying the fundamental solution to the wave equation, and again in our study of restriction theorems for the Fourier transform.

First recall the definition of the Gamma function,

Definition 3.1. For $\mathrm{Re}(z) > 0$ we define

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \tag{32}$$

as well as the Beta function,

$$B(a, b) = \int_0^1 s^{a-1}(1-s)^{b-1} ds \tag{33}$$

Clearly $\Gamma(a+1) = a\Gamma(a)$ and $\Gamma(1) = 1$. Thus $\Gamma(n+1) = n!$. Recall that the following identity holds:

$$B(a, b) = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} \tag{34}$$

We also record for future applications,

$$\Gamma(a)\Gamma(1-a) = B(a, 1-a) = \frac{\pi}{\sin(\pi a)} \tag{35}$$

In particular $\Gamma(1/2) = \pi^{1/2}$.

Exercise. Prove formulas (34) and (35). For help see Hörmander, [**?**] section 3.4.

Definition 3.2. For $\mathrm{Re}(a) > 0$, we denote by $j_a(\lambda)$ the locally integrable function which is identically zero for $\lambda < 0$ and

$$j_a(\lambda) = \frac{1}{\Gamma(a)} \lambda^{a-1}, \qquad \lambda > 0. \tag{36}$$

The following proposition is well known,

Proposition 3.3. *For all a, b, $Re(a)$, $Re(b) > 0$,*

$$j_a * j_b = j_{a+b}$$

Proof : We have,

$$\begin{aligned} j_a * j_b(\lambda) &= \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)}\int_0^\lambda \mu^{a-1}(\lambda-\mu)^{b-1}d\mu \\ &= \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)}\lambda^{a+b-1}\int_0^1 s^{a-1}(1-s)^{b-1}ds \\ &= \frac{B(a,b)}{\Gamma(a)\cdot\Gamma(b)}\lambda^{a+b-1} = \frac{1}{\Gamma(a+b)}\lambda^{a+b-1} = j_{a+b}(\lambda) \end{aligned}$$

■

Proposition 3.4. *There exists a family of distribution j_a, defined for all $a \in \mathbb{C}$, which coincide with the functions j_a for $Re(a) > 0$, such that, $j_a * j_b = j_{a+b}$, $\frac{d}{d\lambda}j_a(\lambda) = j_{a-1}(\lambda)$ and $j_0 = \delta_0$, the Dirac delta function at the origin. Moreover for all positive integers m, $j_{-m}(x) = \partial_x^m\delta_0(x)$.*

Proof : The proof is based on the observation that $\frac{d}{d\lambda}j_a(\lambda) = j_{a-1}(\lambda)$. Thus, for a test function ϕ,

$$\int_{\mathbb{R}} j_{a-1}(\lambda)\phi(\lambda)d\lambda = -\int_{\mathbb{R}} j_a(\lambda)\phi'(\lambda)d\lambda$$

Based on this observation we define, for every $a \in \mathbb{C}$ such that $\mathrm{Re}(a) + m > 0$ as distribution

$$< j_a, \phi >= (-1)^m \int_0^\infty j_{a+m}(\lambda)\,\phi^{(m)}(\lambda)d\lambda$$

In particular,

$$< j_0, \phi >= -\int_0^\infty j_1(\lambda)\phi'(\lambda)d\lambda = -\int_0^\infty \phi'(\lambda)d\lambda = \phi(0)$$

Hence $j_0 = \delta_0$. It is also easy to see that $j_a * j_b = j_{a+b}$ for all $a, b \in \mathbb{C}$. ■

Remark: In applications one often sees the family of distributions $\chi_+^a = j_{a+1}$. Clearly $\chi_+^a * \chi_+^b = \chi_+^{a+b+1}$, $\frac{d}{d\lambda}\chi_+^b(\lambda) = \chi_+^{b-1}(\lambda)$, and $\chi_+^{-1} = \delta_0$. Observe also that $\chi_+^k(\lambda) = \frac{\lambda^k}{k!}\cdot(\lambda > 0)$ for positive integers k, and more generally χ_+^a is homogeneous of degree a, i.e. , $\chi_+^a(t\lambda) = t^a\chi_+^a(\lambda)$, for any positive constant t. This homogeneity clearly makes sense for $\mathrm{Re}(a) > -1$ when χ_+^a is a function. Can you also make sense of it for all $a \in \mathbb{C}$?

3.5. Support of a distribution. The support of a distribution can be easily derived as follows:

Definition 3.6. For $u \in \mathcal{D}'(\Omega)$, we define the complement of the support of u,

$$\Omega\backslash\mathrm{supp}(u) = \{x \in \Omega \mid \exists\, V_x \ni x \text{ open, such that} < u, \phi >= 0 \quad \forall\phi \in \mathcal{C}_0^\infty(V_x)\}.$$

Lemma 3.7. *If $u \in \mathcal{D}'(\Omega)$ and ϕ is a test function with $supp(\phi) \subset \Omega \backslash supp(u)$, then $< u, \phi >= u(\phi) = 0$.*

Proof : This follows easily by a partition of unity argument. The argument can be found in Friedlander, section 1.4. ■

The above lemma may be used to show that any distribution $u \in \mathcal{D}'(\Omega)$ of compact support extends to test functions in $\mathcal{C}^\infty(\mathbb{R}^n)$ by taking an arbitrary cutoff $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ equal to 1 on the support of u and defining $u(\phi) = u(\psi\phi)$. In fact, the following proposition shows that we may regard a compactly supported distribution as an element of the dual to $\mathcal{C}^N(\mathbb{R}^n)$ for some N.

Proposition 3.8. *A distribution $u \in \mathcal{D}'(\mathbb{R}^n)$ has compact support $K \subset \mathbb{R}^n$ iff there exists $N \in \mathbb{N}$ such that ,$\forall \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ we have*

$$|u(\phi)| \leq C_U \sup_{x \in U} \sum_{|\alpha| \leq N} |\partial^\alpha \phi(x)|,$$

where U is an arbitrary open neighborhood of K.

Proof : This is seen by using a cutoff function which is identically 1 on the support of the distribution. ■

Remark: If we endow $\mathcal{C}^\infty(\mathbb{R}^n)$ with the Frechet topology induced by the family of seminorms given by $\phi \to \sup_{K_i} |\partial^\alpha \phi|$, with $\alpha \in \mathbb{N}^n$ and K_i running over a countable collection of compact sets exhausting $\mathbb{R}^n$, then the space of compactly supported distributions can be identified with $\mathcal{C}^\infty(\mathbb{R}^n)^*$, i.e. the space dual to $\mathcal{C}^\infty(\mathbb{R}^n)$.

We have the following useful fact (essentially dual to Taylor expansion) concerning the structure of distributions supported at one point. We will find this result useful at various parts of the notes, although its application can essentially always be replaced by repeating some variant of its proof. We will not give all the details, but the main ideas are present.

Proposition 3.9. *Let $u \in \mathcal{D}'(\mathbb{R}^n)$ and assume that $supp(u) \subset \{0\}$. Then we have $u = \sum_{|\alpha| \leq N} a_\alpha \partial^\alpha(\delta_0)$, for some integer N, complex numbers a_α and δ_0 the Dirac measure in $\mathbb{R}^n$ supported at 0.*

Proof Let u be a distribution supported at the origin, $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ be a test function. It is not true in general that $< u, \phi >$ depends only on the value of ϕ at 0, but it is true that $< u, \phi >$ depends only on the restriction of ϕ to any small neighborhood of 0. So let $\eta \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ be a cutoff function which is equal to 1 on a neighborhood of the origin, and let $\eta_\delta(x) = \eta(\frac{x}{\delta})$ for $\delta > 0$ be a cutoff function with an even smaller support. Then $< u, \phi >=< u, \eta_\delta \phi >$. We wish to prove that $< u, \phi >$ depends only on the first N derivatives of ϕ, where

$$| < u, \psi > | \leq C ||\psi||_{\mathcal{C}^N} = C \sum_{|\alpha| \leq N} \max |\partial^\alpha \psi|, \quad \psi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$$

By Taylor expansion, one reduces to the case where all derivatives $\partial^a \phi = 0$ for $|\alpha| \leq N$. More precisely, we can write

$$\begin{aligned}\phi(x) &= \phi(0) + \int_0^1 \frac{d}{dt}\phi(tx)dt \\ &= \phi(0) + \sum_i \int_0^1 x^i \partial^i \phi(tx)dt \\ &= \phi(0) + \sum_i \partial^i \phi(0)x^i + \int_0^1 (1-t)\frac{d^2}{dt^2}\phi(tx)dt\end{aligned}$$

and continues integrating by parts until one has written ϕ as a multinomial with coefficients corresponding to derivatives of ϕ at 0 plus a sum of terms of the form $x^\tau \phi_\tau$, where τ is a multi-index with $|\tau| > N$ and the functions ϕ_τ are smooth (but obviously not all of compact support).

Expanding ϕ in this way, we need to show that $< u, \eta_\delta \, x^\tau \, \phi_\tau >= 0$. Here we will use the estimate $| < u, \psi > | \leq C||\psi||_{\mathcal{C}^N}$, so we will have to estimate derivatives of the type $\partial^\alpha \, (\eta_\delta x^\tau \phi_\tau)$ with order $|\alpha| \leq N$. Observe that

$$\partial^\alpha(\eta(\frac{x}{\delta})) = \delta^{-|\alpha|}(\partial^\alpha \eta)(\frac{x}{\delta})$$

This type of scaling with derivatives is consistent with dimensional analysis, if we view δ and x to have the same "units" and view each differentiation ∂^i to have the reciprocal units. It is also a computation that comes up extremely often in analysis as we shall see later on. We take advantage of this scaling by absorbing the monomial factor into the cutoff $\eta(\frac{x}{\delta})x^\tau \phi_\tau = \delta^\tau \tilde{\eta}(\frac{x}{\delta})\phi_\tau$. We then obtain

$$\begin{aligned}| < u, \eta_\delta x^\tau \phi_\tau > | &= \delta^\tau | < u, \tilde{\eta}_\delta \phi_\tau > | \\ &\leq C \sum_{|\alpha| \leq N} \delta^\tau \max |\partial^\alpha(\tilde{\eta}_\delta \phi_\tau)|\end{aligned}$$

One could expand these derivatives using the Leibniz formula

$$\partial^\alpha(\tilde{\eta}_\delta \cdot \phi_\tau) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!}\partial^\beta \tilde{\eta}_\delta \partial^\gamma \phi_\tau$$

and generate a tremendous number of terms, but to find the exact formula for these products may not be useful even though it might be worthwhile to go through the details at least once. In practice, however, one avoids details (such as the exact values of constants) which are not at the heart of the matter by understanding what kinds of terms will occur, and in particular one isolates the "worst" terms. In this case, the worst terms occur when a derivative falls upon the cutoff $\tilde{\eta}_\delta = \tilde{\eta}(\frac{x}{\delta})$, which is becoming increasingly sharp as $\delta \to 0$. For such terms, each derivative generates a factor of δ^{-1}. However, at most N derivatives can hit this cutoff, and so we have, for some number C' independent of δ (although potentially dependent on η and ϕ),

$$| < u, \eta_\delta x^\tau \phi_\tau > | \leq C' \delta^{\tau - N}$$

which tends to 0 as $\delta \to 0$ since $\tau > N$. ■

Now that we have introduced the notion of support, it is important to observe that the convolution of two distributions cannot be defined in general, but only when certain conditions on the support of the distributions are satisfied. We note in particular the fact that if $u_1, u_2 \in \mathcal{D}'(\mathbb{R}^n)$ one of which is compactly supported, then the convolution $u_1 * u_2$ can be defined. Indeed, assuming u_2 to be compactly supported, we simply define, $(***)$

$$(u_1 * u_2) * \phi = u_1 * (u_2 * \phi), \qquad \forall \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n).$$

Here, $\operatorname{supp}(u_2 * \phi) \subset \{x + y : x \in \operatorname{supp}(u_2),\, y \in \operatorname{supp}(\phi)\}$, which is a compact set. This definition extends the classical convolution for functions.

4. Pull back of distributions

Let $\Omega \subseteq \mathbb{R}^n$ and $\Omega' \subseteq \mathbb{R}^m$ be open sets, $u \in \mathcal{D}'(\Omega')$ be a distribution on Ω', and $f : \Omega \to \Omega'$. We can sometimes define the pull-back $f^* u \in \mathcal{D}'(\Omega)$. There may be obstructions; for example, if f is a constant function, then $u \circ f$ makes sense only for continuous functions u, and more generally if f maps some set with positive measure into a set of zero measure, then $u \circ f$ does not even make sense for $u \in L^p$, which are only well-defined as functions up to a set of measure zero. To ensure that none of the obstructions mentioned above occur, we assume that f is smooth and that its derivative matrix Df has full rank at every point so that, at least, any open set maps onto another open set.

To consistently define the pull-back of u by f, when possible, we use duality by regarding the pull-back of a function as the operation adjoint to the push-forward of a measure and set

$$< f^* u, \phi > = < u, f_\# \phi >$$

where $f_\# \phi$ is (the density function of) the pushforward of the finite, complex measure $\phi\, dx$ by the map f.

We shall later prove that $f_\# \phi$ is a smooth function and hence that pull-back is well-defined. It is then obvious that $f^* u$ is continuous in the distribution u with respect to weak limits. An immediate consequence of this continuity is that the chain rule for smooth functions u and f generalizes to the case where u is a distribution. For example,

$$\nabla u(f(x)) = u'(f(x)) \nabla f(x)$$

as distributions.

The above definition is certainly consistent with the formalism of measure theory. However, it is not immediately clear that the pushforward measure has a density function which is a valid test function, nor is it clear how to compute using this definition, so let us first discuss a few concrete examples explicitly.

Example 1: When $f : \Omega \to \Omega'$ is a $\mathcal{C}^\infty$ diffeomorphism with inverse g, then we apply the familiar change of variables formula for $y = f(x)$,

$$\begin{aligned}\int u(f(x))\phi(x)dx &= \int u(f(x))\phi(g(f(x)))dx \\ &= \int u(y)\phi(g(y)) \cdot |\det Dg(y)|dy\end{aligned}$$

where Dg is the derivative matrix of g. This calculation motivates the definition of pullback for such diffeomorphisms

$$< f^*u, \phi >=< u, |detDg(y)|\phi \circ g >, \quad \phi \in \mathcal{C}_0^\infty(\Omega)$$

In this case we see that the change of variables formula is equivalent to the definition of pullback.

Example 2: If $f : \Omega \to \mathbb{R}$ has a nonvanishing gradient, then we can explicitly obtain the pullback of the delta function δ_t, namely $f^*(\delta_t) = \frac{1}{|\nabla f|}d\sigma$. Here, $d\sigma$ denotes the canonical surface measure on the embedded hypersurface $f^{-1}(t) = \{f(x) = t\} \subset \mathbb{R}^n$ and ∇f denotes the gradient of f.

In other words, we can compute the value at t of the pushforward measure's density function

$$f_\# \phi(t) = \int_{f^{-1}(t)} \phi(x)\frac{d\sigma(x)}{|\nabla f|},$$

and therefore compute $< f^*u, \phi >=< u, f_\#\phi >$ not only for a δ-function, but also for arbitrary distributions $u \in \mathcal{D}'(\mathbb{R})$. In this way, the pullback formula may be written informally as a sort of decomposition

$$u(f(x)) = \int u(t)\delta(f(x) - t)dt$$

which can be formally derived from the identity $u(y) = \int u(t)\delta(y - t)dt$.

As a sample application of this formula, one can see that the derivative of the volume of the ball of radius R is the surface area of the sphere of radius R from the fact that the gradient of $\nabla|x| = \frac{x}{|x|}$ has norm 1 and from the differentiation

$$\frac{d}{dr}\int H(r - |x|)dx = \int \delta(r - |x|)dx.$$

This formula is clear geometrically: when one compares the volume of a ball of radius r to a slightly larger ball of radius $r + \epsilon$, the change in volume is essentially ϵ times the surface area.

Since the pullback of a delta-function will be very important for us, let us give a proof of this formula. Once we have proven this formula, we have built up the theory enough to carry out the details of the previous calculation in full. One would take difference quotients in the variable r of the distribution $H(r - \cdot)$, and these difference quotients are essentially supported on a thickened sphere. We then use the trivial observation that the pullback of a distribution $< f^*u, \phi >=< u, f_\#\phi >$ is continuous in u with respect to weak limits.

Now let us prove the pullback formula for a δ-function. The geometric picture in the proof is basically a generalization of the special case $f(x) = |x|$ considered above.

Proof of the Pullback Formula for $m = 1$

By taking a partition of unity to decompose ϕ if necessary, we may assume that f may be completed to a coordinate system on the support of ϕ, since this is always possible on a small neighborhood of any point in the support of ϕ by the nonvanishing of $|\nabla f|$, and since finitely many such neighborhoods suffice. We consider the measure $\mu = \phi(y)dy$ and let $\Psi(t) : \mathbb{R} \to \mathbb{C}$ be the distribution function defined by

$$\Psi(t) = f_{\#}\mu(-\infty, t] = \int \phi(y) \cdot (f(y) \leq t)dy.$$

We now wish to show that

$$\Psi'(t) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int (t < f(y) \leq t + \epsilon) \cdot \phi(y)dy$$

exists at **every** point $t \in \mathbb{R}$, from which it will follow[2] that Ψ is absolutely continuous and that $\Psi'(t) = f_{\#}\phi(t)$ given by the formula is in fact the correct density function. For simplicity of notation, let us suppose $\epsilon > 0$.

We now verify by change of coordinates that, very close to a point $y_0 \in f^{-1}(t)$ the thickened hypersurface $\{y \mid t < f(y) \leq t+\epsilon\}$ can be parameterized to have "height" $\frac{\epsilon}{|\nabla f|(y_0)|} + o(\epsilon)$ and "width" $\sim d\sigma_{f^{-1}(t)}(y_0)$, which is at least intuitively plausible from a picture of the generic situation (for example, in the case of the preceding example $f(x) = |x|$).

We may assume (without loss of generality) that $\frac{\partial f}{\partial y^1}(y_0) \neq 0$ and consider the smooth function $h(x^1, \dots, x^n)$ satisfying

$$f(h(x), x') = x^1 \tag{37}$$

for all $x = (x^1, x^2, \dots, x^n) = (x^1, x')$ in a neighborhood of $(t = f(y^0), y_0^1, \dots, y_0^n)$ containing the support of ϕ. We then essentially use f in place of y^1 as a coordinate by making the coordinate transformation $(y^1, y^2, \dots, y^n) = (h(x), x^2, \dots, x^n)$, obtaining:

[2]The Fundamental Theorem of Calculus applies when Ψ is continuous and classically differentiable at **every** point t.

$$
\begin{aligned}
&\lim_{\epsilon} \int \left[\frac{1}{\epsilon}(t < f(y) \leq t+\epsilon) \cdot \phi(y) \right] |dy^1 \wedge dy^2 \wedge \ldots \wedge dy^n| \\
&= \lim_{\epsilon} \int \left[\frac{1}{\epsilon}(t < x^1 \leq t+\epsilon) \cdot \phi(h(x), x') \right] |dh \wedge dx^2 \wedge \ldots \wedge dx^n| \\
&= \lim_{\epsilon} \int \left[\frac{1}{\epsilon}(t < x^1 \leq t+\epsilon) \cdot \phi(h(x), x') \right] |\frac{\partial h}{\partial x^1}(x)||dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n| \\
&= \lim_{\epsilon} \int \left[\frac{1}{\epsilon}(t < x^1 \leq t+\epsilon) \cdot (\phi/|\frac{\partial f}{\partial y^1}|) \circ \big(h(x^1, x'), x'\big) \right] dx^1 \ldots dx^n \\
&= \int (\phi/|\frac{\partial f}{\partial y^1}|) \circ (h(t, x'), x')\, dx'
\end{aligned}
$$

To compute the Jacobian of the transformation in the second line, we have used the shorthand of differential forms, which quickly encapsulates the fact that the volume of an n-dimensional parallelopiped remains unchanged when one vertex is translated within the span of others through the identity

$$
\begin{aligned}
dh \wedge dx^2 \wedge \ldots \wedge dx^n &= (\frac{\partial h}{\partial x^1} dx^1) \wedge dx^2 \wedge \ldots \wedge dx^n \\
&\quad + (\sum_2^n \frac{\partial h}{\partial x^i} dx^i) \wedge dx^2 \wedge \ldots \wedge dx^n \\
&= \frac{\partial h}{\partial x^1} dx^1 \wedge \ldots \wedge dx^n
\end{aligned}
$$

and $\frac{\partial h}{\partial x^1}$ is computed through implicit differentiation of equation (37) which defines h implicitly. It is now clear that $\Psi' = f_{\#}\phi$ is a smooth function of t.

The equation (37) also shows that, for t fixed, the function $\psi_t(x') = h(t, x')$ parameterizes the hypersurface $f^{-1}(t)$ as the graph of the function $x^1 = \psi_t(x')$ when x' varies. We now wish to interpret the integral over $f^{-1}(t)$ in terms of the surface measure

$$
d\sigma_{f^{-1}(t)}(x') = \left(1 + \sum_2^n (\frac{\partial \psi_t}{\partial x^j})^2 \right)^{1/2} dx'
$$

so we compute the surface density by implicitly differentiating $f(\psi_t(x'), x') = t$ to obtain

$$
\frac{\partial f}{\partial y^1} \cdot \frac{\partial \psi_t}{\partial x^j} + \frac{\partial f}{\partial y^j} = 0 \quad j = 2, \ldots, n
$$

Hence, we see that

$$
|\nabla f| = \sqrt{\sum \left(\frac{\partial f}{\partial y^i} \right)^2} = |\frac{\partial f}{\partial y^1}| \cdot \left(1 + \sum_2^n (\frac{\partial \psi_t}{\partial x^j})^2 \right)^{1/2}
$$

Substituting into $\int (\phi/|\frac{\partial f}{\partial y^1}|) \circ \psi_t(x') dx'$ gives the formula in Example 2 above. ∎

The proof above could have been simplified by employing the same change of variables $(y^1, \dots, y^n) = (h(x), x^2, \dots, x^n)$ to show directly that

$$\int u(f(y))\phi(y)dy = \int u(t)f_{\#}\phi(t)dt$$

for all $u \in \mathcal{C}^\infty(\mathbb{R})$, or equivalently by using a smooth approximate delta-function in place of the sharp approximate delta-function $\frac{1}{\epsilon}(0 < t \le \epsilon)$. We have alternatively chosen the above, lengthier proof for its intuitive, geometric appeal and also to demonstrate the use of distribution functions for computing $f_{\#}\phi$. A briefer and more general proof is given in the Appendix along with some computational tools.

Exercise 1. Let $S_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ be the dilation map $S_\lambda(x) = \lambda x$. We say that a distribution $u \in \mathcal{D}'(\mathbb{R}^n)$ is homogeneous of degree a if, $S_\lambda^* u = \lambda^a u$. Show that the definition coincides with the usual one if u is a function. Show that, in $\mathbb{R}^n$, δ_0 is homogeneous of degree $-n$.

Exercise 2. Show that any distribution in $\mathbb{R}^n$ which is both homogeneous of degree $-n$ and also supported at the origin is a constant multiple of δ_0.

The examples above are special cases of a more general formula. We can compute $f_{\#}\phi$ when $f : \Omega \to \mathbb{R}^m$ is a smooth map whose derivative is everywhere surjective by the following explicit formula:

$$(f_{\#}\phi)(y) = \int_{f^{-1}(y)} \phi(x)\frac{d\sigma(x)}{||f^*\omega||(x)} \tag{38}$$

Here $d\sigma$ denotes the induced measure on the codimension m submanifold $f^{-1}(y) \subseteq \Omega$, $f^*\omega = df^1 \wedge \dots \wedge df^m$ is the pullback of the volume form $\omega = dy^1 \wedge \dots \wedge dy^m$ on $\mathbb{R}^m$, and $|| \cdot ||$ denotes the norm induced by the pointwise inner product on m-forms. The measure $d\sigma$ can also be written $\frac{|*f^*\omega|}{||f^*\omega||}$ where $*f^*\omega$ is the Hodge dual of $f^*\omega$. These notions are all reviewed further in the Appendix (A), where the general formula (38) is proven in a different manner than the $m = 1$ case proven above. The proof consists of decomposing the volume form

$$dx^1 \wedge \dots \wedge dx^n = \frac{1}{||f^*\omega||^2}(f^*\omega \wedge *f^*\omega)$$

and then integrating first over the level sets of f.

There are more general conditions under which the operation of pullback f^*u is possible when the singularities of the distribution u are understood in a more precise manner.

Applications

Our first application of these operations will be to prove Gauss's divergence theorem, which involves expressing the integral of some derivative of a function ϕ over the interior of a set Ω in terms of a boundary integral of ϕ. One can express the integral of any derivative over Ω as a limit of integrals of difference quotients of ϕ, but dually one can take the function ϕ fixed and take adjoint difference quotients of the characteristic function of Ω; thus, the divergence theorem turns out to be equivalent to the differentiation of the characteristic function of an open set.

We will now compute the gradient of the characteristic function χ_Ω of a domain Ω with a smooth boundary, but let us first do it in words. One can picture the graph of χ_Ω as an Ω-shaped table. If we perturb χ_Ω slightly to can obtain a smooth approximation $\tilde{\chi}_\Omega$, it is clear that the gradient of $\tilde{\chi}_\Omega$ points inside of Ω and normal to the boundary $\partial\Omega$ (in the direction of maximal increase), and furthermore the gradient remains supported essentially on the boundary of Ω. It is therefore no surprise that we obtain

Proposition 4.1 (Gauss's divergence theorem)**.** *For any test function $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$,*

$$-\int_\Omega \nabla\phi(x)dx = \int \nabla\chi_\Omega(x)\phi(x)dx = \int_{\partial\Omega} \phi(x)\vec{n}d\sigma_{\partial\Omega}(x),$$

where $\vec{n}$ denotes the inward, unit normal vector.

Proof By taking a partition of unity, we may decompose $\phi = \sum_\alpha \phi_\alpha$ where each ϕ_α is a smooth function supported within a small region V_α so that $V_\alpha \cap \Omega = \{x \in V_\alpha : f_\alpha(x) > 0\}$ is an upper contour set of some defining function $f = f_\alpha$ with nonvanishing gradient on the boundary. In V_α we have the equality (as distributions), $\chi_\Omega = H \circ f_\alpha$, i.e.

$$\int \chi_\Omega(x)\phi(x)dx = \int H(f_\alpha(x))\phi(x)dx$$

for any test function ϕ supported in V_α. Therefore, by the chain rule,

$$\nabla\chi_\Omega(x) = \delta(f_\alpha(x))\nabla f_\alpha(x) = \frac{\nabla f_\alpha}{|\nabla f_a|}(x)d\sigma_{\partial\Omega}(x) = \vec{n}d\sigma_{\partial\Omega}(x)$$

which proves the proposition. ■

We could have slightly modified the proof to allow for far less stringent regularity conditions on f by first approximating the characteristic function of Ω and taking a limit (for example by letting a family of pre-Heaviside functions H_ϵ converge to the Heaviside function). For example, when the boundary can be expressed locally as a graph of a Lipschitz function, then the normal vector is well-defined almost everywhere, and we obtain in this way the same formula for a larger variety of sets such as polygons, cubes, etc. The details can be found in Hörmander's book [?].

Extending these ideas, we can outline a quick proof of the more general Stokes' theorem. The proof goes essentially as follows: for an oriented k-dimensional manifold Y (which might be embedded inside a higher dimensional manifold X), a $k-1$ form ω on X, and a test function $\phi \in \mathcal{C}_0^\infty(X)$, we have

$$0 = \int_Y d(\phi\omega) = \int_Y \phi d\omega + \int_Y d\phi \wedge \omega$$

which is clear when ϕ has support in a coordinate patch . If $M \subseteq Y$ is an open subset of Y with smooth boundary ∂M, we take ϕ to approach the characteristic function of M (so that $d\phi$ has support on ∂M), and in the limit we obtain Stokes' theorem:

$$\int_M d\omega = \int_{\partial M} \omega$$

The extension of distribution theory to the setting of manifolds is mostly straightforward and is outlined in Hörmander.

The Appendix on integration over submanifolds included at the end of the notes may help for some of the following Exercises.

Exercise 3. Show that if f, g are two smooth functions on $\mathbb{R}^n$ with non-vanishing differential everywhere, then for all $a, b \in \mathbb{R}^n \times \mathbb{R}^n$:

$$\int \delta_0(f(a) - x)\delta_0(g(b) - x)dx = \delta_0(f(a) - g(b)).$$

Hint: Both sides are to be interpreted as distributions on $\mathbb{R}^n \times \mathbb{R}^n$. One could re-write the definition of pull-back in the form $u(g(b)) = \int u(x)\delta_0(g(b) - x)dx$. Approximating with approximate δ-functions, we can extend to the case $u(x) = \delta_0(f(a) - x)$. Alternatively, use the obvious special case where $f(a) = a$ and $g(b) = b$ are both the identity map and pull back for general f and g.

Exercise 4. A point is drawn at random from the punctured square

$$S = \{(x, y) : \max |x|, |y| < 1, \}\backslash\{(0, 0)\}$$

What is the probability density of the random variable $x \cdot y$?

Exercise 5. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be the map $(\tau, \beta) = f(x, y, z) = (x^2 + y^2 + z^2, z)$, which is nonsingular away from the line $x = y = 0$. For $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^3\backslash\{x = y = 0\})$, show directly (e.g. by computing the distribution function) that the pushforward measure $f_\# \phi$ is given by

$$\left(\int_0^{2\pi} \phi(\tau, \beta, \theta)d\theta\right) \cdot (\tau > \beta^2)\frac{|d\tau \wedge d\beta|}{2}$$

in the coordinates (τ, β, θ) on $\mathbb{R}^3$ where θ is the polar angle in the x, y plane. Check that the formula (38) gives the same result.

Exercise 6. Show that, if δ_0 is the Dirac delta function on $\mathbb{R}$, then when viewed as a distribution on $(t, x) \in \mathbb{R} \times \mathbb{R}^3\backslash\{(0, 0)\}$, we have

$$< \delta_0(t^2 - |x|^2), \phi >= \int_{\mathbb{R}^3} (\phi(|x|, x) + \phi(-|x|, x)) \frac{dx}{2|x|}$$

Why does this formula make sense as a distribution on all of $\mathbb{R}^{3+1}$, even though the derivative $\mathrm{d}(t^2 - |x|^2) = 2t\mathrm{d}t - \sum_{i=1}^3 2x_i \mathrm{d}x^i$ vanishes at the origin? Why is this the only possible definition extending $\delta_0(t^2 - |x|^2)$ to all of $\mathbb{R}^{3+1}$ which remains homogeneous of degree -2?

4.2. Other topics we have not discussed.

We have not included a section on multiplying distributions because one cannot define, in general, a meaningful, associative, product of distributions which continuously extends the usual multiplication of functions. (Try to produce an example of three distributions on the real line whose product, if defined, could not be associative, or an example of two distributions whose product could not be commutative.)

There is no difficulty in multiplying together distributions whose singularities are disjoint – one simply uses a smooth partition of unity to localize in space. When the singularities occur at the same points the matter is more subtle. The multiplication is still possible when the singularities "do not collide" in a sense made precise by the notion of a wavefront set, which measures inside the cotangent bundle of $\mathbb{R}^n$ the position and direction of the singularities of a distribution – in this situation one must try to localize in both space and "frequency" to formalize the multiplication. It is possible to show, for example, that $\delta_0(x_1, x_2) = \delta(x_1) \cdot \delta(x_2) \in \mathcal{D}'(\mathbb{R}^2)$. We will allude to the concept of the wavefront set at various points in the notes, but we will remain vague about what this set actually is; for a rigorous discussion of the wavefront set and the propagation of singularities, see $(***)$. For our purposes, it will be enough to remember that singularities have both location and "directions".

To give a more explicit hint of the most general setting: given two distributions $u(x)$ and $v(x) \in \mathcal{D}'(X)$, one defines a tensor product $u(x_1)v(x_2) \in \mathcal{D}'(X \times X)$ in the obvious way, and multiplication, when possible, is the pull-back $u(x)v(x)$ of the tensor product by the diagonal embedding $X \hookrightarrow X \times X$. Note, however, that the diagonal embedding does not satisfy the conditions we assumed when we originally defined pullback of distributions in that its derivative is not surjective. It is sometimes possible to define pullback even in such circumstances, but again some conditions involving the map and the wavefront set of the distribution to be pulled back must be met, as it is clear, for example, that not all distributions can be restricted to lower dimensional subsets. We will confront this issue later on during some of the calculations involving fundamental solutions and again when we study trace theorems (which allow us to make sense of "boundary values" when dealing with certain generalized functions in PDE) and restriction theorems for the Fourier transform.

CHAPTER 2

Fundamental solutions and the basic linear PDEs

In the Introduction, we introduced the basic concept of a fundamental solution in the particular case of the Laplacian in $\mathbb{R}^3$, and used this solution to deduce some basic facts about Poisson's equation for compactly supported data. We now begin to study several of the basic linear partial differential equations from the same angle but in greater depth.

Given a linear partial differential operator with constant coefficients

$$P(\partial) = \sum_{|\alpha| \leq k} a_\alpha \partial^\alpha$$

with $a_\alpha \in \mathbb{C}$, we say that a distribution E is a fundamental solution if it verifies $P(\partial)E = \delta_0$. If this is the case, then we can always find solution of the equation $P(\partial)u = f$, where $f \in \mathcal{D}'(\mathbb{R}^n)$ is a compactly supported distribution, by setting $u = E*f$. This follows easily from the observation that $\delta_0 * u = u$ for any $u \in \mathcal{D}'(\mathbb{R}^n)$ together with the following proposition (which ultimately stems from the fact that all translations commute in $\mathbb{R}^n$).

Proposition 0.3. *Assume $u, v \in \mathcal{D}'(\mathbb{R}^n)$ one of which is compactly supported. Then,*

$$P(\partial)(u*v) = P(\partial)u*v = u*P(\partial)v.$$

One can prove the following general result.

Theorem 0.4 (Ehrenpreis, Malgrange)**.** *Any linear partial differential operator $P(\partial)$ on $\mathbb{R}^n$ with constant coefficients has a fundamental solution.*

The proof, which involves elementary Fourier and functional analysis, is actually rather peripheral to these notes (although a midterm exercise demonstrates that there could be obstructions to a similar theorem in more general settings). Ultimately a fundamental solution has to be quite explicit to be useful in deriving interesting properties of the underlying equations.

We treat instead specific examples of important, translation-invariant linear differential operators which have special, useful invariance properties. From the simplicity and special symmetries of these operators, we are able to derive explicit formulas for the fundamental solutions. This allows us to derive important qualitative properties of the corresponding equations (existence and regularity of solutions, continuous dependence on data, etc.). These qualitative properties, however, will often have nothing to do with the various symmetries, and clearly remain true if one were to perturb the operator slightly, or even substantially. Therefore keep in mind that, though explicit formulas are very useful, we will ultimately need to develop more robust techniques to understand properties of PDE's.

1. Cauchy-Riemann equations

The operator[1] $\frac{\partial}{\partial \bar{z}} = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$ is fundamental to complex analysis, which studies solutions $f : \mathbb{C} \to \mathbb{C}$ to the linear partial differential equation $\frac{\partial f}{\partial \bar{z}} = 0$. Such functions are called holomorphic and taken together form an algebra over $\mathbb{C}$, in addition to having many other fascinating properties. The pair of partial differential equations relating the real and imaginary parts of f are known as the Cauchy-Riemann equations. We will assume the reader is already familiar with the subject of complex analysis, and proceed to develop basic facts in the subject through the application of distribution theory. For this section, we will use the formalism of differential forms and in particular denote by $\frac{\mathrm{d}\bar{z}\wedge\mathrm{d}z}{2i} = \mathrm{d}x \wedge \mathrm{d}y$ the volume form on $\mathbb{C}$. Given a function $f(z, \bar{z})$ we write, in complex notation, $df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar{z}}d\bar{z}$.

According to our general definition, a fundamental solution for $\frac{\partial}{\partial \bar{z}}$ is a distribution K in $\mathbb{R}^2$ such that $\frac{\partial}{\partial \bar{z}}K = \delta_0$. Unlike the Laplace operator, the $\frac{\partial}{\partial \bar{z}}$ operator does not commute with rotations $u(z) \mapsto \tilde{u} = u(e^{it}z)$ as an operator. In fact, we have $\frac{\partial}{\partial \bar{z}}\tilde{u} = e^{-it}\frac{\partial u}{\partial \bar{z}}(e^{it}z)$ from the calculation

$$\begin{aligned} d[u(e^{it}z)] &= \frac{\partial u}{\partial z}(e^{it}z)d(e^{it}z) + \frac{\partial u}{\partial \bar{z}}(e^{it}z)d(\overline{e^{it}z}) \\ &= e^{it}\frac{\partial u}{\partial z}(e^{it}z)dz + e^{-it}\frac{\partial u}{\partial \bar{z}}(e^{it}z)d\bar{z} \end{aligned}$$

However, if one has any fundamental solution $\frac{\partial u}{\partial \bar{z}} = \delta_0$, then by this computation, one can construct another fundamental solution $e^{it}u(e^{it}z)$ since the δ-function is rotationally invariant. By averaging over the group of rotations, we can assume without loss of generality that these two fundamental solutions are the same so that $K(e^{it}z) = e^{-it}K(z)$, which motivates us to seek a fundamental solution of the form $K(re^{i\theta}) = g(r)e^{-i\theta}$.

Since δ_0 is homogeneous of degree -2 and $\frac{\partial}{\partial \bar{z}}$ lowers the degree of homogeneity by 1, we would suspect that $g(r) = \frac{c}{r}$ for some constant $c \in \mathbb{C}$, so that the fundamental solution would be homogeneous of degree -1. Thus, we are led to guess that $K(z) = \frac{c}{z}$ for some constant c. Indeed, since $\frac{1}{z}$ is locally integrable, it defines a distribution everywhere in $\mathbb{R}^2$ with $\frac{\partial}{\partial \bar{z}}\frac{1}{z}$ supported at the origin. Moreover, since $\frac{\partial}{\partial \bar{z}}\frac{1}{z}$ is homogeneous of degree -2, we deduce from the characterization of distributions supported on a point that it must be a constant multiple of δ_0; i.e. $\frac{\partial}{\partial \bar{z}}\frac{1}{z} = C\delta_0$ for some constant $C \in \mathbb{C}$ (possibly 0).

[1] An easy way to remember this definition is to write $f(x, y) = f(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i})$. Note also that $\frac{\partial f}{\partial z} = \frac{1}{2}(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y})$ and $df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar{z}}d\bar{z}$.

We may determine the constant by applying the distribution to any test function we wish, and we will choose our test function to be the characteristic function of the unit disk $D = \{|z| \leq 1\}$. Technically, doing so leaves the realm of distribution theory that we have covered, but we will have no problem justifying our computations: the point is that one factor is at least continuous wherever the other is a bit singular, which allows one to pass from smooth approximations. In the notation of real variables, it is possible to evaluate

$$\int H(1-|z|)\cdot \frac{\partial}{\partial \bar{z}}\frac{1}{z}dxdy = -\int \frac{1}{x+iy}\frac{\partial}{\partial \bar{z}}H(1-|x+iy|)dxdy$$

by integrating by parts. By applying the product rule to

$$\frac{\partial}{\partial \bar{z}}|z|^2 = \frac{\partial}{\partial \bar{z}}(z\bar{z})$$

we obtain that $\frac{\partial}{\partial \bar{z}}|z| = \frac{z}{2|z|}$, and hence

$$\begin{aligned} -\int \frac{1}{x+iy}\frac{\partial}{\partial \bar{z}}H(1-|x+iy|)dxdy &= \int \frac{1}{2|z|}\delta(1-|z|)dxdy \\ &= \pi \end{aligned}$$

Equivalently, we can apply Stokes' theorem[2] to compute that

$$\begin{aligned} \int_D \frac{\partial}{\partial \bar{z}}\left(\frac{1}{z}\right)\frac{\mathrm{d}\bar{z}\wedge \mathrm{d}z}{2i} &= \int_D \mathrm{d}\left(\frac{1}{z}\frac{dz}{2i}\right) \\ &= \int_{\partial D}\frac{1}{z}\frac{dz}{2i} \\ &= \int_0^{2\pi} e^{-i\theta}\frac{ie^{i\theta}d\theta}{2i} \\ &= \pi \end{aligned}$$

giving the proposition:

Proposition 1.1. *Let $K(z) = \frac{1}{\pi}\frac{1}{z}$, then $\frac{\partial K}{\partial \bar{z}} = \delta_0$*

Having found a fundamental solution, we will immediately obtain a representation formula for holomorphic functions. We first note that a variation of the above calculation allows us to compute $\frac{\partial \chi_\Omega}{\partial \bar{z}}$ when Ω is an open set with Lipschitz boundary [3] and with $\frac{1}{z}$ replaced by an arbitrary test function shows that

$$\int \frac{\partial \chi_\Omega}{\partial \bar{z}}\cdot \phi(z)\frac{\mathrm{d}\bar{z}\wedge \mathrm{d}z}{2i} = -\int_{\partial\Omega}\phi(z)\frac{\mathrm{d}z}{2i} \quad \phi \in \mathcal{C}_0^\infty(\mathbb{C})$$

[2] One must take care that the parameterization $\theta \mapsto e^{i\theta}$ gives the correct orientation of the circle. A naive replication of the following calculation with the clockwise parameterization $\theta \mapsto e^{-i\theta}$ would have led to a sign error.

[3] We proved the divergence theorem assuming the boundary was smooth, however a very slight variation of the proof works for Lipschitz boundary. See Hörmander, for instance.

For example, this follows from integrating the identity

$$d(\chi_\Omega \phi dz) = (\frac{\partial}{\partial \bar{z}}\chi_\Omega \phi + \chi_\Omega \frac{\partial}{\partial \bar{z}}\phi) d\bar{z} \wedge dz$$

One of the easiest ways to check that such integral identities involving distributions are valid is by allowing the singular distribution χ_Ω to be approximated by smooth functions χ_Ω^ϵ. In this case, $\frac{\partial}{\partial \bar{z}}\chi_\Omega$ is a measure since $\partial\Omega$ is Lipschitz and ϕ is a continuous function, so one can already see that the computation is valid.

In the special case when $\frac{\partial f}{\partial \bar{z}} = 0$ in Ω and f is (say) $\mathcal{C}^1$ in a neighborhood of the closure of Ω, we can pass from smooth approximations $\phi_n \to f$ to conclude that $\int_{\partial\Omega} f(z)dz = 0$[4].

Now, applying $\frac{\partial}{\partial \bar{z}}$ to the product

$$\frac{\partial}{\partial \bar{z}}\left[\chi_\Omega \cdot \frac{1}{\pi(z-z_0)}\right] = \delta_{z_0} + \frac{1}{\pi(z-z_0)}\frac{\partial \chi_\Omega}{\partial \bar{z}}$$

gives a compactly supported distribution supported only at z_0 and the boundary of Ω.

We understand this equality by integrating against $\phi \frac{d\bar{z}\wedge dz}{2i}$, giving the identity:

Theorem 1.2. *Let $\Omega \subseteq \mathbb{C}$ be a bounded domain with Lipschitz boundary $\partial\Omega$, $z_0 \in \Omega$ and $\phi \in \mathcal{C}^\infty(\mathbb{C})$, then*

$$\frac{1}{2\pi i}\int_{\partial\Omega} \frac{\phi(z)}{z-z_0} dz = \phi(z_0) + \frac{1}{2\pi i}\int_\Omega \frac{\partial \phi}{\partial \bar{z}} \frac{1}{z-z_0} d\bar{z} \wedge dz$$

We have only stated the theorem for smooth functions, but the theorem holds much more generally by approximation. In particular, we can pass from smooth ϕ to f when f is holomorphic in Ω and $\mathcal{C}^1$ in a neighborhood of the closure of Ω, and by doing so we obtain as a corollary

Corollary 1.3 (Cauchy Integral Formula)**.** *Let f, Ω and $z_0 \in \Omega$ as above, then*

$$\frac{1}{2\pi i}\int_{\partial\Omega} \frac{f(z)}{z-z_0} dz = f(z_0)$$

Remark: Some care must be taken when applying the Cauchy Integral formula and calculating the integral over the boundary. For one thing, the assumption that f remains well-behaved at the boundary is essential for the passing from smooth approximations as the example of $\frac{1}{z}$ on the unit disk with the origin removed illustrates. In thise case, the Cauchy Integral Formula cannot apply to (say) points z_0 very close to 0 – the boundary integral $\int_{\partial D} \frac{1}{z} \cdot \frac{1}{z-z_0} dz$ clearly has size not much larger than the arclength $\int_{|z|=1} 1|dz| = 2\pi$. The other important issue which our

[4]More directly, one can calculate that $d(fdz) = df \wedge dz = \frac{\partial f}{\partial z} dz \wedge dz + \frac{\partial f}{\partial \bar{z}} d\bar{z} \wedge dz = 0$ and integrate over Ω.

use of Stokes' theorem subsumes is that the orientation of the boundary must be taken into account if.

By analyzing the Cauchy Integral Formula one can show that holomorphic functions (under the above conditions) possess a convergent power series expansion about any interior point of Ω, and in particular are smooth. Even more usefully, one can make this smoothness quantitative by deducing estimates of the form $||\partial^\alpha f||_{L^\infty(K)} \lesssim ||f||_{L^\infty(\partial\Omega)}$ for compact sets K contained in Ω. The same estimates also indicate how the solution to $\frac{\partial f}{\partial \bar{z}} = 0$ varies continuously upon its boundary values (when the solution exists). This analyticity is one example of a more general phenomenon: the regularity of a fundamental solution away from the origin corresponds to regularity of solutions to the PDE. It would be nice in general, however, to achieve a regularity result such as this one (perhaps not as strong) without relying upon the explicit formulas. We will revisit holomorphic functions shortly.

Exercise 1. We say that $u \in \mathcal{D}'(\Omega)$ is a weak solution to the Cauchy-Riemann equations if $\frac{\partial u}{\partial \bar{z}} = 0$ in the distribution theoretic sense. Prove that a continuous function which is a weak solution is in fact a classical holomorphic function (and hence analytic). (Hint: the class of holomorphic functions is closed under translation and linear combination, so it may be useful to consider a mollification of u. Then use the a-priori estimates.)

Exercise 2. From Exercise 1, deduce:

Theorem 1.4 (Schwartz Reflection Principle)**.** *: Let Ω be an open subregion of $\mathbb{C}$ intersecting the real line in an interval I. If f is continuous on Ω and holomorphic on $\Omega \backslash I$, then f is holomorphic on Ω.*

Hint: $f(x+iy) = lim_{\epsilon \to 0}(|y| > \epsilon) \cdot f(x+iy)$ in the weak sense.

Exercise 3. Also prove:

Theorem 1.5 (Morera's theorem)**.** *: Let f be a continuous function in the open disk D. Then f is a holomorphic function in D if and only if for any right triangle interior T with boundary ∂T contained in D one has*

$$\text{“}\int \chi_T \frac{\partial f}{\partial \bar{z}} d\bar{z} \wedge dz\text{”} \equiv \int_{\partial T} f(z) dz = 0$$

Note, the leftmost integral makes no sense classically, as no regularity assumptions about $\frac{\partial f}{\partial \bar{z}}$ have been made. (Hint: this can be proven either by using the linearity in χ_T to pass to general test functions, or by taking advantage of translation invariance in the assumptions)

Exercise 4. The Cauchy Integral Formula immediately implies an estimate of the form $|f(z_0)| \leq C \max_{z \in \partial\Omega} |f(z)|$ for some positive constant $C = C(z_0)$ independent of f. However, essentially because $\delta_{z_0} : f \to f(z_0)$ is also a ring homomorphism, we are able to choose $C = 1$ independent of z_0 and thereby prove

Theorem 1.6 (Maximum Modulus Principle)**.** *If Ω is open with compact closure $\overline{\Omega}$, and f is a holomorhic function in a neighborhood of $\overline{\Omega}$, then*

$$\max_{z\in\Omega}|f(z)| = \max_{z\in\partial\Omega}|f(z)|$$

One argument in this spirit is due to Landau.

On the other hand, the estimate $|f(z_0)| \leq C\max_{z\in\partial\Omega}|f(z)|$ alone implies (by the Hahn Banach theorem) that the linear functional $\delta_{z_0} : f \to f(z_0)$ defined initially for continuous boundary values of holomorphic functions extends to a continuous linear functional on $\mathcal{C}(\partial\Omega)$ and is therefore represented by a measure (i.e. for f holomorphic in Ω and continuous up to the boundary, we have $f(z_0) = \int_{\partial\Omega} f(z)\mathrm{d}\mu_{z_0}(z)$ for some finite measure μ_{z_0} on the boundary which is not necessarily unique a priori). In fact, we have already calculated this measure by proving the Cauchy Integral Formula. On the other hand, the Maximum Modulus Principle can be proven without making use of the Cauchy Integral Formula (the reader is asked to provide such a proof in Exercise 8 of the following section on the Laplace operator). One sees, therefore, that a-priori estimates and the existence of integral representation formulas come hand in hand in expressing the uniqueness of solutions and continuous dependence on data. However, notice:

Exercise 5. If arbitrary boundary data $f \in \mathcal{C}(\partial\Omega)$ could be realized by a holomorphic function u, $\frac{\partial u}{\partial \bar{z}} = 0$ in Ω then the continuous functional $\delta_{z_0} : f \mapsto u(z_0)$ for $z_0 \in \Omega$ would be a continuous ring homomorphism defined on the algebra $\mathcal{C}(\partial\Omega)$ according to the Maximum Modulus Principle above. However, any continuous, linear functional on $\mathcal{C}(\partial\Omega)$ extending δ_{z_0} as above cannot be a ring homomorphism [5]. Hence, there exist continuous functions which cannot be realized as boundary values of holomorphic functions.

Therefore, we cannot conclude from a priori estimates (like the Maximum Modulus Principle) or representation formulas alone that solutions exist. Understanding the obstructions to the existence of solutions is an interesting problem in PDE, and usually involves understanding the underlying geometry or topology of the equation of interest. For example, in the present case, a holomorphic function cannot map a closed curve to another curve with a reversed orientation.

2. Laplace Operator Δ

. As we have seen in the introduction, the Laplace operator (or Laplacian) $\Delta = \sum_{i=1}^{n}\partial_i^2$ on $\mathbb{R}^n$ is one of the simplest and most important linear differential operators. Solutions to $\Delta u = 0$ are called "harmonic functions". In two dimensions, Δ is related to the study of holomorphic functions (for example, from the identity $\Delta = 4\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z}$ and our preceding regularity results, we see that real and imaginary parts of holomorphic functions are harmonic). The operator is also often denoted

[5] In fact, all such continuous ring homomorphisms on the algebra $\mathcal{C}(\partial\Omega)$ are point masses on the boundary itself.

$\nabla \cdot \nabla =$ "$(\partial_1, \partial_2, \ldots, \partial_n) \cdot (\partial_1, \partial_2, \ldots, \partial_n)$"; this notation makes clear the rotational symmetry of the operator, and also the integration by parts identity

$$-\int \Delta u \; v dx = \int \nabla u \cdot \nabla v dx \quad u, v \in \mathcal{C}_0^\infty(\mathbb{R}^n)$$

which can be taken as an alternative definition of the operator[6].

During the exercises in the introduction, we used spherical symmetry to find the fundamental solution to the Laplace equation. We can now verify rigorously that the fundamental solution we derived formally is a true fundamental solution.

Proposition 2.1. *Define, for all $n \geq 3$, $K_n(x) = \big((2-n)\omega_n\big)^{-1}|x|^{2-n}$ while, for $n = 2$, $K_2(x) = (2\pi)^{-1}\log|x|$. Here ω_n denotes the area of the unit sphere $\mathbb{S}^{n-1}$. Then, for all $n \geq 2$,*

$$\Delta K_n = \delta_0.$$

Proof : By a direct calculation, $\Delta K_n = (\partial_r^2 + \frac{(n-1)}{r}\partial_r)K_n$ vanishes away from the origin and therefore can be expressed as a sum of derivatives of δ_0. Therefore, ΔK_n is a distribution supported at the origin in $\mathbb{R}^n$ and homogeneous of degree $-n$, implying it is a constant multiple of δ_0. To determine the constant, we may use any test function, and (with the same considerations as in the Cauchy-Riemann equations) we choose the characteristic function of the unit ball $H(1-|x|)$. By abuse of notation, let us write $K_n(x) = K_n(|x|)$.

$$\begin{aligned}
\int \Delta K_n(x) H(1-|x|) dx &\equiv -\int \nabla K_n(x) \cdot \nabla H(1-|x|) dx \\
&= \int \delta(1-|x|) \nabla K_n(|x|) \cdot \frac{x}{|x|} dx \\
&\equiv \int \delta(1-|x|) \frac{\mathrm{d}K_n}{\mathrm{d}r}(|x|) \frac{x}{|x|} \cdot \frac{x}{|x|} dx \\
&= \int_{|x|=1} \frac{\mathrm{d}K_n}{\mathrm{d}r}(1) d\sigma \\
&= 1
\end{aligned}$$

■

With the fundamental solution $K_n(x)$ in hand, we can solve the inhomogeneous equation

$$\Delta V = \rho, \quad \rho \in \mathcal{C}_0^\infty(\mathbb{R}^n)$$

with the formula $V(x) = K_n * \rho(x) = \int K_n(x-y)\rho(y)dy$. This solution is also often denoted by $\Delta^{-1}\rho$.

---

[6] This latter definition generalizes to the Laplace-Beltrami operator on Riemannian manifolds, where the gradient, dot product, and volume form must be taken with respect to the metric. This geometric point of view gives us another way of seeing the rotational invariance of the standard Laplacian.

With some basic knowledge of differential geometry, we can give another proof. In polar coordinates $x = r\omega, r > 0, |\omega| = 1$, Δ takes the form,

$$\Delta = \partial_r^2 + \frac{n-1}{r}\partial_r + r^{-2}\Delta_{\mathbb{S}^{n-1}},$$

where $\Delta_{\mathbb{S}^{n-1}}$ is the Laplace -Beltrami operator on the unit sphere $\mathbb{S}^{n-1}$.

Exercise 1. Show that the Laplace-Beltrami operator on a Riemannian manifold with metric g [7] is given, in local coordinates x^i, by

$$\Delta_g \phi = \frac{1}{\sqrt{|g|}}\partial_i\big(g^{ij}\sqrt{|g|}\partial_j\phi\big).$$

Here g^{ij} are the components of the inverse metric g^{-1} relative to the coordinates x^i. The volume element dS_g on M is given, in local coordinates, by $dS_g = \sqrt{|g|}dx^1dx^2\ldots dx^n$. Observe that, on compact manifold M,

$$\int_M \Delta_g u\, v dS_g = \int_M u\, \Delta_g v dS_g.$$

Exercise 2. Calculate the Laplace-Beltrami operator for the unit sphere $\mathbb{S}^{n-1}$ and check the polar decomposition formula for Δ. For the particular case $n = 3$, relative to the coordinates $x^1 = r\cos\theta^1, x^2 = r\sin\theta^1\cos\theta^2, x^3 = r\sin\theta^1\sin\theta^2$, $\theta^1 \in [0,\pi), \theta^2 \in [0, 2\pi)$ show that,

$$\Delta_{\mathbb{S}^2} = \partial^2_{\theta^1} + \mathrm{cotan}\theta^1\,\partial_{\theta_1} + \frac{1}{\sin^2\theta^1}\partial^2_{\theta^2}.$$

Moreover the area element dS_ω takes the form, $dS_\omega = r^2\sin\theta^1 d\theta^1 d\theta^2$.

Proof (geometric derivation): For a smooth function $\phi(x) = \phi(r\omega)$, in polar coordinates $r = |x|$, $\omega \in \mathbb{S}^{n-1}$ unit sphere in $\mathbb{R}^n$, we have

$$\begin{aligned}\Delta\phi &= \big(\partial_r^2 + \frac{n-1}{r}\partial_r + r^{-2}\Delta_{\mathbb{S}^{n-1}}\big)\phi \\ &= r^{-(n-1)}\partial_r\big(r^{n-1}\partial_r\phi\big) + r^{-2}\Delta_{\mathbb{S}^{n-1}}\phi\end{aligned}$$

We now pass to polar coordinates $x = r\omega$ so that the volume element may be written $dx = r^{n-1}drdS_\omega$. Integrating by parts on the Riemannian manifold $\mathbb{S}^{n-1}$, we calculate that

$$\begin{aligned}< \Delta K_n, \phi > &= < K_n, \Delta\phi > \\ &= \int_{|\omega|=1}\int_0^\infty K_n(r)\partial_r\big(r^{n-1}\partial_r\phi\big)drdS_\omega + \int_{|\omega|=1}\int_0^\infty K_n(r)\Delta_{\mathbb{S}^{n-1}}\phi drdS_\omega \\ &= \big((2-n)\omega_n\big)^{-1}\int_{|\omega|=1}\int_0^\infty r^{-n+2}\partial_r\big(r^{n-1}\partial_r\phi\big)drdS_\omega + 0 \\ &= -\int_0^\infty r^{-n+1}\big(r^{n-1}\partial_r\tilde{\phi}\big)dr = -\int_0^\infty \partial_r\tilde{\phi}(r)dr = \phi(0)\end{aligned}$$

[7] Where $\nabla_g u$ is defined implicitly by $\mathrm{d}u(X) = g(\nabla_g u, X)$, and Δ_g is defined by the identity $-\int \Delta_g u v dS_g = \int g(\nabla_g u, \nabla_g v)dS_g$ for $u, v \in \mathcal{C}_0^\infty(M)$,

where in the above calculation we define $\tilde{\phi}(r)$ to be the average of ϕ over $|x| = r$. We infer that, for $n \geq 3$, $\Delta K_n = \delta_0$ as desired. The case $n = 2$ can be treated in the same manner.

Remark : Observe that, up to a constant, the expression of $K_n(x)$ can also be easily guessed by looking for spherically symmetric solutions $K = K(|x|)$. Indeed, the equation $\Delta K = 0$ reduces to the ODE, $\quad K''(r) + \frac{n-1}{r} K'(r) = 0.$

■

Having found a fundamental solution, we can immediately deduce a representation formula as before. Let $\phi \in \mathcal{C}_0^\infty$ be a smooth test function, let Ω a bounded, open set with Lipschitz boundary, and let $x \in \Omega$.

We have

$$\begin{aligned} \phi(x) &= \int_\Omega \phi(y)\delta(x-y)dy \\ &= \int \chi_\Omega \phi \Delta K_n(x-y)dy \end{aligned}$$

Our strategy is to integrate by parts, allowing at most one derivative to hit the characteristic function. We recall from our discussion of pullbacks of distributions that $\nabla \chi_\Omega = \vec{n} d\sigma_{\partial\Omega}$ where $\vec{n}$ is the interior unit normal and $d\sigma$ is the surface measure on the boundary. In contrast to a classical integration by parts, one proceeds as though there are no boundary terms since the product $\chi_\Omega \phi \Delta K_n(x-y)$ has compact support. For a function f with a continuous first derivative at the boundary $\partial\Omega$, we let $\frac{\partial f}{\partial \nu}$ denote the outward unit normal.

$$\begin{aligned} \phi(x) &= -\int \nabla(\chi_\Omega \phi) \cdot \nabla K_n(x-y)dy \\ &= -\int \phi \nabla \chi_\Omega \cdot \nabla K_n(x-y)dy - \int \chi_\Omega \nabla \phi \cdot \nabla K_n(x-y)dy \\ &= \int_{\partial\Omega} \phi \frac{\partial}{\partial \nu} K_n(x-y)d\sigma(y) - \int \nabla \phi \cdot ((\nabla(\chi_\Omega K_n) - K_n \nabla \chi_\Omega) \\ &= \int_{\partial\Omega} \phi \frac{\partial}{\partial \nu} K_n(x-y)d\sigma(y) - \int_{\partial\Omega} K_n(x-y)\frac{\partial \phi}{\partial \nu} d\sigma(y) + \int_\Omega K_n(x-y)\Delta\phi dy \end{aligned}$$

We thus derive the representation formula,

$$\begin{aligned} \phi(x) \quad &= \quad \int_\Omega K_n(x-y)\Delta \phi dy \\ &+ \quad \int_{\partial\Omega} \phi \frac{\partial}{\partial \nu} K_n(x-y)d\sigma(y) - \int_{\partial\Omega} K_n(x-y)\frac{\partial \phi}{\partial \nu} d\sigma(y) \end{aligned} \tag{39}$$

In particular (by approximation), if u is harmonic within Ω (and, say, C^2 in a neighborhood of $\overline{\Omega}$)

Proposition 2.2.

$$u(y) = \int_{\partial\Omega} \left(u(x) \frac{\partial K_n}{\partial \nu}(x-y) - \frac{\partial u}{\partial \nu}(x) K_n(x-y) \right) d\sigma(x) \tag{40}$$

With our representation formula in hand, we can repeat much of the same analysis that had been remarked for the Cauchy-Riemann equations. We find that, thanks to the real analyticity of the fundamental solution, harmonic functions as above are in fact real analytic, with quantitative a priori estimates on derivatives in terms of the boundary values of u and $\frac{\partial u}{\partial \nu}$. We can also use these estimates to show that continuous functions satisfying $\Delta u = 0$ are actually classical solutions. But to proceed with the analysis of harmonic functions from this formula may be misleading because the interior values of a harmonic function are uniquely determined by its boundary values alone, and therefore the normal derivative cannot be prescribed arbitrarily.

Indeed, the Maximum Principle for harmonic functions which we now state implies that harmonic functions in the interior of a domain are determined by their boundary values alone.

Theorem 2.3 (Maximum Principle)**.** *If $u : \Omega \to \mathbb{R}$ is $\mathcal{C}^2$ on a connected, open set Ω and $\Delta u \geq 0$ in Ω, then u cannot obtain an interior maximum unless u is a constant. In particular, when Ω is bounded,*

$$\sup_{\overline{\Omega}} u(x) = \sup_{\partial\Omega} u(x) \tag{41}$$

and as a consequence, $\sup_{\overline{\Omega}} |u(x)| = \sup_{\partial\Omega} |u(x)|$ when u is harmonic.

Proof : The first statement is called the strong maximum principle for C^2, **subharmonic functions** (functions satisfying $\Delta u \geq 0$ in the classical sense); the theorem implies that a subharmonic function in a domain Ω remains in the interior strictly below any harmonic function with everywhere greater boundary values, hence the term "subharmonic". The strong maximum principle will be an easy consequence of a theorem to be proved later (the mean value inequality for subharmonic functions), although we will leave the proof for the reader. At the moment, however, we can at least prove the "weak maximum principle" (that is, inequality (41)).

Indeed, when $\Delta u > 0$ is strictly positive, the (strong) maximum principle is obvious because the function u is at any point strictly convex in at least one direction. By an approximation (e.g. replacing u by, say, $u + \epsilon x_1^2$), we can obtain the weak maximum principle when we only assume $\Delta u \geq 0$ in a bounded domain Ω and that u extends continuously to $\partial\Omega$.

It is clear that a "strong minimum principle for superharmonic functions[8]" holds upon replacing u with $-u$, which in particular implies the last equality stated in the theorem. ∎

The concepts of superharmonic and subharmonic functions described in the proof are useful even for the analysis of harmonic functions themselves because they are much easier to construct explicitly (with exponentials, polynomials, etc.) and can be used to bound harmonic functions according to the maximum principle above.

[8] Superharmonic means $\Delta u \leq 0$.

By the same reasoning that followed the discussion of the maximum modulus principle for holomorphic functions, there must be a representation formula for harmonic functions of the form $u(y) = \int_{\partial\Omega} u(x)\mathrm{d}\mu_y(x)$ for some finite measure μ_y depending on $y \in \Omega$. We can obtain such a representation formula as follows: if a harmonic function $\psi_y(x)$ can be found which coincides with the fundamental solution $K_n(x-y)$ on the boundary of Ω, then the function $G(x,y) = K_n(x-y) - \phi_y(x)$ satisfies

$$\Delta_x G(x,y) = \delta_y(x) \text{ in } \Omega \quad G(x,y) = 0 \text{ on } \partial\Omega$$

There can be only one such function by the maximum principle. This function $G(x,y)$ above is called the Green's function for Ω, and was introduced formally in the exercises in the Introduction. By computing $u(y) = \int \chi_\Omega(x)u(x)\Delta G(x,y)dx$ as in our previous representation formula (this time the boundary condition for G cancels a boundary term: $\chi_\Omega \nabla G(x,y) = \nabla(\chi_\Omega G(x,y)))$ we obtain our desired representation formula:

Proposition 2.4. *If u is harmonic in Ω and C^2 in a neighborhood of $\overline{\Omega}$ and $G(x,y)$ is as above, then*

$$u(y) = \int_{\partial\Omega} \left[u(x) \frac{\partial G}{\partial \nu}(x,y) \right] d\sigma(x) \tag{42}$$

Note: we have not proven that the function defined by the right hand side of the formula is defined for arbitrary domains, nor that it defines a harmonic function, nor even that it realizes the boundary values in the integrand as y tends towards the boundary. When the boundary is sufficiently nice (say, Lipschitz), all of these things can be proven and arbitrary continuous boundary values can be achieved by harmonic functions (in contrast to the Cauchy-Riemann equations).

The probability measure $\frac{\partial G}{\partial \nu}(x,y)\mathrm{d}\sigma(x)$ appearing in (42) describes the probability distribution of the first contact with the boundary of a random walk beginning at the point y. Thus, the value of a harmonic function at the point y may be considered the expected value which the boundary data obtains at the first contact point of a random walk beginning at y. From this interpretation some features of harmonic functions (the maximum principle and mean value property below, for example) are obvious, but we will not explore this interpretation here.

Example: For the half-space $x^n > 0$ in $\mathbb{R}^n$, one can obtain an explicit formula for the Green's function $\Delta G(x,y) = \delta_y$, $G(x^1, \dots, x^{n-1}, 0) = 0$ by placing a negative point source at the point $y^* = (y^1, \dots, y^{n-1}, -y^n)$, and defining $G(x,y) = K(x - y) - K(x - y^*)$. Then $G(x,y) = 0$ on $x^n = 0$ since such points are equidistant from both y and y^* and the fundamental solution depends only on Euclidean distance. The same method can also be used to construct a Green's function for a ball $|y| \leq 1$. In this case, one uses the conformal reflection $y \to y^* = \frac{y}{|y|^2}$, which fixes the sphere $|y| = 1$. The Green's function then takes the form $G(x,y) = K(x-y) - K(|x|(x - y^*))$. Many more examples can be obtained in two dimensions using holomorphic functions.

Among the main results of our analysis up to this point (the maximum principle, some of the various a priori estimates which can be deduced from Green's formula, and the existence of solutions on $\mathbb{R}^n$ for compactly supported data), many hold for other operators closely analogous to the Laplacian. For example, by changing variables, we see that when u is a harmonic function and $v(\Psi(x)) = u(x)$ for a diffeomorphism Ψ of $\mathbb{R}^n$, then v will satisfy an equation of the form $L[v] = \sum_{i,j=1}^n a^{ij}(y)\partial_i\partial_j v + b^i(y)\partial_i v = 0$ where the smooth functions $a^{ij}(y)$ are the coefficients of a symmetric, positive definite matrix $a^{ij}(\Psi(x)) = D\Psi(x)(D\Psi)^t(x)$ and the first order terms depend on second derivatives of Ψ.

More general operators of the form $L = \sum_{i,j=1}^n a^{ij}(x)\partial_i\partial_j + \sum_{i=1}^n b^i\partial_i$ where the matrices $(a^{ij}(x))$ symmetric and positive definite are called elliptic, and it is no surprise that they share many properties in common with the Laplacian, but they generally do not necessarily possess the same amount of symmetry [9] as the Laplace operator does, and therefore they require more robust methods to analyze successfully. However, there are also methods, for extending and transporting results and estimates for the Laplace operator to more general elliptic operators.

The following theorem embodies the rotational and translational symmetry of the Laplace operator, and in fact characterizes harmonic functions as well as the Laplace operator itself. Therefore, it can be used to prove results for the Laplace operator and harmonic functions which are beyond the reach of other methods, and therefore its applications are also limited to these purposes. The theorem shows how the Laplacian controls the change in spherical averages of varying radius.

Theorem 2.5. *[Mean Value Property] When u is harmonic in the ball of radius $R^* > R$ about x, $u(x)$ is equal to its average over the sphere of radius R centered at the point x*

$$u(x) = \frac{1}{n\omega_n}\frac{1}{R^{(n-1)}}\int_{|y-x|=R} u(y)\, d\sigma(y)$$

with the "$=$" replaced by "$\leq$" when $\Delta u \geq 0$. In fact, for all $u \in \mathcal{C}^\infty(\mathbb{R}^n)$, $0 < R_1 < R_2 < R^*$,

$$\frac{1}{R_2^{(n-1)}}\int_{|x|=R_2} u(x)\, d\sigma(x) - \frac{1}{R_1^{(n-1)}}\int_{|x|=R_1} u(x)\, d\sigma(x) = \int_{R_1}^{R_2}\left[\int_{|y|\leq\tau} \Delta u(y)\, dy\right]\tau^{-(n-1)}\, d\tau$$

Proof We prove the last formula, since the first identity of the theorem is an immediate consequence (by letting the inner radius tend to 0). In fact, the latter formula shows that spherical averages increase with the radius when $\Delta u \geq 0$.

[9] When one refers to the "symmetries" of a partial differential operator, one often has in mind a collection of vector fields which commute with the operator, or their flows which leave the operator invariant. The symmetries of Δ on $\mathbb{R}^n$ are the symmetries of the underlying Euclidean geometry: the infinitesimal translations $\partial_{x_1}, \ldots, \partial_{x_n}$ together with the infinitesimal rotations. The corresponding flows generate the group of rigid motions of Euclidean space. In this diffeomorphism-invariant sense, a coordinate change of the Laplacian as above has the same amount of symmetry. We will see later that such symmetries can be very helpful when analyzing a differential operator. In any case, it is obvious that the condition of ellipticity alone does not imply the existence of such operator-preserving flows.

Although the formula has been stated in integral form, we prove a differential version using the auxiliary function $\psi(\tau) = \frac{1}{\tau^{(n-1)}} \int_{|x|=\tau} u(x)\mathrm{d}\sigma(x) = \frac{1}{\tau^{(n-1)}} \int \delta(\tau - |x|)u(x)dx$. We denote $r = |x|$.

$$
\begin{aligned}
\frac{\mathrm{d}\psi}{\mathrm{d}\tau} &= \int \frac{1}{\tau^{(n-1)}} \left(\frac{\partial}{\partial \tau} - \frac{(n-1)}{\tau} \right) \cdot \delta(\tau - r)u(x)dx \\
&= \frac{1}{\tau^{(n-1)}} \int \left(-\frac{\partial}{\partial r} - \frac{(n-1)}{r} \right) \cdot \delta(\tau - r)u(x)dx \\
&= \frac{1}{\tau^{(n-1)}} \int \left(-\frac{\partial}{\partial r} - \frac{(n-1)}{r} \right) (-\frac{\partial}{\partial r})H(\tau - r)u(x)dx \\
&= \frac{1}{\tau^{(n-1)}} \int \Delta H(\tau - r)u(x)dx \\
&= \frac{1}{\tau^{(n-1)}} \int_{|x|\leq\tau} \Delta u(x)dx
\end{aligned}
$$

In the second equality, we used the fact that $\delta(\tau - r)$ is a distribution of order 0, and that $\frac{(n-1)}{r}$ and $\frac{(n-1)}{\tau}$ coincide to 0th order on its support, as well as the fact that any distribution pulled back by the map $(\tau, x) \to \tau - r$ remains fixed by any vector field in the null space of $\mathrm{d}\tau - \mathrm{d}r$ (hence, $(\partial_\tau + \partial_r)\delta(\tau - r) = 0$). In the fourth line, we recognized that the operator $\frac{\partial^2}{\partial r^2} + \frac{(n-1)}{r}\frac{\partial}{\partial r}$ coincides with the Laplacian when applied to spherically symmetric functions.

Integrating in τ from R_1 to R_2 gives the desired formula. ∎

Remark 2.6. What we have essentially computed is that

$$\int u(x)d\mu(x) = \int \Delta u \Delta^{-1} \mu dx$$

where μ is the measure in Exercise 5 of the Introduction.

A special case of the above formula has important applications to complex analysis. When f is a nonzero holomorphic function in a disk $\mathcal{D}$, we have the identity

$$\frac{1}{2\pi} \Delta \log |f(z)| = \sum_{\rho_k} \delta_{\rho_k}(z)$$

where ρ_k runs over the finite collection zeros of f counted with multiplicity – the measure on the right hand side is known in algebraic geometry as the zero divisor of f. One can see this identity locally near a zero ρ_j by writing $f(z) = e^{g(z)}(z - \rho_j)^n$ for some function g holomorphic in a neighborhood of ρ_j, and by recalling that real parts of holomorphic functions are harmonic and that the fundamental solution for the Laplacian in two dimensions is given by $\frac{1}{2\pi} \log |z|$. This calculation implies in particular that $\log |f(z)|$ is a subharmonic function (a fact which can be used in combination with the maximum principle to give strong estimates on holomorphic functions). Applying the general formula in the Theorem (2.5) to $u = \frac{1}{2\pi} \log |f(z)|$, we obtain

Proposition 2.7 (Jensen's formula)**.** *Let f be a function which is holomorphic in a neighborhood of the closed ball of radius R centered at zero and whose (discrete) collection of zeros $\{\rho_j\}$ satisfy $0 < |\rho_j| \neq R$. Then we have,*

$$\frac{1}{2\pi R}\int_0^{2\pi} \log|f(Re^{i\theta})|d\theta - \log|f(0)| = \sum_{|\rho_j|<R} \log\left|\frac{R}{\rho_j}\right|$$

Proof Beginning with the general formula in the Theorem (2.5) we let $u \to \frac{1}{2\pi}\log|f(z)|$ by the standard mollifier construction, and notice that our assumptions on $f(z)$ are enough to guarantee that the integral formula in the theorem is still valid for $R_2 = R$ and $0 < R_1 < \min|\rho_j|$. Letting R_1 tend to 0 gives the left hand side of Jensen's formula.

We now calculate the right hand side of the general formula explicitly with Fubini's theorem (or integration by parts):

$$\begin{aligned}\int_0^R \left[\int_{|y|\leq\tau} \frac{1}{2\pi}\Delta\log|f(z)|dy\right]\tau^{-1}d\tau &= \int_0^R \left[\int_0^\tau \sum_{\rho_j} \delta_{|\rho_j|}(t)dt\right]\tau^{-1}d\tau \\ &= \int_0^R \sum_{\rho_j} \delta_{|\rho_j|}(t)\left[\int_t^R \tau^{-1}d\tau\right]dt \\ &= \sum_{|\rho_j|<R} \log\left|\frac{R}{\rho_j}\right|\end{aligned}$$

∎

The theorems in Exercises 1 - 5 are basic theorems in the study of the Laplace equation (along with their generalizations to other elliptic PDE), and their proofs may be found in some form in either Evans or Gilbarg and Trudinger should the reader wish to consult a reference.

Exercise 1.[Hopf Lemma] states that a harmonic function on a bounded, open set $u : \Omega \to \mathbb{R}$ must satisfy $\frac{\partial u}{\partial \nu}(x_0) > 0$ at a boundary point x_0 where the boundary is smooth and $u(x_0) > u(x)$ for $x \in \Omega\backslash\{x_0\}$. Prove this fact. One approach is to design an appropriate superharmonic perturbation of u close to x_0 and use the weak maximum principle to bound u below the superharmonic perturbation.

Exercise 2. Prove the strong maximum principle for subharmonic functions. (By now you may be able to see more than one proof)

Try to obtain this result also for C^2 "subsolutions" to an elliptic equation – in other words, supposing

$$Lu = \sum_{i,j=1}^n a^{ij}(x)\partial_i\partial_j u + \sum_{i=1}^n b^i(x)\partial_i u \geq 0$$

where the matrices $a^{ij}(x)\xi_i\xi_j \geq \lambda|\xi|^2$ are uniformly positive definite.

Exercise 3. Prove the following properties of the Green's function of an open set with nice boundary Ω: $G(x,y) = \int \delta_x(z)G(z,y)dz = \int \delta_y(z)G(z,x)dz = G(y,x)$, $G(x,y) < 0$ for $x \in \Omega\backslash\{y\}$ and $\frac{\partial G}{\partial \nu}(x_0) > 0$ at any boundary point x_0, and $\int_{\partial\Omega} \frac{\partial G}{\partial \nu}(x,y)\mathrm{d}\sigma(x) = 1$.

How would you interpret any of these facts either physically or probabilistically?

Exercise 4.[Harnack's Inequality] Prove that, If $K \subseteq \Omega$ is compact, then there is a constant C depending on K such that for all *non-negative* harmonic functions u in Ω

$$\sup_K u \leq C \inf_K u$$

Theorem 2.8 (Liouville's Theorem)**.** *A bounded harmonic function on all of* $\mathbb{R}^n$ *is a constant.*

Exercise 5. Prove Liouville's theorem.

Exercise 6. The co-area formula, written distribution theoretically (and a bit vaguely) as $h(f(x)) = \int h(t)\delta_0(f(x)-t)\mathrm{d}t$, allows us to decompose a general density $h(f(x))$ into surface measures on the level sets of f, which can then be analyzed individually. For example, when $f(x) = |x|$, the co-area formula reduces to integration in polar coordinates. This formula is equivalent to the definition of pull-back of a distribution and is proven in the Appendix; use it to prove the following identity:

Let $u \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, and $\rho(x) = \tilde{\rho}(|x|)$ be a spherically symmetric density. Then if $\rho_\epsilon \equiv \epsilon^{-n}\rho(\frac{x}{\epsilon})$ we have the formula

$$\frac{\partial}{\partial \epsilon} u * \rho_\epsilon(x) = \epsilon^{-(n-1)} \int_0^\infty \left[\int_{|y| \leq \epsilon t} \Delta u(x-y)\mathrm{d}y \right] t\tilde{\rho}(t)\mathrm{d}t$$

Exercise 7. Prove (at least the $n \geq 3$ case of) the following

Proposition 2.9. *For any* $f \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, $n \geq 3$ *the equation* $\Delta u = f$ *has a unique smooth solution which vanishes at infinity, i.e. tends to zero as* $|x| \to \infty$*. The solution is represented by* $\int_{\mathbb{R}^n} K_n(x-y)f(y)dy$*. For* $n = 2$ *the same equation has a smooth solution* $u(x)$ *with* $\lim_{|x|\to\infty} \frac{|u(x)|}{|x|} = 0$ *and* $|\partial u(x)| \to 0$ *as* $|x| \to \infty$*. The solution is represented by* $\int_{\mathbb{R}^2} \frac{1}{2\pi} \log|x-y| f(y)dy$ *, and is unique (in this class) up to an additive constant.*

Exercise 8. Prove the Maximum Modulus Principle for holomorphic functions.

We will return to the study of the Laplace equation (and some of its generalizations) in Chapter 3.

3. D'Alembertian operator

Recall that the D'alembertian $\Box = -\partial_t^2 + \Delta$ is the simplest differential operator in $\mathbb{R}^{1+n}$ invariant under translations and Lorentz transformations, i.e. the Poincaré. group. The easiest way to see is to write $\Box = m^{\alpha\beta}\partial_\alpha\partial_\beta$, with m the Minkowski metric. Since m is invariant under the Poincaré group so is $\Box$. Thus it makes sense to look for a fundamental solution of the form $\phi(t,x) = f(\rho)$ where[10] $\rho = t^2 - |x|^2 = -m_{\alpha\beta}x^\alpha x^\beta$ is invariant under Lorentz transformations. Also, because the distribution δ_0 on $\mathbb{R}^{n+1}$ is homogeneous of degree $-(n+1)$ and applying $\Box$ lowers the degree of homogeneity by 2, we conclude that f must be homogeneous of degree $-\frac{n-1}{2}$. Therefore, a good candidate for a fundamental solution must have the form $E = c_n(t^2 - |x|^2)^{-\frac{n-1}{2}}$, for some constant c_n, in the region $t > |x|$. We are therefore led to look for a distribution E_+, homogeneous of degree $-n+1$, which coincides with $c_n(t^2 - |x|^2)^{-\frac{n-1}{2}}$ in the region $t > |x|$. This may seem difficult at first, due to the high degree of the singularity of $(t^2 - |x|^2)^{-\frac{n-1}{2}}$ along $|x| = |t|$, until we realize that we can make use of the homogeneous family of distributions j_a defined by proposition **??**. We need to choose in fact $a = -\frac{n-1}{2} + 1$ and take E_+ to be proportional to$j_{-\frac{n-1}{2}+1}(t^2 - x^2)$, understood as the pull back $f^*(j_{-\frac{n-1}{2}})$, with $f = t^2 - |x|^2$. It is more convenient in the context to change notation a little bit and write,

$$\chi_+^a := j_{a+1}$$

Thus,

$$E = \chi_+^{-\frac{n-1}{2}}(t^2 - |x|^2).$$

Note that the expression $\chi_+^{-\frac{n-1}{2}}(t^2 - |x|^2)$ is not exactly rigorous, since the gradient of $t^2 - |x|^2$ vanishes at the origin, and hence $\chi_+^{-\frac{n-1}{2}}(t^2 - |x|^2)$ defines a distribution only on $\mathbb{R}^{n+1} - \{0\}$. A rigorous formulation requires a bit more care, but the particular degree of homogeneity of the distribution basically allows for a unique extension to the whole space, See Exercise 3 of section (4) for the $n = 3$ case. Now the distribution we have produced has the right properties except for the fact that it supported in the entire region $|x| \le |t|$. For deterministic physical reasons we prefer a distribution supported only in the future region $|x| \le t$. This defines our candidate for a *forward* fundamental solution $E_+^{(n+1)}(t,x) = c_n H(t)\chi_+^{-\frac{n-1}{2}}(t^2 - |x|^2)$ with $H(t)$ the Heavyside function supported on $t \ge 0$ and c_na normalizing constant to be determined in the verification. Using the chain rule it is easy to show that, $\Box E_+^{(n+1)} = m^{\alpha\beta}\partial_\alpha\partial_\beta\big(\chi_+^{-\frac{n-1}{2}}(f)\big)$ must vanish outside the origin. By the usual homogeneity considerations we deduce that $\Box E_+^{(n+1)}$ is proportional to the δ function at the origin. It thus only remains to determine the normalizing factor c_n. We have the following result,

Theorem 3.1. *The distribution $E_+^{(n+1)}$ defined by*

$$E_+^{(n+1)}(t,x) = -\frac{1}{2}\pi^{\frac{1-n}{2}}H(t)\chi_+^{-\frac{n-1}{2}}(t^2 - |x|^2) \tag{43}$$

[10] Recall that we denote $t = x^0$ and we use the summation convention w.r.t the indices $\alpha, \beta = 0, 1, \ldots, n$.

is the unique fundamental solution of the wave equations supported in the forward region $|x| \leq t$.

We shall prove this theorem later for the moment a few remarks are in order. First, observe a fundamental difference between the cases when $n > 1$ is odd and the cases when n is even. Indeed in the former case $E_+^{(n+1)}$ is supported only on the boundary of the region $|x| < t$, i.e the future light cone $|x| = t$ while in the latter case $E_+^{(n+1)}$ is supported in the entire forward region $|x| \leq t$. More precisely, using the chain rule and the fact that $\frac{\mathrm{d}}{\mathrm{d}\lambda}\chi_+^s(\lambda) = \chi_+^{s-1}(\lambda)$, we can write the fundamental solution a bit more explicitly away from the origin: in dimensions $n = 3 + 2k$, the fundamental solution looks like a derivative of a measure supported on the forward light cone

$$c_n H(t)\left(\frac{-1}{2r}\partial_r\right)^k \delta(t^2 - |x|^2) = c_n H(t)\left(\frac{1}{2t}\partial_t\right)^k \delta(t^2 - |x|^2)$$

while in $n = 2 + 2k$ dimensions, it is of the form

$$c_n H(t)\left(\frac{-1}{2r}\partial_r\right)^k \frac{1}{\sqrt{t^2 - |x|^2}} \cdot (|x| \leq t) = c_n H(t)\left(\frac{1}{2t}\partial_t\right)^k \frac{1}{\sqrt{t^2 - |x|^2}} \cdot (|x| \leq t)$$

(the above distributions being equal since $\frac{1}{t}\partial_t + \frac{1}{r}\partial_r$ is in the null space of $\mathrm{d}(t^2 - r^2)$). In the most important particular case, when $n = 3$, we have,

$$E_+^{(1+3)} = \frac{1}{2\pi}H(t)\delta(t^2 - |x|^2) = \frac{1}{4\pi}\delta(t - |x|) \tag{44}$$

Also, for $n = 2$,

$$E_+^{(1+2)} = \frac{1}{2\pi^{1/2}}H(t)(t^2 - |x|^2)^{1/2} = \tag{45}$$

It is important to observe that the knowledge of the fundamental solution in odd dimensions allows one to determine it for even dimensions. This is called the method of descent. This can be done by simply applying $E_+^{(2k+1+1)}$ to a test functions which are independent of one of the spatial variables. Never mind that this test function does not have compact support, it will work because the fundamental solution has compact support in x for any t. As an example the reader is invited to deduce (50) from (44). As another simple remark, observe that though E_+ is compactly supported in x for every fixed t. Thus E_+ can be applied to any smooth functions whose compact support in t.

Exercise 1. Deduce the fundamental solution for dimension $n = 1$. Show in fact[11] that the general solution to the Cauchy problem takes on the form $\psi_1(t + x) + \psi_2(t - x)$.

The fundamental solution allows us to solve the general Cauchy problem,

$$\Box\phi = f, \qquad \phi(0, x) = f(x), \partial_t\phi(0, x) = g(x) \tag{46}$$

[11] In $1+1$ dimensions, the wave operator factors into $\Box = (\partial_t - \partial_x) \cdot (\partial_t + \partial_x)$

To see how to do this consider a point $p = (t_0, x_0)$ with $t_0 > 0$ and observe that, for any test function ϕ we have in the upper half space, $\mathcal{D}_+ = \{(t,x)\ t \geq 0\}$,

$$\phi(t_0, x_0) = \int_{\mathcal{D}_+} \phi(t,x)\delta_p(t,x)dtdx = \int_{\mathbb{R}^{1+n}} \chi_+\phi\,(m^{\alpha\beta}\partial_\alpha\partial_\beta)E_p(t,x)dtdx$$

where χ_+ is the characteristic function of $\mathcal{D}_+$ and $E_p(t,x) = E_+(t - t_o, x - x_0)$. Therefore, integrating by parts,

$$\begin{aligned}
\phi(t_0, x_0) &= -\int_{\mathbb{R}^{1+n}} m^{\alpha\beta}\partial_\alpha\chi\,\phi\,\partial_\beta E_p - \int_{\mathbb{R}^{1+n}} m^{\alpha\beta}\,\chi\,\partial_\alpha\phi\,\partial_\beta E_p \\
&= -\int_{\mathbb{R}^{1+n}} m^{\alpha\beta}\partial_\alpha\chi\,\phi\,\partial_\beta E_p + \int_{\mathbb{R}^{1+n}} \chi\, m^{\alpha\beta}\partial_\alpha\partial_\beta\phi\,E_p \\
&+ \int_{\mathbb{R}^{1+n}} m^{\alpha\beta}\partial_b\chi\partial_\alpha\phi\,E_p
\end{aligned}$$

i.e.,

$$\begin{aligned}
\phi(t_0, x_0) &= \int_{\mathcal{D}_+} \Box\phi E_p - \int_{\mathbb{R}^{1+n}} \partial^\alpha\chi\,\phi\,\partial_\alpha E_p + \int_{\mathbb{R}^{1+n}} \partial^\alpha\chi\partial_\alpha\phi\,E_p \\
&= \int_{\mathcal{D}_+} \Box\phi + \int_{\mathbb{R}^{1+n}} \partial_t\chi\,\phi\,\partial_t E_p - \int_{\mathbb{R}^{1+n}} \partial_t\chi\partial_t\phi\,E_p \\
&= \int_{\mathcal{D}_+} \Box\phi + \int_{\mathbb{R}^{1+n}} \delta(t)\,\phi\,\partial_t E_p - \int_{\mathbb{R}^{1+n}} \delta(t)\partial_t\phi\,E_p \\
&= \int_{\mathcal{D}_+} \Box\phi E_p - \int_{\mathbb{R}^{1+n}} \delta(t)\,\phi\,\partial_{t_0} E_p - \int_{\mathbb{R}^{1+n}} \delta(t)\partial_t\phi\,E_p \\
&= \int_{\mathcal{D}_+} \Box\phi E_p - \partial_{t_0}\big(\int_{\mathbb{R}^n} f(x)E_p(0,x)dx\big) - \int_{\mathbb{R}^n} g(x)E_p(0,x)dx
\end{aligned}$$

The final formula takes the form,

$$\begin{aligned}
\phi(t_0, x_0) &= \int_0^{t_0}\int_{\mathbb{R}^n} E_+(t_0 - t, x_0 - x)\Box\phi(t,x)dtdx \\
&- \partial_{t_0}\big(\int_{\mathbb{R}^n} E_+(t_0, x - x_0)f(x)dx\big) - \int_{\mathbb{R}^n} E_+(t_0, x - x_0)g(x)dx
\end{aligned} \tag{47}$$

Leaving aside the issue of uniqueness, which we shall treat separately later on, we deduce the following.

Theorem 3.2. *[Kirchoff-Hadamard] The initial value problem $\Box\phi = F$, $\phi(0,x) = f(x)$, $\partial_t\phi(0,x) = g(x)$ has a unique solution for arbitrary smooth functions f, g, F, given by formula* (47).

Exercise 2. Compare formula (47) with (39) for the Laplacian. Explain what may go wrong if we try to prove a result for the Laplace equation similar to that of theorem 3.2 above.

Exercise 3. Show that in the particular case of dimension $1+3$ formula (47) takes the more familiar Kirchoff formula form,

$$\begin{aligned} \phi(t,x) &= \partial_t\big((4\pi t)^{-1}\int_{|x-y|=t} f(y)da(y)\big) + (4\pi t)^{-1}\int_{|x-y|=t} g(y)da(y) \\ &+ \int_0^t ds\frac{1}{t-s}\int_{|x-y|=t-s} \Box\phi(s,y)da(y) \end{aligned} \tag{48}$$

The traditional way to derive the Kirchoff formula (48) is to first prove it in the homogeneous case, i.e. $\Box\phi = 0$. In fact it suffices to prove it for the case $f = 0$ and arbitrary g using the beautiful method of spherical means, see [**J**] for a clean derivation. Once the homogeneous case is treated one can derive the general formula using the Duhamel principle. This goes as follows: Let $W(t)g$ denote the solution $\phi(t,\cdot)$ of the homogeneous problem with data $f = 0$ and arbitrary g. Think of it as an family of operators, parametrized by t, which take smooth functions in $\mathbb{R}^n$ to smooth functions in $\mathbb{R}^n$. We then have to verify that the solution of the equation $\Box\phi = F$ is given by the formula

$$\phi(t,x) = \int_0^t W(t-s)F(s,\cdot)ds \tag{49}$$

Exercise 4. Prove the claim.

Exercise 5. What happens if we replace in the formulation of the Cauchy problem the hypersurface $t = 0$ with a more general hypersurface Σ_0 given by $t = h(x)$?. Show a similar formula with that in (47) can be deduced if the hypersurface is space-like, i.e. $|\nabla h(x)| < 1$. What happens if the surface becomes time-like, i.e. $|\nabla h(x_0)| > 1$ at some point $(t_0, x_0) \in \Sigma_0$. Show that the Cauchy problem with prescribed initial values and normal derivatives on the light cone $t = |x|$ does not, in general, admit a solution in the spatial interior of the cone. What happens when you try to derive a representation formula for data on the light cone with $H(t-|x|)$ replacing $H(t)$?

Exercise 6. Suppose ρ is compactly supported and that ρ is smooth outside of a compact set K. Let u_+ be the solution $E_+ * \rho$ to $\Box u = \rho$ in $\mathbb{R}^{n+1}$. Show that u is smooth outside of the set of light cones emanating from K given by $\{k + (|x|, x) : k \in K, x \in \mathbb{R}^n\}$.

Proof of theorem 3.1 in $\mathbb{R}^{1+3}$. First remark that we can write $\Box$ in terms of spherical coordinates as follows,

$$\Box = -\partial_t^2 + \Delta = -\partial_t^2 + \partial_r^2 + \frac{2}{r}\partial_r + r^{-2}\Delta_{\mathbb{S}^{n-1}}$$

we have to check that

$$E_+^{(1+3)}(t,x) = -\frac{1}{2}\pi^{-1}H(t)\delta_0(t^2 - |x|^2) = -\frac{1}{4\pi}r^{-1}\delta(t-r)$$

with $r = |x|$. Thus, since $\Box\phi = -r^{-1}(\partial_t + \partial_r)(\partial_t - \partial_r)(r\phi) + \Delta_{\mathbb{S}^2}\phi$, we have with $\psi(t, r\omega) = (\partial_t - \partial_r)\big(r\phi(t, r\omega)\big)$,

$$\begin{aligned} < E_+, \Box\phi > &= \frac{1}{4\pi}\int_{\mathbb{S}^2}\int_0^\infty\int_0^\infty \delta(t-r)(\partial_t + \partial_r)\psi dt dr dS_\omega \\ &= \frac{1}{4\pi}\int_{\mathbb{S}^2} dS_\omega\Big(\int_0^\infty \frac{d}{dr}\psi(r,r)dr\Big) \\ &= -\psi(0,0) = \phi(0) \end{aligned}$$

Thus, $\Box E_+ = \delta_0$ as desired.

In what follows we give yet another derivation of the fundamental solution for the wave equation in the special case $\mathbb{R}^{1+3}$. This is the so called geometrics optics derivation. We look for solutions of $\Box\phi = 0$ of the form,

$$E = A\delta(u) \tag{50}$$

for given real functions A and u to be determined. Here $\delta(u)$ is simply the pull back of δ_0 by u as discussed in subsection 4, Example 2. A simple calculation leads to,

$$m^{\alpha\beta}\partial_\alpha\partial_\beta\big(A\delta(u)\big) = m^{\alpha\beta}\big(\partial_\alpha\partial_\beta A\,\delta(u) + (2\partial_\alpha A\partial_\beta u + \Box u)\,\delta'(u) + \partial_\alpha u\partial_\beta u\,\delta''(u)\big)$$

To cancel the coefficient of $\delta''(u)$ we need to chose u such that,

$$m^{\alpha\beta}\partial_\alpha u\partial_\beta u = 0. \tag{51}$$

This is the famous Eikonal equation in Minkowski space. A simple family of solutions is given by $u(t,x) = t - t_0 - |x - x_0|$ for a given point (t_0, x_0), whose level hypersurfaces are simply backward light cones with vertex at (t_0, x_0). For our purposes we choose $u = t - |x|$. Next, to cancel the coefficient of $\delta'(u)$, we need to choose A such that[12],

$$2\partial_\alpha A\partial_\beta u + \Box u = 0.$$

One can easily check that the choice $A = |x|^{-1}$ will do. Finally it only remains to calculate the term containing $\delta(u)$, i.e.,

$$(\Box A)\delta(u) = (-\Delta|x|^{-1})\delta(u) = -4\pi\delta_0(x)\delta(u) = -\delta_0(t,x)$$

where the first $\delta_0(x)$ is the delta function in $\mathbb{R}^3$ while the final δ_0 is the desired delta function in $\mathbb{R}^{1+3}$. Hence $E_+^{1+3} = -\frac{1}{4\pi}\frac{1}{|x|}\delta(t - |x|)$ as desired.

Exercise 7 Justify that last step involving products of distributions.

Uniqueness of the fundamental solution E_+. It suffices to prove uniqueness of solutions to the general Cauchy problem in theorem (3.2).

Exercise 8. Verify the above statement.

We start with the simple calculation involving the energy momentum tensor. To calculate efficiently it helps to remember that we are using the summation convection with respect to the space-time indices $\alpha, \beta = 0, 1, \dots, n$. We will also be using

[12]It turns out that the equation below can be interpreted as a transport equation along the generators of the backward null cone $t - |x| = 0$.

the standard geometric convention of raising, or lowering, the indices relative to the metric. Thus, if U^α is vector (so called contravariant) we define the covariant vector $U_\beta = m_{\beta\alpha} U^\alpha$. Similarly, if V_α is a covariant vector, we define $V^\beta = m^{\beta\alpha} V_\alpha$.

Proposition 3.3. *Let*

$$Q_{\alpha\beta}[\phi] = \partial_\alpha\phi\partial_\beta\phi - \frac{1}{2} m_{\alpha\beta}\big(m^{\gamma\delta}\phi_\gamma\phi_\delta\big) = \partial_\alpha\phi\partial_\beta\phi - \frac{1}{2} m_{\alpha\beta}\phi^\gamma\phi_\gamma \tag{52}$$

the so called energy momentum tensor of $\square$. *Then*[13]

$$\partial^\beta Q_{\alpha\beta} = \square\phi\,\partial_\alpha\phi. \tag{53}$$

In particular, if $\square\phi = 0$, $\partial^\beta Q_{0\beta} = 0$. Now consider a point $p(t_0, x_0) \in \mathbb{R}^{1+n}$ and the solution $u(t,x) = t - t_0 - |x - x_0|$ to (51) introduced above. Let also $t_1 < t_t < t_0$ and consider the distribution $H(u)H(t-t_1)H(t_2-t)$, with $H(t)$ the Heaviside function, and perform an integration by parts, as we have done many times before, to derive the identity,

$$\begin{aligned}
0 &= \int_{\mathbb{R}^{1+n}} H(u)H(t-t_1)H(t_2-t)\partial^\beta Q_{0\beta} \\
&= -\int_{\mathbb{R}^{1+n}} H'(u)H(t-t_1)H(t_2-t)\partial^\beta u Q_{0\beta} \\
&+ \int_{\mathbb{R}^{1+n}} H(u)H(t-t_1)H'(t_2-t)Q_{00} \\
&- \int_{\mathbb{R}^{1+n}} H(u)H'(t-t_1)H(t_2-t)Q_{00}
\end{aligned}$$

This identity can be rewritten in the form,

$$\int_{\mathcal{D}(t_2)} Q_{00} + \int_{\mathcal{N}(t_1,t_2)} Q_{0\beta}L^\beta = \int_{\mathcal{D}(t_1)} Q_{00} \tag{54}$$

where $\mathcal{D}(t_1), \mathcal{D}(t_2)$ are t-sections through the solid light cone $|x - x_0| \le t - t_0$, $\mathcal{N}(t_1, t_2)$ represents the portion of the light cone $|x - x_0| = t - t_0$ between the sections $t = t_1$ and $t = t_2$ and

$$L^\beta = -\partial^\beta u = m^{\beta\gamma}\partial_\gamma u.$$

It is easy to check that,

$$Q_{00} = \frac{1}{2}(|\partial_t\phi|^2 + \sum_{1=1}^n |\partial_i\phi|^2) = \frac{1}{2}(|\partial_t\phi|^2 + |\nabla\phi|^2).$$

and that we have,

Exercise 9. Show that

$$Q_{0\beta}L^\beta \ge 0.$$

We thus deduce

[13] In view of the raising and lowering of indices convention, $\partial^\beta Q_{\alpha\beta} = m^{\beta\gamma}\,\partial_\gamma Q_{\alpha\beta}$. This works well, since $\partial_\gamma m_{\alpha\beta} = 0$.

Theorem 3.4 (Energy inequality)**.** *For every solution of the wave equation $\Box\phi = 0$, in a neighborhood of the solid region bounded by the surfaces $t - t_0 - |x - x_0| = 0$, $t = t_1$ and $t = t_2$, we have,*

$$\int_{D(t_2)} \frac{1}{2}(|\partial_t\phi|^2 + |\nabla\phi|^2) \leq \int_{D(t_1)} \frac{1}{2}(|\partial_t\phi|^2 + |\nabla\phi|^2)$$

In particular any smooth solution which vanishes at $D(t_1)$ must also vanish at $D(t_2)$.

Exercise 10. Deduce from the energy inequality the finite propagation speed for the Cauchy problem, as we have deduced earlier from the explicit solution. Note that strong Huygens' principle in odd spatial dimensions $n = 3+2k$ discussed earlier is a very special phenomenon related to the precise form of the wave operator $\Box$. However the phenomenon of finite speed of propagation exhibited by the wave equation is a more robust feature shared by many related equations for which an energy inequality as above holds true.

Finally, we check below the validity of our forward fundamental solution in all dimensions.

Proof [Theorem 3.1 all n] We prove the formula (modulo the absolute value of the constant c_n.)

$$\begin{aligned}
&\int\int \Box\chi_+^{-\frac{(n-1)}{2}}(t^2 - |x|^2)\cdot H(1 - |t|)dxdt \\
&\qquad = \int\int \partial_t^2\chi_+^{-\frac{(n-1)}{2}}(t^2 - |x|^2)\cdot H(1 - |t|)dxdt \\
&\qquad = \int\int \partial_t\chi_+^{-\frac{(n-1)}{2}}(t^2 - r^2)\delta(1 - t)dxdt \\
&\qquad - \int\int \partial_t\chi_+^{-\frac{(n-1)}{2}}(t^2 - r^2)\delta(1 + t)dxdt \\
&\qquad = 2\int\int \partial_t\chi_+^{-\frac{(n-1)}{2}}(t^2 - r^2)\delta(1 - t)dxdt \\
&\qquad = 2\int\int (\frac{-t}{r})\partial_r\chi_+^{-\frac{(n-1)}{2}}(t^2 - r^2)\delta(1 - t)dxdt
\end{aligned}$$

Taking the definition of $\delta(1-t)$ the above becomes an integral over the hypersurface $t = 1$ which we put in polar coordinates (writing the volume form $r^{n-1}drd\Omega_{n-1}$ with $d\Omega_{n-1}$ the surface measure on the unit sphere in $\mathbb{R}^n$). We remark here that, as we will see later on, a distribution does not quite have to be a continuous function in order to have a meaningful restriction to a lower dimensional submanifold.

Dropping the unimportant factor of 2, we proceed

$$\begin{aligned}\int\int(\frac{-t}{r})\partial_r\chi_+^{-\frac{(n-1)}{2}}(t^2-r^2)\delta(1-t)dxdt \\ = \int\int\frac{-1}{r}\partial_r\chi_+^{-\frac{(n-1)}{2}}(1-r^2)r^{n-1}drd\Omega_{n-1} \\ = |S^{n-1}|\int_0^\infty -r^{n-2}\partial_r\chi_+^{-\frac{(n-1)}{2}}(1-r^2)dr\end{aligned}$$

Our proposition has therefore reduced to showing that the number

$$-\int_0^\infty r^{n-2}\partial_r\chi_+^{-\frac{(n-1)}{2}}(1-r^2)dr$$

is positive. Despite not having defined this number[14], we prove its positivity by induction, separating into cases based on the parity n. For $n=1+2k$, we integrate by parts and notice the boundary term vanishes to find

$$\begin{aligned}\int_0^\infty r^{n-2}\partial_r\chi_+^{-\frac{(n-1)}{2}}(1-r^2)dr &= \int_0^\infty r^{2k}(\frac{-\partial_r}{2r})^{k+1}H(1-r^2)dr \\ &\equiv I_k \\ &= 2^{-(k+1)}\int\partial_r\left(r^{2k-1}\cdot H(r)\right)(\frac{-\partial_r}{2r})^k H(1-r^2)dr \\ &= 2^{-(k+1)}(2k-1)\int_0^\infty r^{2(k-1)}(\frac{-\partial_r}{2r})^k H(1-r^2)dr \\ &= 2^{-(k+1)}(2k-1)I_{k-1}\end{aligned}$$

Where the last integral should be positive by induction on k provided we can calculate

$$\begin{aligned}I_1 &= \int_0^\infty r^2(\frac{-\partial_r}{2r})H(1-r^2)dr = \int_0^\infty \delta(1-r^2)dr \\ &= \int_0^\infty \frac{1}{1+r}\delta(1-r)dr = 1/2 > 0\end{aligned}$$

This proves the proposition in odd spatial dimensions once $c_n>0$ has been chosen appropriately. The case $n=2+2k$ is similar. ■

4. Heat Operator $\mathcal{H}$.

We consider the heat operator $\mathcal{H}=\partial_t-\Delta$ acting on functions defined on $\mathbb{R}\times\mathbb{R}^n=\mathbb{R}^{n+1}$. It makes sense to look for spherically symmetric solutions to $\mathcal{H}u=0$: that is to say, functions $u(t,x)=u(t,|x|)=u(t,r)$. It is possible to find in this way a class of locally integrable solutions $E_c(t,x)=cH(t)\,t^{-\frac{n}{2}}e^{-|x|^2/4t}$, with $H(t)$ the Heaviside function (although it is easier to proceed via the Fourier transform).

[14]Observe, however, that (by the product rule) $r^{n-2}\cdot H(r)$ has a decent amount of regularity, and its only singularity is located away from $r=1$.

Indeed $\mathcal{H}(E_c) = 0$ for all $(t, x) \neq (0,0)$. We show below that, in the whole space, $\mathcal{H}(E_c)$ is proportional to δ_0 and that we can determine the constant $c = c_n = 2^{-n}\pi^{-\frac{n}{2}}$ such that the corresponding $E = E_c$ is a fundamental solution of $\mathcal{H}$, i.e. $\mathcal{H}(E) = \delta_0$.

We could very easily reason by considering the parabolic scaling $(t, x) \to (\alpha^2 t, \alpha x)$, that $\mathcal{H}E_c = C\delta_0$ for some constant C (possibly 0). To determine the constant, we could use any test function, and it would be simple to take $H(1-t)$. More or less, this is exactly how we will proceed.

Let $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^{n+1})$,

$$\begin{aligned} < \mathcal{H}(E), \phi > &= < E, \mathcal{H}^t\phi >= -\int E(t,x)(\partial_t + \Delta)\phi(t,x)dxdt \\ &= -\lim_{\epsilon\to 0^+}\int_\epsilon^\infty \int_{\mathbb{R}^n} E(t,x)(\partial_t+\Delta)\phi(t,x)dxdt \\ &= \lim_{\epsilon\to 0^+}\int_\epsilon^\infty \int_{\mathbb{R}^n} (\partial_t - \Delta)E(t,x)\phi(t,x)dxdt + \lim_{\epsilon\to 0^+}\int_{\mathbb{R}^n} E(\epsilon,x)\phi(\epsilon,x)dx \\ &= \lim_{\epsilon\to 0^+}\int_{\mathbb{R}^n} E(\epsilon,x)\phi(\epsilon,x)dx = c_n \lim_{\epsilon\to 0^+}\epsilon^{-n/2}\int_{\mathbb{R}^n} e^{-|x|^2/4\epsilon}\phi(\epsilon,x)dx \end{aligned}$$

We now perform the change of variables $x = 2\epsilon^{1/2}y$,

$$\begin{aligned} < \mathcal{H}(E), \phi > &= 2^n c_n \lim_{\epsilon\to 0^+}\int_{\mathbb{R}^n}\phi(\epsilon, 2\epsilon^{1/2}y)e^{-|y|^2}dy = 2^n c_n \phi(0,0)\int_{\mathbb{R}^n} e^{-|y|^2}dy \\ &= \phi(0,0) \end{aligned}$$

Modulo the fact that $\int_{\mathbb{R}^n} e^{-|y|^2}dy = \pi^{n/2}$ (which will be shown later), we have proven that

$$E(t,x) = (4\pi t)^{-n/2}H(t)\ e^{-|x|^2/4t} \tag{55}$$

is a fundamental solution for $\mathcal{H}$. Notice that, for any fixed $t > 0$, $E(t,x)$ has support on all of $\mathbb{R}^n$, implying that the heat equation (in contrast to the wave equation) exhibits "infinite speed of propagation". This phenomenon is related to the parabolic scaling of the heat operator, which, in constrast to $-\partial_t^2 + \sum_i \partial_i^2$, endows time and space with different "units". Also notice that $E(t,x)$ is smooth for $t > 0$; this fact will lead to instantaneous smoothing for the initial value problem $\mathcal{H}\psi = 0$ on $(0,\infty)\times\mathbb{R}^n$, $\psi(0,x) = \psi_0(x)$.

Exercise 1. Derive a representation formula for the initial value problem. Why is it impossible to solve the heat equation backwards in time for arbitrary initial data?

Exercise 2. Show that the above representation formula for the Cauchy problem does indeed give a classical solution for sufficiently smooth data. Check that the correct boundary value is obtained. (This is in contrast to the situation with the Cauchy-Riemann equations).

Exercise 3. Write down a maximum principle for $\mathcal{C}^2$ solutions to the $\mathcal{H}\psi = 0$ in the interior of $(0,T]\times\overline{\Omega}$, for Ω open and bounded.

Exercise 4. Let $\Psi : [0,T] \times \mathbb{R}^n$ be a solution to the heat equation with decays rapidly at spatial infinity for fixed time. Show that the "energy"

$$e(t) = \int_{\mathbb{R}^n} |\Psi(t,x)|^2 dx$$

decreases in time. Deduce a uniqueness theorem for the heat equation.

Exercise 5. One often denotes by $e^{t\Delta}\psi_0$ the restriction to a hypersurface of fixed time $t > 0$ of the (canonical) solution to $\mathcal{H}\psi = 0$ on $(0,\infty) \times \mathbb{R}^n$, $\psi(0,x) = \psi_0(x)$. Show that $e^{(t+s)\Delta} = e^{t\Delta}e^{s\Delta}$ as time-evolution operators on rapidly decaying, smooth functions.

4.1. Schrödinger operator $\mathcal{S}$. The Schrödinger operator, $\mathcal{S} = i\partial_t + \Delta$ has a fundamental solution which looks, superficially, exactly like that of the Heat operator,

$$E(t,x) = (4\pi i t)^{-n/2} H(t)\ e^{i|x|^2/4t} \tag{56}$$

Yet, of course, the presence of i in the exponential factor $e^{i|x|^2/4t}$ makes a world of difference.

Exercise 1. Show that (for the appropriately chosen branch cut of log) the locally integrable function E is indeed a fundamental solution for $\mathcal{S}$.

Exercise 2. Similarly to Exercise 5 for the heat equation, one denotes the time-evolution operator for the Schrödinger equation by $e^{it\Delta}$. Show that $e^{it\Delta}$ is a unitary operator in the sense that the quantity

$$\int_{\mathbb{R}^n} |\Psi(t,x)|^2 dx$$

remains constant in time.

CHAPTER 3

Fourier transform

1. Basic properties.

Recall that if $f \in L^1(\mathbb{R}^n)$, then the Fourier transform $\mathcal{F}(f) = \hat{f}$ is defined as the continuous function

$$\hat{f}(\xi) = \int f(x) e^{-ix\xi} dx \tag{57}$$

In case that $\hat{f} \in L^1(\widehat{\mathbb{R}}^n)$, we have the inversion formula

$$f(x) = (2\pi)^{-n} \int \hat{f}(\xi) e^{ix\xi} d\xi, \tag{58}$$

whose proof we shall indicate later. To distinguish between the two conceptually, we refer to the $\mathbb{R}^n$ on which f lives as the "physical space" and the set of points ξ on which $\hat{f}$ lives as "frequency space". We denote the frequency space by $\widehat{\mathbb{R}}^n$ and endow it with the normalized measure $\frac{d\xi}{(2\pi)^n}$.

The inversion formula supplies us with a valuable heuristic understanding of what the Fourier transform does. We see that $f(x)$ can be written as some kind of linear combination of plane waves $(x \mapsto e^{ix\cdot\xi})$ and the measure $\hat{f}(\xi)\frac{\mathrm{d}\xi}{(2\pi)^n}$ describes the distribution of f over the space of frequencies. If we view the plane waves $e^{ix\cdot\xi}$ as eigenvectors of the translation operators on $\mathbb{R}^n$, we can consider the Fourier transform an attempt to simultaneously diagonalize translations. Similarly, if we view the plane waves $e^{ix\cdot\xi}$ as the eigenvectors of the operators $\{i\frac{\partial}{\partial x_j} \; : \; j = 1 \dots n\}$, which are self-adjoint with respect to the L^2 inner product (when restricted to the appropriate domain), then we see that differentiation has also been diagonalized by the Fourier transform.

With these heuristics in mind, we can begin to see how the Fourier transform might be useful for analysis. For example, if $\hat{f}$ is concentrated nearby a frequency $\xi' \in \widehat{\mathbb{R}}^n$, we expect f to behave in some ways like the plane waves nearby $e^{ix\cdot\xi'}$. For instance, f may admit a bounded, complex-analytic extension into part of $\mathbb{C}^n$. We also expect that $\partial_{x_j} f(x) \sim i\xi'_j f(x)$ so that differentiation becomes a much easier operation to study. Indeed, when we encounter Littlewood Paley theory later on, the main idea will be to decompose general functions into frequency localized components, analyze these components separately, and then reassemble.

Another important principle regarding the Fourier transform is the duality between smoothness and decay in physical and frequency space. Intuitively, a function f whose graph has sudden jumps or spikes in physical space must "be composed of" arbitrarily large frequencies, whereas when $\hat{f}$ is compactly supported, f must be globally tame. Similarly, when $\hat{f}$ is very smooth, f can decay at infinity thanks to interference (cancelation) among nearby plane waves in the inversion formula. There are not one but many formal manifestations of these basic principles all over Fourier analysis, so let us keep them in mind as we proceed to develop the theory.

The inversion formula takes particularly concrete form in the case of the Gaussian function $G(x) = e^{-|x|^2/2}$.

Lemma 1.1. *The following calculation holds true for functions of one variable and $a, b \in \mathbb{R}, b > 0$,*

$$\int_{-\infty}^{\infty} e^{iax} e^{-bx^2} = \big(\frac{\pi}{b}\big)^{1/2} e^{-a^2/4b} \tag{59}$$

Thus in $\mathbb{R}^n$, for $t > 0$

$$\int_{\mathbb{R}^n} e^{ix\cdot y} e^{-ty^2} = \big(\frac{\pi}{t}\big)^{n/2} e^{-|x|^2/4t} \tag{60}$$

In particular

$$\mathcal{F}(G)(\xi) = (2\pi)^{n/2} G(\xi) \tag{61}$$

Proof : Make the change of variables in the complex domain, $z = b^{1/2}x - \frac{a}{2b^{1/2}}i$, and denote by Γ the contour $\mathrm{Im}(z) = -\frac{a}{2b^{1/2}}$,

$$\int_{-\infty}^{\infty} e^{iax} e^{-bx^2} dx = \frac{e^{-a^2/4b}}{b^{1/2}} \int_{\Gamma} e^{-z^2} dz = \frac{e^{-a^2/4b}}{b^{1/2}} \int_{-\infty}^{\infty} e^{-x^2} dx$$

by a standard contour deformation argument. Now to calculate the integral $J = \int_{-\infty}^{\infty} e^{-x^2} dx = \pi^{1/2}$, we observe that $J^2 = \int_{\mathbb{R}^2} e^{-|x|^2} dx = \pi$ by passing to polar coordinates and from this follows (59) . Formula (60) now follows immediately.

■

We can give another proof of the above identity after reviewing some of the fundamental properties of the Fourier transform.

Proposition 1.2. *The Fourier transform is linear and verifies the following simple properties.*

- *Fourier transform takes translations in physical space $T_{x_0}f(x) = f(x-x_0)$ into modulations in frequency space $\mathcal{F}(T_{x_0}f)(\xi) = e^{-i\xi\cdot x_0}\hat{f}(\xi)$.*
- *Fourier transform takes modulations in physical space $M_{\xi_0}f(x) = e^{ix\cdot\xi_0}f(x)$ into translation in frequency space $\mathcal{F}(M_{\xi_0}f)(\xi) = \hat{f}(\xi-\xi_0)$.*
- *Fourier transform takes conjugation in physical space into conjugation and reflection in frequency, i.e. $\mathcal{F}(\bar{f})(\xi) = \overline{\hat{f}}(-\xi)$.*

- *Fourier transform takes convolution in physical space into multiplication in frequency space,* $\widehat{f*g} = \hat{f}\,\hat{g}$.
- *Fourier transform takes partial derivatives in physical space into multiplication in frequency space,* $\mathcal{F}(\partial_{x_j} f)(\xi) = i\xi_j \hat{f}(\xi)$.
- *Fourier transform takes multiplication by* x_j *in physical space into the partial derivative* ∂_{ξ_j} *in frequency space,* $\mathcal{F}(x_j f)(\xi) = i\partial_{\xi_j}\hat{f}(\xi)$.
- *We also have the simple self duality relation,*

$$\int f(x)\hat{g}(x)dx = \int \hat{f}(\xi)g(\xi)d\xi$$

- *Fourier transform takes scaling in physical space* $S_\lambda f(x) = f(\lambda x)$ *into a dual scaling in Fourier space,* $\mathcal{F}(S_\lambda f)(\xi) = \lambda^{-n}\hat{f}(\xi/\lambda)$. *Observe that* $S_\lambda(f)$ *preserves size, i.e.* $\|S_\lambda f\|_{L^\infty} = \|f\|_{L^\infty}$ *while the dual scaling* $S_\lambda^* f = \lambda^{-n} f(x/\lambda)$ *preserves mass, that is* $\|S_\lambda^* f\|_{L^1} = \|f\|_{L^1}$.

Proof Almost all of the above properties reduce to simple identities about exponentials when we specialize to the case where f and g are point masses in the physical or frequency space[1] (and hence are plane waves in the dual space) – the identities themselves may even be regarded as continuous, (bi)linear extensions of these special cases. ■

Using these properties, we can give another proof of (59). Thanks to the scaling identity, it suffices to consider $b = \frac{1}{2}$ and compute the Fourier transform of $G(x) = e^{-x^2/2}$. Taking the Fourier transform of the identity, $\frac{dG}{dx} = -xG(x)$ and applying the properties above, we see that $\widehat{G}$ satisfies the same differential equation in ξ, and is therefore is of the form $\widehat{G}(0)e^{-\xi^2/2}$. Since we have already shown that $\widehat{G}(0) = \int e^{-x^2/2} = \pi^{1/2}$, this completes the second proof of (59).

Let $G_{\lambda,x_0,\xi_0}(x) = e^{ix\cdot\xi_0} G((x-x_0)/\sqrt{\lambda})$ be a translated, modulated, rescaled Gaussian. Then,

$$\begin{aligned} \mathcal{F}(G_{\lambda,x_0,\xi_0})(\xi) &= \lambda^{n/2} e^{-i(\xi-\xi_0)\cdot x_0} \int e^{-i\sqrt{\lambda}y\cdot(\xi-\xi_0)} G(y)dy \\ &= (\pi\lambda)^{n/2} G\big(\sqrt{\lambda}(\xi-\xi_0)\big) \end{aligned}$$

We can interpret this result as saying that G_{λ,x_0,ξ_0} is localized at spatial position x_0, with spatial spread $\Delta x \approx \sqrt{\lambda}$, and at frequency position ξ_0 with frequency spread $\Delta\xi = 1/\sqrt{\lambda}$. Observe that $\Delta x \cdot \Delta\xi \approx 1$, so our ability to localize simultaneously in both physical and frequency space in this way seems to be limited. Surprisingly, this construction is in some sense the best we can do, and it is our first encounter with the "uncertainty principle" of Fourier analysis, which, in its various manifestations, states that there is a bound on how well one can simultaneously localize in both frequency and physical space.

[1] The Fourier transform of a finite, Borel measure is the continuous function $\hat{\mu}(\xi) = \int e^{-ix\cdot\xi} d\mu(x)$..

We now prove our first important manifestation of the duality between smoothness and decay[2].

Proposition 1.3 (Riemann Lebesgue)**.** *Given an arbitrary $f \in L^1(\mathbb{R}^n)$ we have, $\|\hat{f}\|_{L^\infty} \lesssim \|f\|_{L^1}$. Moreover, $\hat{f}(\xi) \to 0$ as $|\xi| \to \infty$.*

Proof : Only the last statement requires an argument. Observe that if $f \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, then we can use integration by parts to conclude that $\hat{f}$ decays rapidly. Indeed for any multi-index α, $|\alpha| = k \in \mathbb{N}$,

$$\begin{aligned} \xi^\alpha \hat{f}(\xi) &= i^k \int \partial_x^\alpha e^{-ix\xi} f(x)dx = (-i)^k \int e^{-ix\xi} \partial_x^\alpha f(x)dx \\ |\xi^\alpha \hat{f}(\xi)| &\lesssim \int |\partial_x^\alpha f(x)dx| \leq C_\alpha \end{aligned}$$

for some constant C_α. Thus, $|\hat{f}(\xi)| \lesssim (1 + |\xi|)^{-k}$ which proves the statement in this case. For general $f \in L^1(\mathbb{R}^n)$, given $\epsilon > 0$, we can choose $g \in \mathcal{C}_0^\infty$ such that $||f - g||_{L^1} \leq \frac{\epsilon}{2}$. From the preceding, we know that $|\hat{g}(\xi)| \leq \frac{\epsilon}{2}$ if $|\xi| > M = M_\epsilon$ sufficiently large and therefore,

$$\sup_{|\xi|>M} |\hat{f}(\xi)| \leq \|f - g\|_{L^1(\mathbb{R}^n)} + \sup_{|\xi|>M} |\hat{g}(\xi)| \leq \epsilon$$

■

2. The Schwartz Space and the Inversion Formula

Many of the operations on smooth functions extend naturally to distributions (by duality with $\mathcal{C}_0^\infty$), and we would like to see how the Fourier transform extends to distributions. The only possible extension would have to be consistent with the formula

$$< \hat{u}, \phi > = < u, \hat{\phi} >$$

but this formula does not make sense for distributions $u \in \mathcal{D}'(\mathbb{R}^n)$ and test functions $\phi \in \mathcal{C}_0^\infty(\widehat{\mathbb{R}}^n)$ because one cannot guarantee that $\hat{\phi}$ is also compactly supported. In fact, as a manifestation of the uncertainty principle, both ϕ and $\hat{\phi}$ cannot simultaneously be compactly supported unless $\phi = 0$. Thus, if we desire a symmetric theory generalizing the Fourier transform, we are lead to consider a new family of test functions (and corresponding distributions) which behaves well with respect to Fourier duality.

Definition 2.1. A function $\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ is said to be rapidly decreasing if for all multi indices α, β we have

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

[2]We remark that this proposition is not valid for measures which do not have a density function in L^1 – for example, the Fourier transform of a point mass is a plane wave, which has absolutely no decay. Thus, some amount of smoothness is necessary.

This so-called Schwarz space $\mathcal{S}(\mathbb{R}^n)$ of rapidly decreasing functions is endowed in the usual way with a natural Frechet topology. A sequence of functions ϕ_j converges to zero in this topology if, for all multi-indices α, β, $x^\alpha \partial^\beta \phi_j$ converges uniformly to zero. Note that $\mathcal{S}(\mathbb{R}^n)$ contains the compactly supported functions $\mathcal{C}_0^\infty(\mathbb{R}^n)$. Since $\mathcal{C}_0^\infty(\mathbb{R}^n)$ is dense in the $L^p(\mathbb{R}^n)$ spaces, for $1 \leq p < \infty$, $\mathcal{S}(\mathbb{R}^n)$ is also dense in the L^p spaces. It is also easy to check that $\mathcal{C}_0^\infty(\mathbb{R}^n)$ is dense in $\mathcal{S}(\mathbb{R}^n)$.

We have the following important fact, which is the reason for considering the Schwarz space in our context:

Proposition 2.2. *The Fourier transform is an isomorphism of $\mathcal{S}(\mathbb{R}^n)$ onto $\mathcal{S}(\widehat{\mathbb{R}}^n)$ with inverse given by the inversion formula* (58)*. Moreover we have the Plancherel identity, for all $f, g \in \mathcal{S}(\mathbb{R}^n)$,*

$$(f, g)_{L^2} = \int_{\mathbb{R}^n} f(x)\bar{g}(x)dx = (2\pi)^{-n} \int \hat{f}\bar{\hat{g}}d\xi = (\hat{f}, \hat{g})_{L^2(\widehat{\mathbb{R}}^n)} \tag{62}$$

In particular we have the Parseval identity $\|f\|_{L^2(\mathbb{R}^n)} = \|\mathcal{F}(f)\|_{L^2(\widehat{\mathbb{R}}^n)}$.

Proof : Observe that $|\xi^\alpha \partial^\beta \hat{\phi}(\xi)| = |\widehat{x^\beta \partial^\alpha \phi}|$ and that $\partial^\alpha \phi(x)$ decays faster than $|x|^{-|\beta|-n-1}$. Thus we easily infer that $\mathcal{F}$ maps $\mathcal{S}(\mathbb{R}^n)$ into $\mathcal{S}(\widehat{\mathbb{R}}^n)$. Let $Rf(x) = f(-x)$ and define $T = R\mathcal{F}^2$. Observe that T commutes with partial derivatives ∂_j and multiplications by x_j. Indeed, for all $j = 1, \ldots n$,

$$T(\partial_j f) = \partial_j(Tf), \quad T(x_j f) = x_j(Tf) \tag{63}$$

The inversion formula follows from the lemma.

Lemma 2.3. *A linear operator $T : \mathcal{S}(\mathbb{R}^n) \longrightarrow \mathcal{S}(\mathbb{R}^n)$ which verifies* (63) *must be of the form $T\phi = c\phi$ for some constant c.*

Proof : From the commuting property (63), we see that T is linear over the algebra of polynomial functions. As a consequence of this linearity, we can show that the value $T\phi(x_0)$ depends only on the value of $\phi(x_0)$ at the point x_0. For example, in one dimension, if ϕ vanishes at the point x_0, then we may write

$$\phi(x) = (x - x_0) \int_0^1 \phi'(x_0 + t(x - x_0))dt = (x - x_0)\tilde{\phi}$$

with $\tilde{\phi} \in \mathcal{S}(\mathbb{R})$. Applying T to this identity, we see that $T\phi(x_0) = 0$ as well, and we may therefore write $T\phi = f_\phi \phi$ for some function f_ϕ possibly depending on ϕ.

But $f_\phi = f$ does not depend on ϕ. If ψ is any other Schwartz function, the linear combination $\psi(x_0)\phi - \phi(x_0)\psi$ vanishes at the point x_0, and applying T we conclude by the same property that $f_\phi(x_0) = f_\psi(x_0)$ at any point x_0 at which ψ and ϕ are simultaneously nonzero. It is clear that the function f must be smooth for T to map $\mathcal{S}(\mathbb{R}^n)$ into itself, and in order for T to commute with differentiation, f must be a constant. ■

To determine the constants we only have to remark that, in view of lemma 1.1 we have $T(G) = \left((2\pi)^{n/2}\right)^2 G = (2\pi)^n G$. Hence the constant $c = (2\pi)^n$ which ends the proof of the inversion formula, and the proposition, for Schwartz functions.

The Plancherel and Parseval identities are immediate consequences of the inversion formula. ■

Corollary 2.4. *The following properties hold for all functions in $\mathcal{S}(\mathbb{R}^n)$:.*

$$\begin{aligned}
\int \hat{\phi}\psi dx &= \int \phi\hat{\psi} d\xi \\
\int \phi\bar{\psi} dx &= (2\pi)^{-n}\int \hat{\phi}\bar{\hat{\psi}} dx \\
\widehat{\phi * \psi} &= \hat{\phi}\hat{\psi} \\
\widehat{\phi\psi} &= (2\pi)^{-n}\int \hat{\phi}(\xi-\eta)\hat{\psi}(\eta)d\eta = \hat{\phi}*\hat{\psi}
\end{aligned}$$

The last convolution being taken with respect to the measure on $\widehat{\mathbb{R}}^n$

We only completely worked out the proof of the inversion formula for one dimension, although the same proof requires only a miniscule generalization of the Taylor expansion to work for general n. The general case can also be deduced from the case $n = 1$ as follows: the inversion formula is true for tensor products $f_1(x_1)\cdots f_n(x_n)$ and linear combinations thereof, the delta-function is a tensor product $\delta_0(x) = \delta(x_1)\cdots\delta(x_n)$, and an arbitrary function may be written as a linear combination of delta functions $f(x) = \int f(t)\delta(x-t)dt$.

Exercise 2.5. Make the above argument into a rigorous, self-contained proof of the inversion formula for $\mathbb{R}^n$ by using approximate delta-functions.

It is worthwhile to explore the relationship of the above proof of the inversion formula via the Lemma 2.3 with other proofs of the formula. Just as a linear operator between vector spaces of finite dimension can be studied via a matrix representation, we can study the operator T in terms of its Kernel K – the distribution on $\mathbb{R}^n \times \mathbb{R}^n$ such that

$$T\phi(x) = \int \phi(x')K(x,x')dx'$$

In asserting that $T\phi(x_0)$ depends only on $\phi(x_0)$, we had proven that applying T was the same as multiplying by some function; in terms of the kernel, we had established that

$$T\phi(x) = \int \phi(x')f(x)\delta(x-x')dx'$$

In order for T to commute with differentiation – which is not so different from commuting with translation – we concluded that $f(x)$ was a constant.

But we can see directly that an equivalent formulation of the inversion formula is the distribution-theoretic identity

$$\int_{\widehat{\mathbb{R}^n}} e^{i(x-x')\cdot\xi}\frac{d\xi}{(2\pi)^n} = \delta(x-x') \tag{64}$$

which is really the special case of the inversion formula for a δ-function. Viewing the integral on the left hand side as an inner product, the above identity can be regarded as a statement that the plane waves $\xi \to e^{i\xi x}$ are in some sense "orthonormal" as x varies. Thus, in writing

$$\int f(t)\delta(x-t)dt = f(x) = \int \hat{f}(\xi)e^{i\xi}\frac{d\xi}{2\pi}$$

we might regard the Fourier transform as analogous to a change of orthonormal "basis" from δ-functions in physical space to plane waves so that the Plancharel and Parseval identities should follow immediately. For example, in one dimension, distinct plane waves are eigenfunctions for the self-adjoint operator $i\frac{d}{d\xi}$ with distinct eigenvalues, and therefore should be orthogonal as a matter of general principle – this argument can be made rigorous to show the above distribution vanishes away from $x - x' = 0$, as was essentially done in the previous proof through Taylor expansion and linearity over the polynomial ring. Let us mention several other ways to establish this identity and hence prove the inversion formula.[3]

It suffices to show that

$$\frac{1}{(2\pi)^n}\int e^{i\xi\cdot x}d\xi = \delta_0(x)$$

as a distribution in the variable x on $\mathbb{R}^n$ – this translation invariance corresponds to T commuting with $\frac{d}{dx}$ in the previous proof. By viewing the above distribution as a tensor product, it would suffice to consider the case $n = 1$, but let us refrain from doing so. Recall that every distribution supported at the origin is a finite linear combination of derivatives of $\delta(x)$, and hence the δ function itself is, up to a constant, the only distribution homogeneous of degree $-n$ supported at 0 – these facts are easily established by Taylor expansion. As the integral on the left hand side is clearly homogeneous of degree $-n$ in x, we will have proven the identity up to a constant if we can show that

$$\int e^{i\xi\cdot x}d\xi \tag{65}$$

is supported at the origin – in this precise sense, a plane wave $\xi \mapsto e^{i\xi\cdot x}$ is zero "on average".

Heuristically, let us outline a few ways to perform this calculation. Pretend that the integral 65 is a classical integral and that $x \neq 0$ is fixed. If we view the plane waves as eigenfunctions of differential operators, we may integrate in ξ by parts using the identity

$$\int e^{i\xi\cdot x}d\xi = \int \frac{1}{|x|^2}\Delta_\xi e^{i\xi\cdot x}d\xi,$$

or alternatively we can rotate to the case $x = |x|(1, 0, \dots, 0)$ and integrate by parts using the identity $\frac{1}{i|x|}\frac{\partial e^{i\xi\cdot x}}{\partial\xi_1}$. If we would rather view the plane waves as eigenfunctions of translation operators, we may show the integral is zero for $x \neq 0$ by translating in the ξ variable

$$\int e^{i\xi\cdot x}d\xi = \int e^{i(\xi-\xi')\cdot x}d\xi = e^{-i\xi'\cdot x}\int e^{i\xi\cdot x}d\xi$$

[3] As remarked, the plane waves may be viewed as eigenfunctions of the commuting family of self-adjoint operators $f \mapsto i\frac{d}{dx^i}f$. Dually, the delta functions, which are similarly "orthonormal" in the sense that $\int \delta_0(y-x)\bar{\delta}_0(y'-x)dx = \delta(y-y')$, can be viewed as eigenfunctions for the commuting family of operators $f \mapsto x^i f$.

by some appropriate frequency ξ' depending on x. Of course these are all heuristic lines of attack which treat the integral as though it were a classical one, and we cannot treat x as a fixed point, but we can make these arguments rigorous by fixing a test function localized around x and produce a complete argument akin to the following.

Proof Let $\psi \in \mathbb{C}_0^\infty(\mathbb{R}^n)$ be supported away from the origin. Let ϕ be a smooth, rapidly decreasing function with $\phi(0) = 1$.

$$\begin{aligned}\int\int \psi(x)e^{i\xi\cdot x}dxd\xi &= \lim_{\delta\to 0^+}\int\left(\int \psi(x)e^{i\xi\cdot x}dx\right)\phi(\delta\xi)d\xi \\ &= \lim_{\delta\to 0^+}\int \psi(\delta x)\left(\int e^{i\xi\cdot x}\phi(\xi)d\xi\right)dx\end{aligned}$$

This limit is zero by the dominated convergence theorem (the $d\xi$ integral is a rapidly decreasing function of x and $\psi(\delta x) \to 0$). ∎

Without assuming anything about the support of ψ, the above proof would have established the Inversion Formula directly with the constant had we chosen a ϕ (such as a Gaussian) whose Fourier transform was understood. Indeed, if we know the Inversion Formula for a Gaussian, the inversion formula is true for rescalings and translates of Gaussians. As a limiting case, the Inversion Formula holds for any δ function, and hence for an arbitrary function by the decomposition $f(x) = \int f(t)\delta(x-t)dt$.

Exercise 2.6. Create a self-contained, direct proof of the Inversion Formula from the case of a Gaussian.

In the case $n = 1$, there is also a more complex-analytic way to evaluate the distribution-theoretic integral $\int_{-\infty}^{\infty} e^{i\xi x}d\xi$, which not only determines the constant[4], but directly relates the 2π in the inversion formula to the 2π in the Cauchy Integral formula (the circumference of a circle). Basically, one uses complex-analytic extensions of the plane waves and the formula (31) for $\frac{1}{x+i0}$ in order to decompose

$$\begin{aligned}\int_{-\infty}^{\infty} e^{i\xi x}d\xi &= \int_{-\infty}^{0} e^{i\xi x}d\xi + \int_{0}^{\infty} e^{i\xi x}d\xi \\ &= \lim_{y\to 0^+}\int_{-\infty}^{0} e^{i\xi(x-iy)}d\xi + \int_{0}^{\infty} e^{i\xi(x+iy)}d\xi \\ &= i\left(\frac{-1}{x-i0} + \frac{1}{x+i0}\right) \\ &= 2\pi\delta(x)\end{aligned}$$

[4]Incidentally, this argument also determines the integral of a Gaussian indirectly.

3. Extension of the Fourier Transform

As a corollary to the Parseval and Plancherel formulas we can extend our definition of the Fourier Transform to $L^2(\mathbb{R}^n)$ functions by a simple density argument. Indeed for any $u \in L^2$ we can choose a sequence of $\mathcal{S}(\mathbb{R}^n) \subset L^1$ functions u_j converging to u in the L^2 norm. By Plancherel, $\|\mathcal{F}(u_j) - \mathcal{F}(u_k)\|_{L^2} \lesssim \|u_j - u_k\|_{L^2}$. Hence the sequence $\mathcal{F}(u_j)$ forms a Cauchy sequence in L^2 and therefore converges to a limit which we may call $\hat{u}$. Clearly, this definition does not depend on the particular sequence. Moreover one can easily check that the Parseval identity extends to all L^2 functions. Since $\mathcal{F}$ is therefore an isometry onto its image, its image must be closed, but then the image must be all of L^2 since the image also contains the Schwartz functions. Thus we have proved,

Theorem 3.1. *The Fourier transform is an isometry of the Hilbert spaces $L^2(\mathbb{R}^n)$ and $L^2(\widehat{\mathbb{R}}^n)$.*

We can extend the Fourier transform even further to a special class of distributions defined on $\mathbb{R}^n$.

Definition. We define a tempered distribution to be an element in the dual space of the Schwarz space.

Example. While e^x is not a tempered distribution on $\mathbb{R}$ because it grows too quickly, the function $e^x \cos(e^x) = \frac{d}{dx} \sin(e^x)$ is an example. Here we make the usual identification of a function with a distribution.

Note that the tempered distributions embed continuously into the space of ordinary distributions defined earlier. In analogy with the properties of ordinary distributions, for every tempered distribution u, there exists a natural number N and a constant $C = C_{\alpha,\beta}$ such that

$$| < u, \phi > | \le C \sum_{|\alpha|,\, |\beta| \le N} \sup |x^\alpha \partial^\beta \phi|, \phi \in \mathcal{S}(\mathbb{R}^n)$$

We can now define the Fourier transform of a tempered distribution; namely,

$$< \hat{u}, \phi > = < u, \hat{\phi} > .$$

One easily checks that this defines a tempered distribution $\hat{u}$ for every tempered u. Moreover, all the properties of the Fourier transform, which have been verified for Schwartz functions in $\mathcal{S}(\mathbb{R}^n)$ can be easily extended to all tempered distributions. In particular, since all L^p spaces are included in $\mathcal{S}'(\mathbb{R}^n)$ we have a definition of Fourier transform for all such spaces. Observe that, in the case of L^1 this definition coincides with the definition given in (57).

The following simple, and very useful, formulas for the Fourier transform of the Dirac measure δ_0 now make sense:

$$\mathcal{F}(\delta_0) = 1, \qquad \mathcal{F}(1) = (2\pi)^n \delta_0 \tag{66}$$

Observe also that if we denote by $\text{sign}(x)$ the one dimensional tempered distribution given by the locally integrable function $\frac{x}{|x|}$ we have,

$$\widehat{\text{sign}}(\xi) = -2i\,\text{pv}(\xi) \tag{67}$$

Indeed $\text{sign}'(x) = 2\delta_0$. Hence, $i\xi\,\widehat{\text{sign}}(\xi) = 2$. Therefore, for any rapidly decreasing ϕ, we have

$$i\int \text{sign}(x)\widehat{x\phi}(x)dx = 2\hat{\phi}(0) = 2\int \phi(x)dx$$

Also, observe that $\widehat{\text{sign}}(x)$ is an odd distribution so that whenever $\phi(x) = \phi(-x)$ is an even test function, then $< \widehat{\text{sign}}, \phi >= 0$. Now given a general test function ϕ, write $\phi = \frac{1}{2}(\phi(x) + \phi(-x)) + \frac{1}{2}(\phi(x) - \phi(-x)) = \phi_{ev} + \phi_{odd}$. Hence, from the preceding, we infer that

$$< \widehat{\text{sign}}, \phi >=< \widehat{\text{sign}}, \xi\big(\frac{1}{\xi}\phi_{odd}\big) >= -2i < \text{pv}(\frac{1}{\xi}), \phi >$$

as desired.

This fact may also be observed more directly by evaluating the distribution-theoretic integral

$$\int \text{sign}(x)e^{-i\xi x}dx$$

along the same lines as the complex-analytic proof of the Fourier Inversion Formula outlined in the previous section.

Exercise 1. Show that the only harmonic functions which are tempered distributions are polynomials.

Exercise 2. Let $f(x) = e^{-|x|} \in L^1(\mathbb{R})$. Compute $\hat{f}(\xi)$ (and hence $\hat{f}(0) = \int f(x)dx = 2$) using the fact that f satisfies a simple, second order differential equation. Comment on the precise amounts of regularity and decay of f and $\hat{f}$ and how they can be anticipated from the physical space representation. Note that $\hat{f}$ continues meromorphically into the complex plane – by considering correlations against complex plane waves $x \to e^{izx}$, $z \in \mathbb{C}$, you can anticipate the location of the poles from the form of f in physical space.

Exercise 3. Suppose that u is a tempered distribution which is invariant under translation by a subgroup S of $\mathbb{R}^n$ – for instance u could be periodic or a function of less than n of the variables. Why can we assume S is closed? Show that the Fourier transform $\hat{u}$ is supported on the annihilator subgroup $S^\perp$ of plane waves which are invariant under S.

$$S^\perp = \{\xi \mid e^{i\xi\cdot x} = 1 \text{ for all } x \in S\}$$

4. Uncertainty principle and localization

On the real line let the operators X, D defined by,

$$Xf(t) = tf(t), \qquad Df(t) = -if'(t)$$

Observe that,

$$[D, X]f = DXf - XDf = -if$$

This lack of commutation is responsible for the following:

Proposition 4.1 (Heisenberg uncertainty principle)**.** *The following inequality holds,*

$$\|Xf\|_{L^2} \cdot \|Df\|_{L^2} \geq \frac{1}{2}\|f\|_{L^2}^2$$

Proof : Observe, using the commutator relation above,

$$0 \leq \|(aX + ibD)f\|_{L^2}^2 = a^2\|Xf\|_{L^2}^2 + b^2\|Df\|_{L^2}^2 - ab\|f\|_{L^2}^2$$

Now, miniize the right hand side by choosing $a = \|Df\|_{L^2}$ and $b = \|Xf\|_{L^2}$. ■

The uncertainty principle, which can informally be described as[5] $\Delta x \cdot \Delta\xi \geq 1/2$, places a limit on how accurately we can localize a function, or any other relevant object, simultaneously in both space and frequency. Let us investigate these localizations in more detail.

4.2. Physical space localization. If we want to localize a function f to a domain $D \subset \mathbb{R}^n$ we may simply multiply f by the characteristic function χ_D. The problem with this localization is that the resulting function $\chi_D f$ is not smooth even if f is. To correct for this we choose $\phi_D \in \mathcal{C}_0^\infty(D)$ in such a way that ϕ_D is not too different from χ_D. In the particular case when D is a ball $B(x_0, R)$ centered at x_0 we can choose ϕ_D to be 1 on the ball $B(x_0, R)$ and zero outside the ball $B(x_0, 2R)$. This leads to the following bounds for the derivatives of ϕ_D,

$$|\partial^\alpha \phi_D| \lesssim R^{-|\alpha|}.$$

In general given a domain D to which we can associate a length scale R (such as its diameter or distance from a fixed point in its interior), we can find a function $\phi_D \in \mathcal{C}_0^\infty(D)$ such that,

$$|\partial^\alpha \phi_D| \lesssim R^{-|\alpha|}, \tag{68}$$

for all multi-indices $\alpha \in \mathbb{N}^n$.

A general remark: derivative estimates of the form (68) are very common in analysis and almost always arise when the function obeying the estimates comes from a rescaled version of another function, whose derivatives are simply bounded. That is why the particular exponent which appears is consistent with dimensional analysis.

[5]This notation comes from quantum mechanics, where one normalizes $\|f\|_{L^2} = 1$ and interprets both $|f(x)|^2$ and $\frac{|\hat{f}(\xi)|^2}{2\pi}$ as probability densities over states in position and momentum space respectively. In this setting the Heisenberg uncertainty principle gives a lower bound for the product of the standard deviations of position and momentum.

4.3. Frequency space localization. . Just like before we can localize a function to a domain $D \subset \mathbb{R}^n$ in frequency space by $\mathcal{F}^{-1}(\chi_D \hat{f})$. Once more, it often pays to use a smoother version of cut-off, thus we set,

$$\widehat{P_D f}(\xi) = \phi_D \hat{f}(\xi).$$

P_D is an example of a Fourier multiplier operator, that is an operator of the type:

$$\widehat{T_m f}(\xi) = m(\xi)\hat{f}(\xi). \tag{69}$$

with $m = m(\xi)$ a given function called the symbol of the operator. Clearly,

$$T_m f(x) = f{*}K(x) = \int f(x-y)K(y)dy \tag{70}$$

where K, the kernel of T, is the inverse Fourer transform of m,

$$K(x) = (2\pi)^{-n} \int e^{ix\cdot\xi} m(\xi) d\xi$$

Clearly any linear differential operator $P(\partial)$ is a multiplier with symbol $P(i\xi)$.

To compare the action, in physical space, between rough and smooth cut-off operators it suffices to look at the corresponding kernels K. Let $I = [-1, 1] \subset \mathbb{R}$ and χ_I the rough cut-off (while ignoring the 2π constants). The corresponding kernel

$$K(x) = \int_{-1}^{1} e^{ix\cdot\xi} d\xi = 2\frac{sinx}{x}$$

decays very slowly as $|x| \to \infty$. Because of this the operator

$$\mathcal{F}^{-1}(\chi_I \hat{f})(x) = 2\int \frac{sin(x-y)}{(x-y)} f(y)dy$$

has very poor localization properties. Indeed, the operator spreads around to the whole $\mathbb{R}$ any function supported in some set $J \subset \mathbb{R}$. This situation corresponds to a perfect localization in frequency space and a very bad one in physical space. The exact opposite situation occurs when we do the rough cut-off localization $\chi_I f$ in physical space. On the other hand, when we use a smooth cut-off ϕ_I in frequency space, then the frequency cutoff operator $P_I f = \mathcal{F}^{-1}(\phi_I \hat{f})$ is of the form $f \to K * f$ where the kernel

$$K(x) = \int_{\mathbb{R}} e^{ix\cdot\xi} \phi_I(\xi) d\xi$$

is rapidly decreasing. In this case, we can prove that

Lemma 4.4. *Let $I = [-1, 1]$, ϕ_I a smooth cut-off on I and $P_I f = \mathcal{F}^{-1}(\phi_I \hat{f})$. Then, if f is any L^2 function supported on a set $D \subset \mathbb{R}$,*

$$|P_I(f)(x)| \lesssim C_j \|f\|_{L^2} \big(1 + dist(x, D)\big)^{-j}$$

for all $j \in \mathbb{N}$.

Thus P_I spreads the support of any function f by a distance $O(1)$ plus a rapidly decreasing tail.

Exercise. Show that there exists no non-trivial function ϕ such that both ϕ and $\mathcal{F}(\phi)$ are compactly supported.

The above discussion can be easily extended to higher dimensions. In particular we can get a qualitative description of functions in $\mathbb{R}^n$ whose Fourier support is restricted to a ball $B_R = B(0,R)$ centered at the origin. Let ϕ_R be a smooth cut-off for B_R. More precisely we take it of the form

$$\phi_R(\xi) = \phi(\xi/R)$$

ϕ a smooth cut-off for B_1, i.e. ϕ is smooth, identically equal to 1 on B_1 and supported, say, in B_2. It is easy to check the estimate for any multi-index α,

$$\sup_\xi |\partial_\xi^\alpha \phi_R(\xi)| \le c_\phi R^{-|\alpha|},$$

with a constant c_ϕ depending only on the fixed ϕ and its derivatives.

If f is a function whose Fourier support is restricted to B_R then $\hat{f} = \phi_R \hat{f}$. Hence,

$$f(x) = \int_{\mathbb{R}^n} f(y) K_R(x-y)dy \tag{71}$$

where $K_R(x) = \mathcal{F}^{-1}(\phi_R)$.

Lemma 4.5. *The kernel $K_R(x)$ verifies the estimates,*

$$|\partial_x^\alpha K_R(x)| \le C_{N,\alpha} R^{|\alpha|} R^n (1 + |x|R)^{-N}, \tag{72}$$

for all $R > 0$, any $N \in \mathbb{N}$ and multi-index $\alpha \in \mathbb{N}^n$, with a constant $C_{N,\alpha}$ which depends only on N, α, dimension n and choice of the fixed test function ϕ.

Proof Indeed, integrating by parts,

$$\begin{aligned} K_R(x) &= \int_{\mathbb{R}^n} e^{ix\cdot\xi} \phi_R(\xi) d\xi = \int_{\mathbb{R}^n} (\frac{-1}{ix})^\alpha \partial_\xi^\alpha (e^{ix\cdot\xi}) \phi_R(\xi) d\xi \\ &= \int_{\mathbb{R}^n} (\frac{1}{ix})^\alpha e^{ix\cdot\xi} \partial_\xi^\alpha \phi_R(\xi) d\xi \end{aligned}$$

Thus, for any α, $|\alpha| = N$, denoting by $|B_R| = c_n R^n$ the volume of B_R,

$$|x^\alpha K_R(x)| \le \int_{\mathbb{R}^n} |\partial_\xi^\alpha \phi_R(\xi)| \le c_\phi R^{-N} |B_R| \le c_n c_\phi R^{-N+n}$$

Hence, $|K_R(x)| \le C_N R^n (|x|R)^{-N}$, for a constant C_N which depends on N, n and the fixed ϕ. On the other hand, for $|x| \le R^{-1}$, $|K_R(x)| \lesssim R^n$. Hence, for every $N \in \mathbb{N}$,

$$|K_R(x)| \lesssim C_N R^n (1 + |x|R)^{-N}.$$

It is easy to check also that each derivative of K_R costs us a factor of R, proving (72). ∎

Now back to (71) we have

$$\begin{aligned} |\partial^\alpha f(x)| = |\int_{\mathbb{R}^n} f(y)\, \partial^\alpha K_R(x-y) dy| &\lesssim R^{|\alpha|+n} \int_{\mathbb{R}^n} |f(y)|(1 + R|x-y|)^{-N} dy \\ &\lesssim R^{|\alpha|+n} \|f\|_{L^1} \end{aligned}$$

Also, by Hölder's inequality with $\frac{1}{p}+\frac{1}{p'}=1$,

$$
\begin{aligned}
|\partial^\alpha f(x)| &\lesssim \|f\|_{L^p}\|\partial^\alpha K_R\|_{L^{p'}} \lesssim R^{|\alpha|}R^n R^{-n/p'}\|f\|_{L^p}\\
&\lesssim R^{|\alpha|+n/p}\|f\|_{L^p}
\end{aligned}
$$

We have just proved the following version ($L^p - L^\infty$ version) of the very important Bernstein inequality,

Proposition 4.6. *Assue that f is an L^p function which has its fourier transform supported in the ball $B_R = B(0,R)$. Then f has infinitely many derivatives bounded in L^∞ and we have,*

$$\|\partial^\alpha f\|_{L^\infty(\mathbb{R}^n)} \lesssim R^{n/p+|\alpha|}\|f\|_{L^p}.$$

Remark. Observe that the proposition could have been proved by reducing it to the particular case of $R=1$. More precisely assume that the result is true for $R=1$ and consider a function f whose Fourier transform is supported in B_R. Let $g(x) = R^{-n}f(R^{-1}x)$ and observe that, $\operatorname{supp}\hat{g}(\xi) = \operatorname{supp}\hat{f}(R\xi) \subset B_1$ and therefore we have, $\|\partial^\alpha g\|_{L^\infty(\mathbb{R}^n)} \lesssim \|g\|_{L^1} = R^{-n}R^{n/p}\|f\|_{L^p}$. Thus, $\|\partial^\alpha f\|_{L^\infty(\mathbb{R}^n)} \lesssim R^{n/p+|\alpha|}\|f\|_{L^p}$.

As we will very often see during these notes, dimensional analysis can be used to rapidly figure out the exponent which arises above. For example, if we regard the spatial variables as having a scale L, so that the volume element dx has scale L^n, then the frequency variables have units $R \sim L^{-1}$. We see that $\|\partial^\alpha f\|_{L^\infty}$ has a scale $L^{-|\alpha|}$ and $\|f\|_{L^p} = \left(\int |f(x)|^p dx\right)^{1/p}$ has scale $L^{n/p}$ – thus the power of R appearing makes the estimate consistent with dimensional analysis.

Qualitatively, the Berenstein estimate embodies some basic intuition regarding L^p norms and the Fourier Transform. In a sense, the higher L^p norms of a function such as the L^∞ norm control a functions ability to blow up in a localized region of space, whereas lower L^p norms control growth at infinity. The Berenstein estimate says that if a frequency-localized function does not grow at infinity (i.e. has a bounded L^p norm, $1 \le p < \infty$), then due to the absense of high-frequency components, the function is prohibited from having localized singularities as well (i.e. the L^∞ norm is bounded).

A main reason we are interested in such estimates is not that we often run into functions with compactly supported Fourier transforms, but rather that we often decompose more general functions into a sum of parts which are frequency localized. We shall return to this idea in our study of Littlewood-Paley theory.

5. Applications to PDE

Consider the initial value problems for our basic PDE's in $\mathbb{R}\times\mathbb{R}^n$, written in the form

$$\partial_t \phi = \Delta \phi, \qquad \phi)0, x) = f(x) \tag{73}$$

$$\partial_t \phi = i\Delta \phi, \qquad \phi(0, x) = f(x) \tag{74}$$

$$\partial_t^2 \phi = \Delta \phi, \qquad \phi(0, x) = f(x), \quad \partial_t \phi(0, x) = g(x) \tag{75}$$

$$\partial_t^2 \phi = -\Delta \phi, \qquad \phi(0, x) = f(x), \quad \partial_t \phi(0, x) = g(x) \tag{76}$$

In each of these cases we can write down solutions using the Fourier transform method. More precisely we can take the Fourier transform of each equation, set

$$\hat{\phi}(t, \xi) = \int e^{-ix\cdot\xi} \phi(t, x) dx$$

and solve the resulting differential equation in t. Once this is done we obtain our solution simply using the inverse Fourier transform, i.e.

$$\phi(t, x) \quad = \quad \int e^{ix\cdot\xi} \hat{\phi}(t, \xi) \frac{d\xi}{(2\pi)^n}$$

In the case of the heat equation (73) we derive,

$$\phi(t, x) = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{ix\xi} e^{-t|\xi|^2} \hat{u}_0(\xi) \frac{d\xi}{(2\pi)^n} \tag{77}$$

while in the case of the Schrödinger equation,

$$\phi(t, x) = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{ix\xi} e^{-it|\xi|^2} \hat{u}_0(\xi) \frac{d\xi}{(2\pi)^n} \tag{78}$$

Exercise 1. Show how to relate the formulas (77) and (78) to the physical space formulas (55) and (56).

In the particular case of the wave equation (75) we derive,

$$\phi(t, x) \quad = \quad \int_{\mathbb{R}^n} e^{ix\cdot\xi} \big(\cos t|\xi| \hat{f}(\xi) + \frac{\sin t|\xi|}{|\xi|} \hat{g}(\xi) \big) \frac{d\xi}{(2\pi)^n} \tag{79}$$

Exercise 2. Derive a formula similar to (79) for the Laplace equation (76). Show, using these formulas that (75) has solutions for all $f, g \in \mathcal{S}(\mathcal{R}^n)$ while (76) does not. Show however that if we only prescribe $\phi(0, x) = f$ (this is the Dirichlet problem for the Laplacian $\partial_t^2 + \Delta$ in $\mathbb{R}^{n+1}$), then the problem has a unique solution ϕ, which decays to zero as $|t| + |x| \to \infty$, for all functions $f \in \mathcal{S}(\mathbb{R}^n)$.

Exercise 3. Show, in the special case of dimension $1+3$, how to pass from formula (79) to the Kirchoff formula (48)

$$\phi(t, x) \quad = \quad \partial_t \big((4\pi t)^{-1} \int_{|x-y|=t} f(y) da(y) \big) + (4\pi t)^{-1} \int_{|x-y|=t} g(y) da(y) \tag{80}$$

which is consistent with the formulas derived in the previous chapter, based on the explicit calculation of the fundamental solution.

It is interesting to make a comparison between the Fourier based formula (79) and the Kirchoff formula (80). Observe that it is quite easy, using Parseval, to derive the global energy identity from (79),

$$\int_{\mathbb{R}^n} (|\partial_t\phi|^2 + |\nabla\phi|^2) = \int_{\mathbb{R}^n} \big(|\nabla f|^2 + |g|^2\big)dx$$

while obtaining such an identity from (80) seems not at all obvious, in fact quite implausible. On the other hand (80) is perfect for giving us domain of influence information. Indeed we read immediately from the formula that if the data f, g is supported in ball $B_a = \{|x - x_0| \le a\}$ than $\phi(t,x)$ is supported in the ball $B_{a+|t|}$ for any time t. This fact, on the other hand, does not at all seem transparent[6] in the Fourier based formula[7] (79). The fact that different representations of solutions have different, even opposite, strengths and weaknesses has important consequences for constructing parametrices, i.e. approximate solutions, for more complicated, linear variable coefficient or nonlinear wave equations. There are two type of possible constructions, those in physical space, which mimic the physical space formula (80) or those in Fourier space, which mimic formula (79). The first are called Kirchoff-Sobolev, or Hadamard parametrices while the second are called Lax parametrices, or, more generally, Fourier integral operators.

[6] Support information can ve derived by Paley-Wiener type results.

[7] Support information can be extracted however from (79) using the Paley-Wiener method, see [**Ho**] vol. 1.

CHAPTER 4

Basic Functional Inequalities

1. Basic interpolation theory

1.1. Introduction. Consider the Fourier transform as a linear operator $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\widehat{\mathbb{R}}^n)$. According to the Plancherel identity we have $\|\mathcal{F}(f)\|_{L^2} \le \|f\|_{L^2}$. On the other hand, we have $\|\mathcal{F}(f)\|_{L^\infty} \le \|f\|_{L^1}$. Can we get other bounds of the type $\|\mathcal{F}(f)\|_{L^q} \lesssim \|f\|_{L^p}$? It turns out that such estimates can be easily established by interpolating between the two estimates mentioned above. Complex interpolation allows us to conclude an L^p to L^q estimate for any values of p and q such that $p^{-1} + q^{-1} = 1$ and $q \ge 2$. This is known as the Young-Hausdorff inequality. Interpolation theory is particularly useful for linear multiplier operators of the form

$$\widehat{T_m f}(\xi) = m(\xi)\hat{f}(\xi)$$

with bounded multipler m. In view of Parseval's identity it is very easy to check the $L^2 - L^2$ estimate, $\|T_m f\|_{L^2} \lesssim \|f\|_{L^2}$. To obtain additional estimates we typically use the integral representation (70) $T_m f(x) = f{*}K(x) = \int f(x-y)K(y)dy$ where K is the inverse Fourier transform of m. If, for example, we can establish that $K \in L^1$ than we easily deduce that $\|T_m f\|_{L^1} \lesssim \|f\|_{L^1}$, since $\|f{*}K\|_{L^1} \le \|f\|_{L^1} \cdot \|K\|_{L^1}$. We thus have both $L^1 - L^1$ and $L^2 - L^2$ estimates for T_m. and it is tempting to conclude we might have an $L^p - L^p$ estimate for all $1 \le p \le 2$. Such an estimate is indeed true and follows by interpolation. On the other hand, if we can establish that $K \in L^\infty$ then $\|f{*}K\|_{L^\infty} \lesssim \|f\|_{L^1}$ and thus we can prove, by interpolation, the same $L^p - L^q$ estimate as in the Hausdorff-Young inequality.

1.2. Review of L^p spaces. Given a measurable subset $\Omega \subset \mathbb{R}^n$ the space $L^p(\Omega)$, $1 \le p < \infty$, consists in all measurables functions $f : \Omega \to \mathbb{C}$ with finite L^p norm,

$$\|f\|_{L^p} = \left(\int_\Omega |f(x)|^p \, dx \right)^{1/p} < \infty.$$

The space $L^\infty(\Omega)$ consists of all measurable functions, bounded almost everywhere, that is,

$$\|f\|_{L^\infty} = \inf \left\{ \alpha : \int_\Omega (|f(x)| > \alpha) dx = 0 \right\} = \text{ ess sup } |f| < \infty.$$

For all values of $1 \le p \le \infty$ the spaces $L^p(\Omega)$ are Banach spaces. The theory of L^p spaces generalizes when we replace the Lebesgue measure dx with a general, positive measure μ [1] The following is called Hölder's inequality

$$\|fg\|_{L^p} \leq \|f\|_{L^q}\|g\|_{L^r}, \tag{81}$$

whenever $1/p = 1/q + 1/r$. The relationship between the exponents is necessary so that both sides are homogeneous of degree $\frac{1}{p}$ in the measure. In particular, for $p = 1$,

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

where q' verifying $\frac{1}{q'} = 1 - \frac{1}{q}$ is the exponent dual to q. This inequality implies that we can identify each element $g \in L^{q'}$ with the bounded, linear functional on the Banach space L^q given by $f \mapsto \int f(x)g(x)dx$. For all $1 \leq q < \infty$ the space $L^{q'}(\Omega)$ is dual to $L^q(\Omega)$ in the sense that the above identification is an isometry (in particular, every bounded linear functional on L^q arises this way for a unique g), while the dual of $L^\infty(\Omega)$ includes $L^1(\Omega)$, but is vastly larger. Often taking the role of L^∞ is the space $\mathcal{C}_0(\mathbb{R}^n)$ of continuous functions vanishing at infinity (since they constitute the closure of $\mathcal{C}_0^\infty$ in the L^∞ norm), whose dual space is the set of finite, Borel measures on $\mathbb{R}^n$.

The different L^p norms measure different aspects of the size of a function. An estimate of a higher L^p norm such as $\|f\|_{L^\infty} \lesssim 1$ guarantees that $|f|$ does not become too large locally, whereas an estimate of a lower L^p norm such as $\|f\|_{L^1} \lesssim 1$ controls the behavior of f at infinity. The space $L^2(\Omega)$ is especially important because of its self-duality and its Hilbert space structure given by the inner product

$$< f, g >= \int_\Omega f\bar{g}dx$$

Exercise. Show that $C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$ for all $1 \leq p < \infty$.

Given a measurable function f and a positive number α, denote by $\Lambda(f,\alpha)$ the *distribution function* of f defined by

$$\Lambda(f,\alpha) = |\{x \in \Omega : |f(x)| > \alpha\}|$$

For $1 \leq p < \infty$ we have *Chebyschev*'s inequality

$$\Lambda(f,\alpha) \leq \alpha^{-p}\|f\|_{L^p}^p \tag{82}$$

which quantitatively expresses the fact that the upper contour sets of an L^p function have finite measure. It is helpful (at least as a mnemonic) to note that both sides have the same units since f and α have the same units.

Proof

$$\Lambda(f,\alpha) = \int (|f(x)| > \alpha)d\mu(x) \leq \int \left(\frac{|f(x)|^p}{\alpha^p}\right) \cdot (|f(x)| > \alpha)d\mu(x) \leq \alpha^{-p}\|f\|_{L^p}^p$$

■

[1] There are some complications, however, when the whole space cannot be written as a countable union of finite μ-measure subsets. This will not concern us in these notes.

We can write the L^p norm of f in terms of its distribution function. Indeed, the integral $\int |f|^p$ is the measure of the region bounded by the graph $\{(\beta, x) : 0 < \beta < |f(x)|^p\}$, hence

$$\int |f(x)|^p dx = \int \int_0^\infty (|f(x)|^p > \beta) d\beta dx = p \int_0^\infty \alpha^{p-1} \Lambda(f, \alpha) d\alpha \qquad (83)$$

where the last integral is obtained from the substitution $\beta = \alpha^p$.

A measurable function $f : \Omega \to \mathbb{C}$ is said to be simple if its range consists of a finite number of points in $\mathbb{C}$, that is $f = \sum_{i=1}^N a_i \chi_{A_i}$ for $a_i \in \mathbb{C}$ and $A_i \subset \Omega$ measurable. In this section we denote by $\mathcal{S}(\Omega)$ the set of all simple functions in Ω. Recall that $\mathcal{S}(\Omega)$ is dense in $L^p(\Omega)$ for all $1 \leq p \leq \infty$. The proof typically involves approximating a fixed $f(x)$ with linear combinations of characteristic functions $(f(x) \in E_\alpha)$, and letting the collection $\{E_\alpha\}$ tend towards a fine and complete partition of $\mathbb{C}$.

Exercise. Let $f(x, y)$ be a measurable function on $\Omega_1 \times \Omega_2 \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$. Prove the following version of the *Minkowski*'s inequality,

$$\| \int_{\Omega_2} f(x, y) dy \|_{L^p_x(\Omega_1)} \leq \int_{\Omega_2} \|f(x, y)\|_{L^p_x(\Omega_1)} dy$$

for $1 \leq p \leq \infty$ both by using duality and without doing so.

1.3. Three lines lemma. The method of analytic interpolation, for linear operators acting on L^p spaces, is based on a variant of the maximum modulus theorem for a strip-like domain called the three lines lemma. Consider the strip-like domain,

$$D = \{z \in \mathbb{C} : 0 < \mathrm{Re}(z) < 1\}.$$

We will denote by $\mathcal{A}_{BC}$ the set of bounded continuous functions on the closure of D which are analytic on D.

Lemma 1.4 (Three lines lemma)**.** *Let $f \in \mathcal{A}_{BC}$ such that*

$$|f(0 + ib)| \leq M_0, \quad |f(1 + ib)| \leq M_1,$$

for all $b \in \mathbb{R}$. Then for all $0 < a < 1$ and $b \in \mathbb{R}$,

$$|f(a + ib)| \leq M_0^{1-a} M_1^a.$$

Remark. Recall that $\log |f(z)|$ is a subharmonic function when f is holomorphic (and nontrivial). The bounds on the two strips guarantee that $\log |f(z)|$ obtains lesser values than the harmonic function $\phi(a + bi) = (1 - a) \log M_0 + a \log M_1$ on the boundary of D, and the conclusion of the Three Lines Lemma asserts that $\log |f(a + bi)| \leq \phi(a + bi)$ within the domain D, however we cannot simply apply the weak maximum principle, since it is not quite valid for unbounded domains (as the example $z \to e^{-ie^{i\pi z}}$ on D shows).

Proof First, by replacing f with $\frac{f(z)}{M_0^{1-z} M_1^z}$ if necessary, we can assume that $M_0 = M_1 = 1$ and it suffices to show $|f(z)| \leq 1$ throughout. When f decays to 0 as $|\Im z| \to \infty$, then one can simply apply the usual maximum modulus principle to a sufficiently large subset of D to conclude $|f(z)| \leq 1$ throughout. If this is not the case, then (because we have assumed already that $|f(z)|$ does not grow substantially as $|\Im z| \to \infty$) we can apply the same argument to the approximation $F_\epsilon(z) = e^{\epsilon(1-z)z} f(z)$ (which does decay for large $|\Im z|$) and conclude

$$|f(z)| = \lim_{\epsilon} |F_\epsilon(z)| \leq 1$$

throughout D. ■

1.5. Stein-Riesz-Thorin interpolation.

Definition 1.6. We say that a family of linear operators T_z, indexed by $z \in D$, is an *analytic family of operators* if,

(1) T_z maps simple functions into measurable functions;
(2) For any pair of simple functions $f, g \in \mathcal{S}(\Omega)$, the map $z \mapsto \int g(x) T_z f(x) dx$ belongs to $\mathcal{A}_{BC}$.

Remark 1.7. The reason for choosing simple functions as test functions in the previous definition is because they are easy to manipulate and they make a dense set in L^p for every $p \in [1, \infty)$.

Theorem 1.8. *Let T_z be an analytic family of operators and assume there are positive constants M_0, M_1 such that, for every $b \in \mathbb{R}$,*

$$\|T_{ib} f\|_{L^{q_0}(d\mu)} \leq M_0 \|f\|_{L^{p_0}(d\nu)}, \quad \|T_{1+ib} f\|_{L^{q_1}(d\mu)} \leq M_1 \|f\|_{L^{p_1}(d\nu)},$$

with $1 \leq q_0, p_0, q_1, p_1 \leq \infty$. Then, for $z = a + ib \in D$, T_z extends to a bounded operator from $L^p(d\nu)$ to $L^q(d\mu)$ and

$$\|T_z f\|_{L^q(d\mu)} \leq M_0^{1-a} M_1^a \|f\|_{L^p(d\nu)},$$

where

$$\frac{1}{p} = \frac{1-a}{p_0} + \frac{a}{p_1}, \quad \frac{1}{q} = \frac{1-a}{q_0} + \frac{a}{q_1}.$$

Proof : By changing the measures μ and ν themselves by a scalar multiple, we can assume that $M_0 = M_1 = 1$. Adopting a bilinear formulation we have to prove that

$$\left| \int g(x) T_z f(x) dx \right| \leq 1 \tag{84}$$

for every pair of simple functions f, g with $\|f\|_{L^p} = \|g\|_{L^{q'}} = 1$. Fix such a pair f, g and consider the related (analytic) families of simple functions

$$f_z(x) = |f(x)|^{\frac{p}{p(z)} - 1} f(x), \quad g_z(x) = |g(x)|^{\frac{q'}{q'(z)} - 1} g(x),$$

with the exponents,

$$\frac{1}{p(z)} = \frac{1-z}{p_0} + \frac{z}{p_1}, \quad \frac{1}{q'(z)} = \frac{1-z}{q_0'} + \frac{z}{q_1'}.$$

We can easily check that

$$|f_{ib}| \leq |f|^{p/p_0}, \quad |f_{1+ib}| \leq |f|^{p/p_1}, \quad |g_{ib}| \leq |g|^{q'/q_0'}, \quad |g_{1+ib}| \leq |g|^{q'/q_1'}.$$

Here we use the convention that $1/\infty = 0$, and in particular if $p_0 = p_1 = \infty$ then $p = p(z) = \infty$ and $f_z \equiv f$, similarly $q_0' = q_1' = \infty$ then $q' = q'(z) = \infty$ and $g_z \equiv g$. It is immediate to verify that $\|f_z\|_{L^{\mathrm{Re}(p(z))}} = \|f\|_{L^p} = 1$ and $\|g_z\|_{L^{\mathrm{Re}(q'(z))}} = \|g\|_{L^{q'}} = 1$.

Now consider the map defined on D,

$$h(z) = \int g_z(x) T_z f_z(x) dx.$$

It is not difficult to see from our construction and the linearity and analyticity properties of T_z, that $h \in \mathcal{A}_{BC}$. By hypothesis (and Hölder) we have that $|h(ib)| \leq 1$ and $|h(1+ib)| \leq 1$ for every $b \in \mathbb{R}$. It follows from the three-lines lemma that $|h(z)| \leq 1$ and in particular 84. ■

1.9. Young inequality. We often need to estimate integral operators of the form[2].

$$Tf(x) = \int k(x,y) f(y) \mathrm{d}y, \tag{85}$$

The simplest result of this type is given by Young's theorem below.

Theorem 1.10 (Young)**.** *Let $k(x,y)$ be a measurable function and assume that for some $1 \leq r \leq \infty$ we have*

$$\sup_x \|k(x,\,\cdot\,)\|_{L^r} \leq 1, \quad \sup_y \|k(\,\cdot\,,y)\|_{L^r} \leq 1.$$

Then, for $1 \leq p \leq r'$ and

$$1 + \frac{1}{q} = \frac{1}{r} + \frac{1}{p}, \tag{86}$$

we have

$$\|Tf\|_{L^q} \leq \|f\|_{L^p}. \tag{87}$$

Proof : By Hölder inequality,

$$\|Tf\|_{L^\infty} \leq \|f\|_{L^{r'}}. \tag{88}$$

On the other hand the dual operator T^* has the same form as T,

$$T^* g(y) = \int \overline{k(x,y)} g(x) \mathrm{d}x,$$

and hence,

$$\|T^* g\|_{L^\infty} \leq \|g\|_{L^{r'}},$$

which by duality gives the other endpoint

$$\|Tf\|_{L^r} \leq \|f\|_{L^1}. \tag{89}$$

Now, we can use Theorem 1.8, with $T_z \equiv T$, to interpolate between (88) and (89) and obtain (87). ■

[2] In fact, the Schwartz Kernel theorem states that every continuous linear map from $\mathcal{C}_0^\infty(\Omega_1) \to \mathcal{D}'(\Omega_2)$ is of the form (85) for some distribution $k(x,y) \in \mathcal{D}'(\Omega_1 \times \Omega_2)$

As an immediate consequence, when k is translation invariant, $k(x,y) = k(x-y)$, we obtain the well known estimate for convolutions:

$$\|k * f\|_{L^q} \leq \|k\|_{L^r} \|f\|_{L^p}, \tag{90}$$

whenever the exponents $1 \leq p, q, r \leq \infty$ satisfy (86). Note that this relationship between the exponents is necessary so that both sides will have the same degree of homogeneity in the measure.

Exercise. More generally, when $\|k\|_{L^\infty_x L^r_y} \leq 1$ fails and similarly for $\|k\|_{L^\infty_y L^r_x}$, one can reduce to the hypotheses of Theorem (1.10) by changing the measures in the x and y variables. By doing so, what "more general" Young inequality do you obtain?

Exercise. Prove, using complex interpolation, the Hausdorff-Young inequality for the Fourier transform $\mathcal{F}$,

$$\|\mathcal{F}(f)\|_{L^q} \lesssim \|f\|_{L^p}, \qquad \text{for all} \quad q \geq 2, \quad 1/q + 1/p = 1.$$

1.11. Marcinkiewicz interpolation. A slightly weaker condition than L^p integrability for a function f is the so called *weak-L^p* property.

Definition 1.12. For $1 \leq p < \infty$, we say that f belongs to weak-L^p if $\Lambda(f, \alpha) \lesssim \alpha^{-p}$, for every $\alpha > 0$. If $p = \infty$ we let weak-L^∞ coincide with L^∞.

By Chebyschev's inequality (82), any function in L^p is also in weak-L^p. The following is the simplest example of real interpolation. It applies to sublinear operators, that is,

$$|T(f+g)(x)| \lesssim |Tf(x)| + |Tg(x)|,$$

Theorem 1.13. *Consider a sublinear operator T mapping measurable functions on X to measurable functions on Y. Assume that T maps $L^{p_i}(X)$ into weak-$L^{p_i}(Y)$, with bound*

$$\Lambda(Tf, \alpha) \lesssim \alpha^{-p_i} \|f\|_{L^{p_i}}^{p_i},$$

for $i = 1, 2$ and $1 \leq p_1 < p_2 \leq \infty$. Then, for any p, $p_1 < p < p_2$, T maps $L^p(X)$ into $L^p(Y)$, with the bound

$$\|Tf\|_{L^p} \lesssim \|f\|_{L^p},$$

.

Proof : Given $f \in L^p(X)$ and $\alpha > 0$ we write $f = f^\alpha + f_\alpha$, where $f^\alpha(x) = f(x) \cdot (|f(x)| > \alpha)$ and $f_\alpha(x) = f(x) \cdot (|f(x)| \leq \alpha)$ are cutoffs of f . In particular $f^\alpha \in L^{p_1}$ and $f_\alpha \in L^{p_2}$ by Hölder's inequality.

Consider first the case $p_2 < \infty$. By our assumptions on T we have

$$\Lambda(Tf, 2\alpha) \lesssim \Lambda(Tf^\alpha, \alpha) + \Lambda(Tf_\alpha, \alpha) \lesssim \alpha^{-p_1} \|f^\alpha\|_{L^{p_1}}^{p_1} + \alpha^{-p_2} \|f_\alpha\|_{L^{p_2}}^{p_2}. \tag{91}$$

Using the distributional characterization of $\|Tf\|_{L^p}$ and Fubini's theorem, we infer that

$$\int |Tf(x)|^p \,\mathrm{d}x \lesssim \int_{0<\alpha<|f(x)|} |f(x)|^{p_1}\alpha^{p-p_1-1}\mathrm{d}\alpha\mathrm{d}x + \int_{|f(x)|\leq\alpha} |f(x)|^{p_2}\alpha^{p-p_2-1}\mathrm{d}\alpha\mathrm{d}x.$$

But $\int_0^{|f(x)|} \alpha^{p-p_1-1}\mathrm{d}\alpha \simeq |f(x)|^{p-p_1}$, since $p-p_1-1>-1$, and $\int_{|f(x)|}^{\infty} \alpha^{p-p_2-1}\mathrm{d}\alpha \simeq |f(x)|^{p-p_2}$, since $p-p_2-1<-1$, and the conclusion follows.

In the case of $p_2=\infty$ the proof is actually simpler. We only have to observe that $|Tf(x)| \gg \alpha$ implies $|Tf^\alpha(x)| \gg \alpha$, since $|Tf_\alpha(x)| \lesssim \|f_\alpha\|_{L^\infty} \leq \alpha$. Hence we can replace (91) by

$$\Lambda(Tf, C\alpha) \lesssim \Lambda(Tf^\alpha, \alpha) \lesssim \alpha^{-p_1}\|f^\alpha\|_{L^{p_1}}^{p_1},$$

where C is some positive constant, and the proof proceeds as before. ∎

2. Maximal function, fractional integration and applications

2.1. Maximal Function. A function f which is in $L^p(\mathbb{R}^n)$, for some $1\leq p\leq\infty$, may possess very bad regularity properties. Given $\alpha>0$, the set of points x where $|f(x)|>\alpha$ may merely be any measurable set (with finite measure if $p<\infty$). It is often desirable to replace f with a positive function which has (almost) the same integrability properties of f but better local regularity. This is achieved by considering maximal averages of f.

Definition 2.2. Given a measurable function on $\mathbb{R}^n$ we define its *maximal function* by

$$\mathcal{M}f(x) = \sup_{x\in B}\frac{1}{|B|}\int_B |f(y)|\mathrm{d}y.$$

Here the supremum is taken over all possible euclidean balls B containing x (not only those centered at x).

Remark 2.3. It follows immediately from the definition that $\mathcal{M}f$ is lower semi-continuous. Indeed, for every $\alpha\geq 0$, the sets $E_\alpha = \{x\in\mathbb{R}^n : \mathcal{M}f(x)>\alpha\}$ are always open: if $x\in E_\alpha$ then there exists a ball B containing x such that

$$\frac{1}{|B|}\int_B |f(y)|\mathrm{d}y > \alpha \tag{92}$$

and this also means that $\mathcal{M}f(y)>\alpha$ for every $y\in B$, hence $B\subset E_\alpha$.

By the triangle inequality we also see that $f\mapsto\mathcal{M}f$ is a *subadditive* operator,

$$\mathcal{M}(f+g)(x) \leq \mathcal{M}f(x) + \mathcal{M}g(x). \tag{93}$$

The averaging process may improve local regularity, but, because of the supremum, it is not clear whether $\mathcal{M}f$ preserves the integrability properties of f. If f is essentially bounded, then $\mathcal{M}f$ is bounded and

$$\|\mathcal{M}f\|_{L^\infty} \leq \|f\|_{L^\infty}. \tag{94}$$

But, if f is an integrable function, it doesn't follow that $\mathcal{M}f$ is integrable. Take for example $f = \chi_B \in L^1$, the characteristic function of a ball, then $\mathcal{M}f(x) \gtrsim (1+|x|)^{-n}$ which barely fails to be in L^1. Fortunately, the maximal function still retains most of the information about the integrability properties of f.

Theorem 2.4. *If $f \in L^1$ then $\mathcal{M}f$ is weakly in L^1, in the sense that for $\alpha > 0$ we have*

$$|E_\alpha| = \Lambda(\mathcal{M}f(x), \alpha) \lesssim \frac{1}{\alpha}\|f\|_{L^1} \tag{95}$$

If $f \in L^p$ with $1 < p \leq \infty$ then $\mathcal{M}f \in L^p$ and we have

$$\|\mathcal{M}f\|_{L^p} \lesssim \|f\|_{L^p}. \tag{96}$$

Proof : The second part of the statement follows from the first and the L^∞ boundedness of the maximal operator by Marcinkiewicz interpolation, Theorem 1.13. Hence, we only need to prove (95).

Let $f \in L^1$ and fix $\alpha > 0$. By the discussion in Remark 2.3 we can find a family of balls $\mathcal{B} = \{B\}$, such that $E_\alpha = \cup_{B\in\mathcal{B}} B$ and each ball B satisfies (92). If these balls were all disjoint then it would be easy to conclude, since in that case

$$|E_\alpha| \leq \sum_{B\in\mathcal{B}} |B| < \frac{1}{\alpha}\sum_{B}\int_B |f(y)|\mathrm{d}y \leq \frac{1}{\alpha}\int_{R^n} |f(y)|\mathrm{d}y.$$

In general these balls are not disjoint and we have to be more careful.

Let K be a compact subset of E_α, then it is possibile to select a finite subfamily $\mathcal{B}'$ of balls in $\mathcal{B}$ that cover K. Using the covering lemma proved below[3], Lemma 2.5, we can select among the balls in $\mathcal{B}'$ another finite subfamily $\mathcal{B}''$ made of disjoint balls (which may no longer cover K) such that

$$|\cup_{B'\in\mathcal{B}'} B'| \lesssim \sum_{B''\in\mathcal{B}''} |B''|.$$

Then, proceeding as above, we find

$$|K| \lesssim \frac{1}{\alpha}\|f\|_{L^1},$$

and taking the supremum over all possible compact sets K we finally obtain (95). ∎

Lemma 2.5. *Let $B_1, \ldots, B_N$ be a finite collection of balls in $\mathbb{R}^n$, then it is possible to select a subcollection $B_{j_1}, \ldots, B_{j_M}$, $M \leq N$, of disjoint balls such that*

$$\left|\cup_{j=1}^N B_j\right| \lesssim \sum_{k=1}^M |B_{j_k}|.$$

Proof : We can assume that the balls $B_j = B(x_j, r_j)$ are labeled so that the radii are in nonincreasing order, $r_1 \geq r_2 \geq \cdots \geq r_N$.

[3]This is sometimes known as the Vitali Covering Lemma

Take $j_1 = 1$, so that B_{j_1} is the ball with largest radius. Then by induction, define j_{k+1} to be the minimum index among those of the balls B_j which don't intersect with the previously chosen balls $B_{j_1}, \ldots, B_{j_k}$; if there are no such balls then stop at step k.

With this construction we have that each ball B_j intersects one of the chosen balls B_{j_k} with $r_j \leq r_{j_k}$, hence $B_j \subset B(x_{j_k}, 3r_{j_k})$. This implies that

$$\left|\cup_{j=1}^{N} B_j\right| \leq \left|\cup_{k=1}^{M} B(x_{j_k}, 3r_{j_k})\right| \leq 3^n \sum_{k=1}^{M} |B_{j_k}|.$$

■

2.6. Lebesgue differentiation theorem. If a function f is continuous then, clearly,

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\mathrm{d}y = f(x). \tag{97}$$

As an application of Theorem 2.4 we can show that this property continue to hold for locally integrable functions.

Corollary 2.7 (Lebesgue's differentiation theorem)**.** *If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ then* (97) *holds for almost every x.*

Proof : Since the statement is local we can assume that $f \in L^1$.

Let A_r be the averaging operator defined by $A_r f(x) = |B(x,r)|^{-1} \int_{B(x,r)} f(y)\mathrm{d}y$. The proof consist of two steps. First we prove that $A_r f \to f$ in L^1 as $r \to 0$, and then it will be enough to show that $\lim_{r\to 0} A_r f(x)$ exists almost everywhere.

For the first step, given $\epsilon > 0$, using the density of C_0 in L^1, we can always find a compactly supported continuous function g which approximates f in L^1 and have $\|A_r f - A_r g\|_{L^1} \leq \|f - g\|_{L^1} < \epsilon$ uniformly in r. Then by the uniform continuity of g, we know that $A_r g \to g$ in L^1 as $r \to 0$, hence there exists an r_ϵ such that

$$\|A_r f - f\|_{L^1} \leq \|A_r f - A_r g\|_{L^1} + \|A_r g - g\|_{L^1} + \|f - g\|_{L^1} \leq 3\epsilon,$$

for $r < r_\epsilon$.

For the second step, we define the *oscillation* of an L^1 function f by

$$\Omega f(x) = \limsup_{r\to 0} A_r f(x) - \liminf_{r\to 0} A_r f(x).$$

The oscillation is a subadditive operator, $\Omega(f+g) \leq \Omega f + \Omega g$ and is bounded by the maximal function operator, $\Omega f \leq 2\mathcal{M} f$, moreover the oscillation of a continuous function vanishes. If g is a continuous function which appoximate f in L^1 then we have that

$$\Omega f \leq \Omega(f - g) + \Omega g = \Omega(f - g) \leq 2\mathcal{M}(f - g).$$

We can apply now the weak-L^1 property of the maximal function, and for any positive α we find that

$$|\{x : \Omega f(x) > \alpha\}| \leq |\{x : \mathcal{M}(f-g)(x) > \alpha/2\}| \lesssim \frac{1}{\alpha}\|f-g\|_{L^1}.$$

Since $\|f-g\|_{L^1}$ can be arbitrarily small, we infer that set of points where the oscillation of f is positive is of measure zero. ■

2.8. Fractional integration. Let T be an integral operator acting on functions defined over $\mathbb{R}^n$ with kernel k as in (85). If the only information that we have on $k(x,y)$ is a decay estimate of the type

$$|k(x,y)| \lesssim |x-y|^{-\gamma},$$

for some $\gamma > 0$, then Young's inequality, Theorem 1.10, does not allow us to recover a good control on Tf, since the function $|\,x\,|^{-\gamma}$ fails, barely, to be in $L^{n/\gamma}$. However, the convolution has smoothing properties that imply some positive results which are contained in the following important theorem, originally proved by Hardy and Littlewood for $n = 1$ and then extended by Sobolev to $n > 1$.

Theorem 2.9 (Hardy-Littlewood-Sobolev inequality)**.** *Let $0 < \gamma < n$ and $1 < p < q < \infty$ such that*

$$1 - \frac{\gamma}{n} = \frac{1}{p} - \frac{1}{q}, \tag{98}$$

then

$$\||\cdot|^{-\gamma} * f\|_{L^q(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}. \tag{99}$$

Proof : We can split the convolution with the singular kernel into two parts:

$$I_\gamma f(x) = |\cdot|^{-\gamma} * f(x) = \int_{|y|\geq R} \frac{f(x-y)}{|y|^\gamma}\mathrm{d}y + \int_{|y|<R} \frac{f(x-y)}{|y|^\gamma}\mathrm{d}y,$$

where the radius R is a positive constant to be chosen later We estimate the first term simply by Hölder's inequality,

$$\left|\int_{|y|\geq R} \frac{f(x-y)}{|y|^\gamma}\mathrm{d}y\right| \leq \|f\|_{L^p} \left(\int_{|y|\geq R} |y|^{-\gamma p'}\mathrm{d}y\right)^{1/p'} \lesssim R^{\frac{n}{p'}-\gamma}\|f\|_{L^p},$$

where we need the integrability condition $\gamma p' > n$, which by (98) is equivalent to $q < \infty$.

For the second part we perform a dyadic decomposition around the singularity and get an estimate in terms of the maximal function,

$$\begin{aligned}\left|\int_{|y|<R} \frac{f(x-y)}{|y|^\gamma}\mathrm{d}y\right| &\leq \sum_{k=0}^{\infty} \int_{2^{-k-1}\leq \frac{|y|}{R} \leq 2^{-k}} \frac{|f(x-y)|}{|y|^\gamma}\mathrm{d}y \lesssim \\ &\lesssim \sum_{k=0}^{\infty} \frac{1}{(2^{-k}R)^\gamma} \int_{|y|\leq 2^{-k}R} |f(x-y)|\mathrm{d}y \lesssim \\ &\lesssim \sum_{k=0}^{\infty} (2^{-k}R)^{n-\gamma}\mathcal{M}f(x) \simeq R^{n-\gamma}\mathcal{M}f(x),\end{aligned}$$

where we need $\gamma < n$ for the convergence of the last geometric series.

At this point we have found that for every $x \in \mathbb{R}^n$ and every $R > 0$,

$$\big| |\cdot|^{-\gamma} * f(x) \big| \lesssim R^{\frac{n}{p'}-\gamma} \|f\|_{L^p} + R^{n-\gamma} \mathcal{M}f(x),$$

with constants independent of R and x. We optimize this inequality choosing, for each x, a radius $R = R(x)$ such that the two terms on the right hand side are equal,

$$R^{\frac{n}{p'}-\gamma} \|f\|_{L^p} = R^{n-\gamma} \mathcal{M}f(x),$$

i.e.,

$$R(x) = \left(\frac{\|f\|_{L^p}}{\mathcal{M}f(x)} \right)^{p/n},$$

and since $(n-\gamma)p/n = 1 - p/q$, we have

$$|I_\gamma f(x)| \lesssim \|f\|_{L^p}^{1-\frac{p}{q}} \mathcal{M}f(x)^{\frac{p}{q}}.$$

Then take the L^q norm on both sides,

$$\|I_\gamma f\|_{L^q} \lesssim \|f\|_{L^p}^{1-\frac{p}{q}} \|\mathcal{M}f\|_{L^p}^{\frac{p}{q}}.$$

If $p > 1$ we can conclude using the estimates for the maximal function (96). ■

Remark. The Hardy-Littlewood-Sobolev inequality has an equivalent bilinear formulation, which reads

$$\iint \frac{f(x)g(y)}{|x-y|^\gamma} \mathrm{d}x\mathrm{d}y \lesssim \|f\|_{L^{p_1}} \|g\|_{L^{p_2}},$$

for $0 < \gamma < n$ and $1 < p_1, p_2 < \infty$ such that

$$\frac{1}{p_1'} + \frac{1}{p_2'} = \frac{\gamma}{n}.$$

It is important to understand that the relation among the exponents can be quickly derived from scaling arguments. If we assign a length scale to L to the variable x, the expression

$$\| |\cdot|^{-\gamma} * f \|_{L^q(\mathbb{R}^n)} = \| \int |x-y|^{-\gamma} f(y) dy \|_{L^q_x(\mathbb{R}^n)}$$

has the units $L^{-\gamma} \cdot L^n \cdot L^{\frac{n}{q}}$, whereas $\|f\|_{L^p}$ has the units $L^{\frac{n}{p}}$. The exponents γ, q, p must relate in such a way that the exponents of both quantities match up. Indeed, if they did not, then one could deduce the failure of the estimate (99) by considering an arbitrary, nontrivial f and rescaling it to derive a contradiction.

Remark. In our proof of (99) we have not fully used the power of the Maximal function (for example, by only considering balls centered at x). In fact, the same estimate holds upon replacing the kernel $|x-y|^{-\gamma}$ with any kernel $k(x,y)$ sharing the same distribution function. A proof along these lines requires one to build up the machinery of Lorentz spaces along with a more general form of the Marcinkiewicz interpolation theorem. For this we refer to (***)

Using the Hardy-Littlewood-Sobolev inequality, we now show that it is possible to give a very short proof of the Sobolev inequality,

$$\|f\|_{L^q} \lesssim \|\partial f\|_{L^p},$$

for $n/q = n/p - 1$, in the non sharp regime $p > 1$. As with the Hardy-Littlewood Sobolev inequality, the exponents are easily deduced by considering the length scaling of both sides. The Sobolev inequality quantifies the intuitive fact that a function can only blow up somewhere if its derivatives blow up even worse.

Proof. Assume $f \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. For every unit vector ω we have

$$f(x) = -\int_0^\infty \frac{d}{dr} f(x+\omega r)\mathrm{d}r$$

To consider all possible directions in which f could grow, we integrate over the whole sphere, and recalling that the volume element in $\mathbb{R}^n$ in polar coordinates is $\mathrm{d}y = r^{n-1}\mathrm{d}r\mathrm{d}\sigma_\omega$, we find that

$$|f(x)| \lesssim \int \frac{|\partial f(y)|}{|x-y|^{n-1}}\mathrm{d}y = \left(|\cdot|^{1-n} * |\partial f|\right)(x).$$

We take the L^q norm and use (99) to get

$$\|f\|_{L^q} \lesssim \|\,|\cdot|^{1-n} * |\partial f|\,\|_{L^q} \lesssim \|\,\partial f\,\|_{L^p},$$

whenever $p > 1$ and

$$1 - \frac{n-1}{n} = \frac{1}{p} - \frac{1}{q}.$$

2.10. Sobolev Inequalities. In the previous section we have seen how to estimate the $L^q(\mathbb{R}^n)$ norm of a function in terms of an L^p norm, $1 - \frac{n-1}{n} = \frac{1}{p} - \frac{1}{q}$, $p > 1$, of the gradient of f. We shall now prove a stronger version of this.

Theorem 2.11 (Galgliardo-Nirenberg-Sobolev)**.** *The inequality*

$$\|f\|_{L^q(\mathbb{R}^n)} \lesssim \|\partial^m f\|_{L^p(\mathbb{R}^n)}, \quad f \in C_0^\infty(\mathbb{R}^n), \tag{100}$$

holds for

$$\frac{1}{q} = \frac{1}{p} - \frac{m}{n} > 0, \quad m \in \mathbb{N}, \quad (1 \le p < q < \infty). \tag{101}$$

While for $q = \infty$, we have

$$\|f\|_{L^\infty(\mathbb{R}^n)} \lesssim \sum_{k=0}^m \|\partial^k f\|_{L^p(\mathbb{R}^n)}, \quad f \in C_0^\infty(\mathbb{R}^n), \tag{102}$$

when $\frac{n}{p} - m < 0$.

Remark. We don't need to remember the precise condition (101); it can be deduced by a simple dimensional analysis. Since the estimate is homogeneous, it has to be invariant under dilations, and (101) simply says that both sides in (100) have the same scaling. Also the condition $\frac{n}{p} - m < 0$ is a comparison of the scalings of the two sides of (102) which excludes a very localized and spiky counterexample.

Remark. The following non-sharp version of estimate (100) also holds for all $1 \le p < q < \infty$ and $1/p - m/n < 1/q$,

$$\|f\|_{L^q(\mathbb{R}^n)} \lesssim \sum_{|\alpha| \le m} \|\partial^\alpha f\|_{L^p} \quad f \in C_0^\infty(\mathbb{R}^n), \tag{103}$$

Exercise. Show by an example that the inequality (102) can fail to be true for $p = n/m$. Prove (103) for $m = 1$, using the results of theorem 2.11.

Exercise. Show by a scaling argument that if the inequality (103) holds true for $1/p = 1/q + m/n < 0$ then the homogeneous inequality (100) is also true.

Proof [Proof of (100)]: We obtain the cases with $m > 1$ by repeated iterations of the case $m = 1$. Hence, we can assume $m = 1$ and, by (101),

$$1 \le p < n, \quad \frac{n}{n-1} \le q = \frac{np}{n-p} < \infty.$$

Once we have the estimate for $p = 1$ and $q = n/(n-1)$, then we get the cases with $p > 1$ and $q > n/(n-1)$ by simply applying Hölder inequality. Indeed, let $q = \lambda n/(n-1)$, for some $\lambda > 1$, then

$$\|f\|_{L^q}^\lambda = \||f|^\lambda\|_{L^{\frac{n}{n-1}}} \lesssim \||f|^{\lambda-1}\partial f\|_{L^1} \le \||f|^{\lambda-1}\|_{L^{p'}} \|\partial f\|_{L^p},$$

and we just have to check that

$$(\lambda - 1)p' = \frac{\frac{n-1}{n}q - 1}{1 - \frac{1}{n} - \frac{1}{q}} = q.$$

But this essentially needs no verification – by the scaling of the inequality, the exponents must work out.

It only remains to prove the special case $m = 1$, $p = 1$, $q = n/(n-1)$. Following Nirenberg, [**?**], one can show the stronger result that for $f \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ we have

$$\|f\|_{L^{\frac{n}{n-1}}(\mathbb{R}^n)} \lesssim \prod_{j=1}^n \|\partial_j f\|_{L^1(\mathbb{R}^n)}^{1/n}. \tag{104}$$

When $n = 1$, this comes easily from writing

$$f(x) = \int_{-\infty}^x f'(y)\mathrm{d}y.$$

When $n = 2$, we do the same with respect toeach variable and then multiply and integrate:

$$\begin{aligned} \iint |f(x_1,x_2)|^2 \mathrm{d}x_1\mathrm{d}x_2 &\le \iint \int |\partial_1 f(y_1,x_2)|\mathrm{d}y_1 \int |\partial_2 f(x_1,y_2)|\mathrm{d}y_2 \mathrm{d}x_1\mathrm{d}x_2 \\ &= \|\partial_1 f\|_{L^1} \|\partial_2 f\|_{L^1}. \end{aligned}$$

When $n \ge 3$ things become more tricky and, to separate the variables, we have to make a repeated use of Hölder inequality. Let just look at the case $n = 3$. To ease the notation set $f_j = \partial_j f$ and $\int \phi(x)\mathrm{d}x_j = \int_j \phi(\hat{x}_j)$. We start with

$$|f(x)|^{\frac{3}{2}} \leq \left(\int_1 |f_1(\cdot, x_2, x_3)|\right)^{\frac{1}{2}} \left(\int_2 |f_2(x_1, \cdot, x_3)|\right)^{\frac{1}{2}} \left(\int_3 |f_3(x_1, x_2, \cdot)|\right)^{\frac{1}{2}}.$$

Then integrate with respect to x_1. The first factor on the right hand side doesn't depend on x_1, while we use Hölder to separate the second from the third,

$$\int_1 |f(\cdot, x_2, x_3)|^{\frac{3}{2}} \leq \left(\int_1 |f_1(\cdot, x_2, x_3)|\right)^{\frac{1}{2}} \left(\int_{1,2} |f_2(\cdot, \cdot, x_3)|\right)^{\frac{1}{2}} \left(\int_{1,3} |f_3(\cdot, x_2, \cdot)|\right)^{\frac{1}{2}}.$$

Proceed similarly with the integration with respect tox_2,

$$\int_{1,2} |f(\cdot, \cdot, x_3)|^{\frac{3}{2}} \leq \left(\int_{1,2} |f_1(\cdot, \cdot, x_3)|\right)^{\frac{1}{2}} \left(\int_{1,2} |f_2(\cdot, \cdot, x_3)|\right)^{\frac{1}{2}} \left(\int_{1,2,3} |f_3(\cdot)|\right)^{\frac{1}{2}},$$

and finally do the same with x_3,

$$\int_{1,2,3} |f(\cdot)|^{\frac{3}{2}} \leq \left(\int_{1,2,3} |f_1(\cdot)|\right)^{\frac{1}{2}} \left(\int_{1,2,3} |f_2(\cdot)|\right)^{\frac{1}{2}} \left(\int_{1,2,3} |f_3(\cdot)|\right)^{\frac{1}{2}}.$$

When $n > 3$ the procedure is exacly the same. ■

Proof [Proof of (102)]: It clearly suffices to look at the case $m = 1$, since the cases $m > 1$ will follow from it applying (100). Assume thus $m = 1$ and $p > n$, we want to prove that

$$|f(0)| \lesssim \|f\|_{L^p} + \|Df\|_{L^p}.$$

Suppose first that f has support contained in the unit ball $B = \{|x| < 1\}$, then

$$f(0) = -\int_0^1 \frac{d}{dr} f(r\omega)\mathrm{d}r, \quad \omega \in \mathbb{S}^{n-1}. \tag{105}$$

Integrate with respect toω and then apply Hölder,

$$|f(0)| \lesssim \int_B \frac{|\partial f(x)|}{|x|^{n-1}}\mathrm{d}x \lesssim \|\partial f\|_{L^p} \left(\int_B \frac{\mathrm{d}x}{|x|^{(n-1)p'}}\right)^{1/p'} \lesssim \|\partial f\|_{L^p}, \tag{106}$$

where the integrability condition needed here is $(n-1)p' < n$, which is precisely $p > n$.

In general, fix a cutoff function $\phi \in C_0^\infty$ with support in B and $\phi(0) = 1$, then in view of the above, $|f(0)| = |\phi(0)f(0)| \lesssim \|\partial(\phi f)\|_{L^p} \lesssim \|f\|_{L^p} + \|\partial f\|_{L^p}$. ■

2.12. Classical Sobolev spaces. The Sobolev inequalities of theorem (2.11) lead us to the introduction of Sobolev spaces.

Definition 2.13. Let Ω be an open subset of $\mathbb{R}^n$. Fix $1 \leq p \leq \infty$ and let $s \in \mathbb{N}$ be a non-negative integer. The space $W^{s,p}(\Omega)$ consists of all locally integrable, real (or complex) valued functions u on Ω such that for all multiindex α with $|\alpha| \leq s$

the weak[4] derivatives $\partial^\alpha u$ belong to $L^p(\Omega)$. These spaces come equiped with the norms,

$$\|u\|_{W^{s,p}(\Omega)} = \Big(\sum_{|\alpha|\le s} \|\partial^\alpha u\|^p_{L^p(\Omega)}\Big)^{1/p}, \qquad \text{for} \quad 1 \le p < \infty$$

$$\|u\|_{W^{s,\infty}(\Omega)} = \sum_{|\alpha|\le s} \|\partial^\alpha u\|_{L^\infty(\Omega)}$$

We also denote by $W_0^{k,p}(\Omega)$ the closure of $\mathcal{C}_0^\infty(\Omega)$ in $W^{k,p}(\Omega)$.

In the particular case $p = 2$ we write $H^s(\Omega) = W^{s,2}(\Omega)$. Clearly $H^0(\Omega) = L^2(\Omega)$. We also write $H_0^s(\Omega) = W_0^{s,2}(\Omega)$. These spaces are especially important because of their Hilbert space structure.

In the particular case $p = \infty$ we work with the smaller space $C^s(\overline{\Omega}) \subset W^{s,\infty}(\Omega)$, the set of functions which are s times continuously differentiable and have bounded $\| \ \|_{W^{s,\infty}}$ norm.

Exercise. Show that for each $s \in \mathbb{N}$ and $1 \le p \le \infty$ the spaces $W^{s,p}(\Omega)$ are Banach spaces.

There is a lot more to be said about Sobolev spaces in domains $\Omega \subset \mathbb{R}^n$. For instance, We refer the reader to Evans ([**E**], ch. 5). For the time being we specialize to the case $\Omega = \mathbb{R}^n$.

Exercise. Show that the spaces $W^{k,p}(\mathbb{R}^n)$ and $W_0^{k,p}(\mathbb{R}^n)$ coincide. That means that $\mathcal{C}_0^\infty$ is dense in $W^{k,p}(\mathbb{R}^n)$.

The Sobolev inequalities proved in the previous subsection can be interpreted as embedding theorems. Indeed (100) and (103) can be interpreted as saying that the Sobolev space $W^{m,p}(\mathbb{R}^n)$ is included in the Lebesgue space $L^q(\mathbb{R}^n)$ as long as $\frac{1}{p} - \frac{m}{n} \le \frac{1}{q}$.

Proposition 2.14. *The following inclusions are continuous*

$$W^{m,p}(\mathbb{R}^n) \subset L^q(\mathbb{R}^n), \qquad \text{if} \quad \frac{1}{p} - \frac{m}{n} \le \frac{1}{q}.$$

Moreover, for $q = \infty$, $W^{m,p}(\mathbb{R}^n)$ embeds into the space of bounded continuous functions on $\mathbb{R}^n$ provided that $m > n/p$.

Proof : Follows from theorem 2.11 and the density of $\mathcal{C}_0^\infty(\mathbb{R}^n)$ in $W^{m,p}(\mathbb{R}^n)$. ∎

2.15. Hölder spaces. Together with Sobolev spaces Hölder spaces play a very important role in Analysis, especially in connection to elliptic equations. Before introducing these spaces we recall the definitions of the spaces $C^m(\overline{\Omega})$ of m times

[4]That is derivatives in the sense of distributions.

continuously differentiable functions $u : \Omega \to \mathbb{R}$ on an open domain Ω for which the $W^{m,\infty}$ norm is bounded,

$$\|u\|_{C^m(\overline{\Omega})} = \sum_{|\alpha| \leq m} \|\partial^\alpha u(x)\|_{L^\infty(\Omega)} < \infty.$$

Definition 2.16. Let Ω be an open domain in $\mathbb{R}^n$. We say that a function $u : \Omega \to \mathbb{R}$ is Hölder continuous with exponent $0 < \gamma \leq 1$ if,

$$[u]_{C^{0,\gamma}(\overline{\Omega})} = \sup_{x \neq y \in \Omega} \frac{|u(x) - u(y)|}{|x - y|^\gamma} < \infty. \tag{107}$$

The Hölder space $C^{k,\gamma}(\overline{\Omega})$ consists of all functions $u \in C^k(\overline{\Omega})$ for which the norm,

$$\|u\|_{C^{k,\gamma}(\overline{\Omega})} = \|u\|_{C^k(\overline{\Omega})} + \sum_{|\alpha|=k} [\partial^\alpha u]_{C^{0,\gamma}(\overline{\Omega})}. \tag{108}$$

is finite.

Exercise 2.17. The space $C^{k,\gamma}(\overline{\Omega})$ is a Banach space.

Exercise 2.18. Show that $C^{0,1}((a,b))$, the space of Lipschitz functions on an interval, consists exactly of those distributions whose derivative belongs to L^∞.

Exercise 2.19. Let $f(x) = (a \leq x \leq b)$ be the characteristic function of an interval. Show that the seven-fold convolution $f * \cdots * f$ is in the Hölder class $C^{5,1}$.

The following stronger version of the Sobolev embedding in L^∞ is important in elliptic theory. As usual, the relationship between the exponents involved can be deduced from dimensional analysis.

Theorem 2.20 (Morrey's inequality)**.** *Assume $n < p \leq \infty$. Then, for all $u \in C_0^\infty(\mathbb{R}^n)$,*

$$\|u\|_{C^{0,\gamma}(\mathbb{R}^n)} \lesssim \|u\|_{W^{1,p}(\mathbb{R}^n)} \tag{109}$$

provided that $\gamma = 1 - n/p$.

Proof : In one dimension, this is an easy application of the fundamental theorem of calculus and Hölder's inequality. For the general case, see [**E**], section 5.6.2. ∎

2.21. Fractional H^s- Sobolev spaces. Consider the Sobolev space

$$H^s(\mathbb{R}^n) = \{u \in L^2 : \partial^\alpha u \in L^2, \quad \forall\, |\alpha| \leq s\}.$$

Proposition 2.22. *The Sobolev space $H^s(\mathbb{R}^n)$ coincides with the set of all tempered distributions $u \in \mathcal{S}'(\mathbb{R}^N)$ for which $\hat{u}$ is locally integrable and,*

$$\|u\|_{H^s}^2 = \int_{\mathbb{R}^N} (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 < \infty \tag{110}$$

Proof : Follows easily from the Parseval identity, the density of $\mathcal{C}_0^\infty$ in each space, and the fact that $\widehat{\partial_j f}(\xi) = i\xi_j \hat{f}(\xi)$. ■

Observe that the equivalent definition of proposition 2.22 makes sense not only for positive integers but for all real numbers s. We can thus talk about Sobolev spaces H^s for all real values of s. We shall also make use of the following homogeneous Sobolev norm, for all $s \geq 0$,

$$\|u\|_{\dot{H}^s}^2 = \int_{\mathbb{R}^N} |\xi|^{2s} |\hat{u}(\xi)|^2 < \infty \tag{111}$$

Question. Why does $\|u\|_{\dot{H}^s}$ have units $L^{\frac{N}{2}-s}$ if we consider the physical space variable to have the unit L?

Exercise. For $s \in (0,1)$ the space $H^s(\mathbb{R}^n)$ coincides with the space of locally integrable functions such that,

$$\left(\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{|u(x) - u(x+y)|^2}{|y|^{n+2s}} dxdy + \|u\|_{L^2(\mathbb{R}^n)}^2 \right)^{1/2} < \infty \tag{112}$$

Exercise. Prove that, for $s > n/2$ the Sobolev space $H^s(\mathbb{R}^n)$ embeds in the space of bounded continuous functions.

2.23. A Trace Theorem. In order to make sense of boundary values of generalized functions for partial differential equations, it is important to prove that the operation of restriction, which obviously makes sense for continuous functions, continues to make sense even when the function is not continuous. Such theorems are called trace theorems. Consider, for simplicity, the case of the hyperplane $x_n = 0$ in $\mathbb{R}^n$ and define the trace operator,

$$Tf(x_1, \ldots, x_n) = f(x_1, \ldots, x_{n-1}, 0). \tag{113}$$

Clearly the operator makes sense for any continuous functions f, in particular for any test function, in $\mathbb{R}^n$.

Theorem 2.24. *The following estimate holds true, uniformly for any test function $f \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, $n \geq 2$ and any $s > 1/2$.*

$$\|Tf\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})} \lesssim \|f\|_{H^s(\mathbb{R}^n)} \tag{114}$$

Therefore T extends as a linear map $T : H^s(\mathbb{R}^n) \to H^{s-1/2}(\mathbb{R}^{n-1})$.

Remark. Observe that the result is dimensionally sharp, which is somewhat surprising if one compares it to the usual embedding of the Sobolev spaces $H^s(\mathbb{R})$ in $L^\infty(\mathbb{R})$, in which case we know that the sharp case, $s = 1/2$, is false. In fact the above trace theorem is also false for the case $s = 1/2$.

Proof : Take f smooth and $g(x') = f(x', 0)$. Let $\tilde{f}$ be the Fourier transform of f in x_n only, and $\hat{f}$, $\hat{g}$ be the Fourier transforms of f and g in $\mathbb{R}^n$ and $\mathbb{R}^{n-1}$, respectively.

I.e.

$$\tilde{f}(x', \xi_n) = \int_{-\infty}^{\infty} f(x', x_n) e^{-ix_n\xi_n} dx_n$$

By applying Fourier inversion (with $x_n = 0$) and then the Fourier transform, we get

$$g(x') = f(x', 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(x', \xi_m) d\xi_n$$
$$\hat{g}(\xi') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi', \xi_m) d\xi_n$$

We can then see, using our knowledge of fractional H^s spaces and Cauchy-Schwartz:

$$\begin{aligned}
\|g\|_{H^{s-1/2}(\mathbb{R}^{n-1}} &\lesssim \int_{\mathbb{R}^{n-1}} |\hat{g}(\xi')|^2 (1 + |\xi'|^2)^{s-1/2} d\xi' \\
&\lesssim \int_{\mathbb{R}^{n-1}} \left| \int_{-\infty}^{\infty} \hat{f}(\xi) d\xi_n \right|^2 (1 + |\xi'|^2)^{s-1/2} d\xi' \\
&\lesssim \int_{\mathbb{R}^{n-1}} (1 + |\xi'|^2)^{s-1/2} \left(\int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 (1 + |\xi|^2)^s d\xi_n \right) J(\xi') d\xi'
\end{aligned}$$

with,

$$J(\xi') = \int_{-\infty}^{\infty} (1 + |\xi|^2)^{-s} d\xi_n$$

And since $s > 1/2$, we have

$$\begin{aligned}
J(\xi') &= \int_{-\infty}^{\infty} (1 + |\xi|^2)^{-s} d\xi_n = \int_{-\infty}^{\infty} (1 + |\xi'|^2 + |\xi_n|^2)^{-s} d\xi_n \\
&= (1 + |\xi'|^2)^{-s+1/2} \int_{-\infty}^{\infty} (1 + y^2)^{-s} dy
\end{aligned}$$

Plugging this into our above estimate for $\|g\|_{H^{s-1/2}}$ proves the result. ■

Similar results hold for traces to higher co-dimension hypersurfaces. Here is such a result, which can be proved by elementary means.

Proposition 2.25. *Consider the trace operator T in $\mathbb{R}^3$ which takes continuous functions $f(t, x^1, x^2)$ to $Tf(t) = f(t, 0, 0)$. We have, for any test function f,*

$$\|\partial_t(Tf)\|_{L^2(\mathbb{R})} \lesssim \|\partial^2 f\|_{L^2(\mathbb{R}^3)} \tag{115}$$

Thus, T extends, as a bounded linear operator, to $H^2(\mathbb{R}^3)$ with values in $H^1(\mathbb{R})$.

Proof

$$\begin{aligned}
\int_{\mathbb{R}} |\partial_t f(t, 0, 0)|^2 dt &= \int_0^{\infty} dx^1 \int_0^{\infty} dx^2 \big(\partial_1 \partial_2 \int_{\mathbb{R}} \partial_t f(t, x) \partial_t f(t, x) dt \big) \\
&= 2 \int_0^{\infty} dx^1 \int_0^{\infty} dx^2 \big(\int_{\mathbb{R}} \partial_1 \partial_2 \partial_t f(t, x) \partial_t f(t, x) dt \big) \\
&+ 2 \int_0^{\infty} dx^1 \int_0^{\infty} dx^2 \big(\int_{\mathbb{R}} \partial_1 \partial_t f(t, x) \partial_2 \partial_t f(t, x) dt \big)
\end{aligned}$$

Clearly,

$$\Big|\int_0^\infty dx^1\int_0^\infty dx^2\big(\int_{-\infty}\partial_1\partial_t f(t,x)\partial_2\partial_t f(t,x)dt\big)\Big|\lesssim\|\partial^2 f\|^2_{L^2(\mathbb{R}^3)}$$

On the other hand, integrating by parts in t,

$$\int_{\mathbb{R}}\partial_1\partial_2\partial_t f(t,x)\partial_t f(t,x)dt=-\int_{\mathbb{R}}\partial_1\partial_2 f(t,x)\partial_t^2 f(t,x)dt$$

Hence,

$$\Big|\int_0^\infty dx^1\int_0^\infty dx^2\big(\int_{-\infty}\partial_1\partial_2\partial_t f(t,x)\partial_t f(t,x)dt\big)\Big|\quad\lesssim\quad\|\partial^2 f\|_{L^2(\mathbb{R}^3)}.$$

as desired. ∎

Exercise. Prove the same result using Fourier transform and extend it to all dimensions and general H^s spaces. **Exercise** Extend the result to bounded intervals in t.

2.26. Extensions. To extend results which hold true for functions in $\mathbb{R}^n$ to domains in $\mathbb{R}^n$ we need to extend the functions in a controlled manner. I will restrict the discussion to the case of the half space $\mathbb{R}^n_+=\{x\in\mathbb{R}^n/x^n\ge 0\}$. Consider the Sobolev space $W^{1,p}(\mathbb{R}^n_+)$. We want to prove the following.

Proposition 2.27. *There exists a bounded linear operator $E:W^{1,p}(\mathbb{R}^n_+)\to W^{1,p}(\mathbb{R}^n)$, such that for any continuous f,*

$$Ef|_{\mathbb{R}^n_+}=f$$

and,

$$\|Ef\|_{W^{1,p}(\mathbb{R}^n)}\lesssim\|f\|_{W^{1,p}(\mathbb{R}^n_+)}$$

Proof It suffices to prove the result for functions $f\in C^1(\mathbb{R}^n_+)$. Given such a function we define, using its higher order reflection, its extension $barf$ which coincided with f in $\mathbb{R}^n_+$ and, for all $x_n<0$,

$$\bar f(x',x^n)=-3f(x',-x^n)+4f(x',-\frac{1}{2}x^n)$$

Observe first that $\bar f$ is also C^1. Indeed $\bar f$ is continuous across $x^n=0$ and so are its derivatives with respect to the variables $x'=(x^1,\dots,x^{n-1})$. On the other hand, for $x^n<0$,

$$\partial_n\bar f(x',x^n)=3\partial_n f(x',-x^n)-2\partial_n f(x',-\frac{1}{2}x^n)$$

Hence, letting x^n tend to zero with $x^n<0$

$$(\partial_n\bar f)^-(x',0)=\partial_n f(x',0).$$

Using these calculations we immediately derive the desired estimate.

∎

Exercise Extend the result to the $W^{s,p}$ spaces, with $s \in \mathbb{N}$. What about fractional H^s spaces ?.

3. Littlewood-Paley theory

In its simplest manifestation Littlewood-Paley theory is a systematic and very useful method to understand various properties of functions f, defined on $\mathbb{R}^n$, by decomposing them in infinite dyadic sums $f = \sum_{k\in\mathbb{Z}} f_k$, with frequency localized components f_k, i.e. $\widehat{f_k}(\xi) = 0$ for all values of ξ outside the dyadic annulus $2^{k-1} \le |\xi| \le 2^{k+1}$. Such a decomposition can be easily achieved by choosing a test function $\chi(\xi)$ in Fourier space, supported in $\frac{1}{2} \le |\xi| \le 2$, and such that, for all $\xi \neq 0$,

$$\sum_{k\in\mathbb{Z}} \chi(2^{-k}\xi) = 1. \tag{116}$$

Indeed choose $\phi(\xi)$ to be a real radial bump function supported in $|\xi| \le 2$ which equals 1 on the ball $|\xi| \le 1$. Then the function $\chi(\xi) = \phi(\xi) - \phi(2\xi)$ verifies the desired properties.

We now define

$$\widehat{P_k f}(\xi) = \chi(\xi/2^k)\hat{f}(\xi) \tag{117}$$

or, in physical space,

$$P_k f = f_k = m_k * f \tag{118}$$

where $m_k(x) = 2^{nk} m(2^k x)$ and $m(x)$ the inverse Fourier transform of χ. Clearly, from (116)

$$f = \sum_{k\in\mathbb{Z}} P_k f \tag{119}$$

as desired. Denoting the map "multiplication by 2^{-k}" by 2^{-k}, we also have the important scaling identity

$$(P_k f) \circ 2^{-k} = P_0(f \circ 2^{-k}) \tag{120}$$

Observe that the Fourier transform of $P_k f$ is supported in the dyadic interval $2^{k-1} \le |\xi| \le 2^{k+1}$ and therefore,

$$P_{k'} P_k f = 0, \qquad \forall\, k, k' \in \mathbb{Z}, \quad |k - k'| > 2.$$

Therefore,

$$P_k f = \sum_{k'\in\mathbb{Z}} P_{k'}(P_k f) = \sum_{|k-k'|\le 1} P_{k'} P_k f$$

Thus, since P_{k-1}, P_k, P_{k+1} do not differ much between themselves we can write $P_k = \sum_{|k-k'|\le 1} P_{k'} P_k \approx P_k^2$. It is for this reason that the cut-off operators P_k are called, improperly, LP projections.

Denote $P_J = \sum_{k\in J} P_k$ for all intervals $J \subset \mathbb{Z}$. We write, in particular, $P_{\le k} = P_{(-\infty,k]}$ and $P_{<k} = P_{\le k-1}$. Clearly, $P_k = P_{\le k} - P_{<k}$.

The following properties of these LP projections lie at the heart of the classical LP theory:

Theorem 3.1. *The LP projections verify the following properties:*

LP 1. *Almost Orthogonality. The operators P_k are selfadjoint and verify $P_{k_1}P_{k_2} = 0$ for all pairs of integers such that $|k_1 - k_2| \ge 2$. In particular,*

$$\|f\|_{L^2}^2 \approx \sum_k \|P_k f\|_{L^2}^2 \tag{121}$$

LP 2. *L^p-boundedness: For any $1 \le p \le \infty$, and any interval $J \subset \mathbb{Z}$,*

$$\|P_J f\|_{L^p} \lesssim \|f\|_{L^p} \tag{122}$$

LP 3. *Finite band property. We can write any partial derivative $\partial P_k f$ in the form $\partial P_k f = 2^k \tilde{P}_k f$ and the symbol of $\tilde{P}_k$ is a cut-off operator*[5] *which verifies property LP2. In particular, for any $1 \le p \le \infty$*

$$\|\partial P_k f\|_{L^p} \lesssim 2^k \|f\|_{L^p} \tag{123}$$

$$2^k \|P_k f\|_{L^p} \lesssim \|\partial f\|_{L^p} \tag{124}$$

LP 4. *Bernstein inequalities. For any $1 \le p \le q \le \infty$ we have the Bernstein inequalities,*

$$\|P_k f\|_{L^q} \lesssim 2^{kn(1/p-1/q)} \|f\|_{L^p}, \qquad \forall\, k \in \mathbb{Z} \tag{125}$$

$$\|P_{\le 0} f\|_{L^q} \lesssim \|f\|_{L^p}. \tag{126}$$

In particular,

$$\|P_k f\|_{L^\infty} \lesssim 2^{kn/p} \|f\|_{L^p}.$$

LP5. *Commutator estimates Consider the commutator*

$$[P_k\,,\, f] \cdot g = P_k(f \cdot g) - f \cdot P_k g$$

with $f, g \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. We have,

$$\|\, [P_k\,,\, f] \cdot g\|_{L^p} \lesssim 2^{-k} \|\nabla f\|_{L^\infty} \|g\|_{L^p}.$$

LP6. *Square function inequalities. Let $\mathbf{S}f$ be the vector valued function $\mathbf{S}f = (P_k f)_{k \in \mathbb{Z}}$. The quantity*

$$Sf(x) = |\mathbf{S}f(x)| = \big(\sum_{k \in \mathbb{Z}} |P_k f(x)|^2 \big)^{1/2} \tag{127}$$

is known as the Littlewood-Paley square function. For every $1 < p < \infty$ there exists constant(s), depending on p, such that for all $f \in \mathcal{C}_0^\infty$

$$\|f\|_{L^p} \lesssim \|Sf\|_{L^p} \lesssim \|f\|_{L^p} \tag{128}$$

[5] Associated with a slightly different test function $\tilde{\chi}$ which remains supported in $\frac{1}{2} \le |\xi| \le 2$, but may fail to satisfy (116).

Proof : Only the proof of LP6 is not straightforward and we postpone it until next section. The proof of LP1 is immediate. Indeed we only have to check (121). Clearly,

$$\begin{aligned}\|f\|_{L^2}^2 &= \|\sum_k P_k f\|_{L^2}^2 = \sum_{|k-k'|\le 1} < P_k f, P_{k'} f >_{L^2}\\ &\le \sum_{|k-k'|\le 1} \|P_k f\|_{L^2}\|P_{k'} f\|_{L^2}\\ &\lesssim \sum_k \|P_k f\|_{L^2}^2\end{aligned}$$

To show that $\sum_k \|P_k f\|_{L^2}^2 \lesssim \|f\|_{L^2}^2$ we only need to use Parseval's identity together with the definition of the projections P_k.

It suffices to prove LP2 for intervals of the form $J = (-\infty, k] \subset \mathbb{Z}$, that is to prove L^p boundedness for $P_{\le k}$. If $\chi(\xi) = \phi(\xi) - \phi(2\xi)$ then $\widehat{P_{\le k} f} = \phi(\xi/2^k)\hat{f}(\xi)$. Thus

$$P_{\le k} f = \bar{m}_k * f,$$

where $\bar{m}_k(x) = 2^{nk}\bar{m}(2^k x)$ and $\bar{m}(x)$ is the inverse Fourier transform of ϕ. Observe that $\|\bar{m}_k\|_{L^1} = \|\bar{m}\|_{L^1} \lesssim 1$. Thus, using the convolution inequality (90),

$$\|P_{\le k} f\|_{L^p} \le \|\bar{m}_k\|_{L^1}\|f\|_{L^p} \lesssim \|f\|_{L^p}$$

as desired.

To prove LP3 we write $\partial_i(P_k f) = 2^k(\partial_i m)_k * f$ where $(\partial_i m)_k(x) = 2^{nk}\partial_i m(2^k x)$. Clearly $\|(\partial_i m)_k\|_{L^1} = \|\partial_i m\|_{L^1} \lesssim 1$. Hence,

$$\|\partial_i(P_k f)\|_{L^p} \lesssim 2^k\|f\|_{L^p}$$

which establishes (123). To prove (124) we write $\hat{f}(\xi) = \sum_{j=1}^n \frac{\xi_j}{i|\xi|^2}\widehat{\partial_{x_j} f}(\xi)$. Hence,

$$2^k\widehat{P_k f}(\xi) = \sum_{j=1}^n 2^k \frac{\xi_j}{i|\xi|^2}\chi(\xi/2^k)\widehat{\partial_{x_j} f}(\xi) = \sum_{j=1}^n 2^k \psi_j(\xi/2^k)\widehat{\partial_{x_j} f}(\xi)$$

where $\psi_j(\xi) = \frac{\xi_j}{i|\xi|^2}\chi(\xi)$. Hence, in physical space,

$$2^k P_k f = \sum_{j=1}^n (\overline{{}^j m})_k * \partial_j f$$

with $(\overline{{}^j m})_k(x) = 2^{nk}\cdot \overline{{}^j m}(2^k x)$ and $\overline{{}^j m}$ the inverse Fourier transform of ψ_j. Thus, as before,

$$2^k\|P_k f\|_{L^p} \lesssim \sum_{j=1}^n \|\partial_j f\|_{L^p} = \|\partial f\|_{L^p}$$

as desired.

Property LP4 is an immediate consequence of the physical space representation (118) and the convolution inequality (90).

$$\|P_k f\|_{L^q} = \|m_k * f\|_{L^q} \lesssim \|m_k\|_{L^r}\|f\|_{L^p}$$

where $1+q^{-1}=r^{-1}+p^{-1}$. Now,

$$\|m_k\|_{L^r}=2^{nk}\big(\int_{\mathbb{R}^n}|m(2^kx)|^rdx\big)^{1/r}=2^{nk}2^{-nk/r}\|m\|_{L^r}\lesssim 2^{nk(1-1/r)}\lesssim 2^{nk(1/p-1/q)}$$

It only remains to prove LP5. In view of (118) we can write,

$$P_k(fg)(x)-f(x)P_kg(x) \quad = \quad \int_{\mathbb{R}^n} m_k(x-y)\big(f(y)-f(x)\big)g(y)dy$$

On the other hand,

$$\begin{aligned} |f(y)-f(x)| &\lesssim \big|\int_0^1 \frac{d}{ds}f(x+s(y-x))ds\big| \\ &\lesssim |x-y|\|\partial f\|_{L^\infty} \end{aligned}$$

Hence,

$$|P_k(fg)(x)-f(x)P_kg(x)| \quad \lesssim \quad 2^{-k}\|\partial f\|_{L^\infty}\int_{\mathbb{R}^n}|\bar{m}_k(x-y)||g(y)|dy$$

where $\bar{m}_k(x)=2^{nk}\bar{m}(2^kx)$ and $\bar{m}(x)=|x|m(x)$. Thus,

$$\|P_k(fg)-fP_kg\|_{L^p}\lesssim 2^{-k}\|\partial f\|_{L^\infty}\|g\|_{L^p}$$

We leave the proof of property LP6 for the next section. ■

Remark. It could have simplified matters in the preceding proof to prove properties **LP2-4** only in the case $k=0$, and deduce the more general estimates from the scaling identity (120). In particular, note that the Bernstein inequality is simply the statement that lower L^p norms control higher L^p norms when f is localized in frequency space (as opposed to the other way around, which occurs when f is localized in physical space). This accords with our intuition for L^p norms: while a frequency localized function may be too large at ∞ in physical space to be integrable, one need not worry about sudden jumps or spikes where the function blows up locally, and hence only the former phenomenon needs to be controlled.

Definition. We say that a Fourier multiplier operator $\tilde{P}_k$ is similar to a standard LP projection P_k if its symbol $\tilde{\chi}_k$ is a bump function adapted to the dyadic region $|\xi|\sim 2^k$. More precisely we can write $\tilde{\chi}_k(\xi)=\tilde{\chi}(\frac{\xi}{2^k})$ for some bump function $\tilde{\chi}$ supported in the region $c^{-1}2^k\lesssim|\xi|\le c2^k$ for some fixed $c>0$.

Remark. Observe that the inequality $\|P_kf\|_{L^p}\lesssim\|f\|_{L^p}$ holds for every other operator $\tilde{P}_k$ similar to P_k. The same holds true for the properties LP3, LP4 and LP5.

Remark: We have the following pointwise relation of the operator $\tilde{P}_k$ with the maximal function:

$$|\tilde{P}_{\le k}f|\lesssim\mathcal{M}f(x) \tag{129}$$

Indeed we have, as before,

$$\tilde{P}_{\le k}f=\tilde{m}_k*f,$$

where $\tilde{m}_k(x) = 2^{nk}\tilde{m}(2^k x)$ and $\tilde{m}(x) \in \mathcal{S}(\mathbb{R}^n)$. Therefore,

$$\begin{aligned}
|\tilde{P}_{\le k} f| &\lesssim 2^{nk}\int |f(y)|\tilde{m}\big(2^k(x-y)\big)|dy \lesssim 2^{nk}\int |f(y)|(1+2^k|x-y|)^{-n-1}dy\\
&\lesssim 2^{nk}\int_{B(x,2^{-k})} |f(y)|(1+2^k|x-y|)^{-n-1}dy\\
&+ 2^{nk}\sum_{j=0}^{\infty}\int_{2^j\le 2^k|x-y|\le 2^{j+1}} |f(y)|(1+2^k|x-y|)^{-n-1}dy\\
&\lesssim 2^{nk}\Big(\int_{B(x,2^{-k})}|f(y)|dy + \sum_{j\ge 0} 2^{-(n+1)j}\int_{|x-y|\le 2^{j+1-k}}|f(y)|dy\Big)\\
&\lesssim \mathcal{M}f(x) + \sum_{j>0} 2^{-(n+1)j}2^{nk}2^{n(j+1-k)}\frac{1}{|B(x,2^{-k+j+1})|}\int_{B(x,2^{-k+j+1})}|f(y)|dy\\
&\lesssim \mathcal{M}f(x) + 2^n\sum_{j>0}2^{-j}\mathcal{M}f(x) \lesssim \mathcal{M}f(x)
\end{aligned}$$

as desired.

Properties LP3-LP4 go a long way to explain why LP theory is such a useful tool for partial differential equations. The finite band property allows us to replace derivatives of the dyadic components f_k by multiplication with 2^k. The $L^p \to L^\infty$ Bernstein inequality is a dyadic remedy for the failure of the embedding of the Sobolev space $W^{\frac{n}{p},p}(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$. Indeed, in view of the finite band property, the Bernstein inequality does actually imply the desired Sobolev inequality for each LP component f_k, the failure of the Sobolev inequality for f is due to the summation $f = \sum_k f_k$.

In what follows we give a few applications of LP -calculus.

3.2. Interpolation inequalities. The following inequality holds true for arbitrary functions in $\mathcal{C}_0^\infty(\mathbb{R}^n)$ and any integers $0 \le i \le m$:

$$\|\partial^i f\|_{L^p} \lesssim \|f\|_{L^p}^{1-i/m}\|\partial^m f\|_{L^p}^{i/m} \tag{130}$$

To prove it we decompose $f = P_{\le k}f + P_{>k}f = f_{\le k} + f_{>k}$. Now, using LP2-LP4, for any fixed value of $k \in \mathbb{Z}$,

$$\begin{aligned}
\|\partial^i f\|_{L^p} &\le \|\partial^i f_{\le k}\|_{L^p} + \|\partial^i f_{>k}\|_{L^p}\\
&\le 2^{ki}\|f\|_{L^p} + 2^{k(i-m)}\|\partial^m f\|_{L^p}
\end{aligned}$$

Thus,

$$\|\partial^i f\|_{L^p} \le \lambda^i\|f\|_{L^p} + \lambda^{i-m}\|\partial^m f\|_{L^p}$$

for any $\lambda \in 2^{\mathbb{Z}}$. To finish the proof we would like to choose λ such that the two terms on the right hand side are equal to each other, i.e.,

$$\lambda_0 = \left(\frac{\|\partial^m f\|_{L^p}}{\|f\|_{L^p}}\right)^{1/m}$$

since we are restricted to $\lambda \in 2^{\mathbb{Z}}$ we choose the dyadic number $\lambda \in 2^{\mathbb{Z}}$ such that, $\lambda \leq \lambda_0 \leq 2\lambda$ Hence,

$$\|\partial^i f\|_{L^p} \quad \leq \quad \lambda_0^i \|f\|_{L^p} + \big(\frac{2}{\lambda_0}\big)^{m-i} \|\partial^m f\|_{L^p} \lesssim \|f\|_{L^p}^{1-i/m} \|\partial^m f\|_{L^p}^{i/m}.$$

In general when an estimate for functions on $\mathbb{R}^n$ fails to be dimensionally consistent (in that the scalings of the two sides are not the same), such an estimate can be "amplified" into one which appears even stronger (or proven false).

Exercise. Assuming the inequality

$$\|\partial^i f\|_{L^p} \lesssim \|f\|_{L^p} + \|\partial^m f\|_{L^p}$$

deduce the estimate (130) by considering the rescalings $f \to f \circ \lambda$.

3.3. Non-sharp Sobolev inequalities. We shall prove the following slightly improved version of the inequality (103), for functions $f \in C_0^\infty(\mathbb{R}^n)$ and exponents $1 \leq p < q < \infty$ with $1/p - m/n < 1/q$,

$$\|f\|_{L^q(\mathbb{R}^n)} \lesssim \|f\|_{L^p} + \|\partial^m f\|_{L^p}$$

We decompose $f = P_{\leq 0} f + \sum_{k \in \mathbb{N}} P_k f = f_{<0} + \sum_{k>0} f_k$. Thus, using LP4 and then LP3,

$$\begin{aligned}
\|f\|_{L^q} \quad &\leq \quad \|f_{<0}\|_{L^q} + \sum_{k>0} \|f_k\|_{L^q} \lesssim \|f\|_{L^p} + \sum_{k>0} 2^{kn(1/p-1/q)} \|f\|_{L^p} \\
&\lesssim \quad \|f\|_{L^p} + \sum_{k>0} 2^{kn(m/n-\epsilon)} \|f\|_{L^p} \lesssim \|f\|_{L^p} + \sum_{k>0} 2^{-kn\epsilon} \|\partial^m f\|_{L^p} \\
&\lesssim \quad \|f\|_{L^p} + \|\partial^m f\|_{L^p}
\end{aligned}$$

3.4. Spaces of functions. The Littlewood-Paley theory can be used both to give alternative descriptions of Sobolev spaces and introduce new, more refined, spaces of functions. We first remark that, in view of the almost orthogonality property LP1,

$$\begin{aligned}
\|f\|_{L^2}^2 \quad &= \quad \|\sum_{k \in \mathbb{Z}} P_k f\|_{L^2}^2 \lesssim \sum_{k \in \mathbb{Z}} \|P_k f\|_{L^2}^2 \\
\sum_{k \in \mathbb{Z}} \|P_k f\|_{L^2}^2 \quad &\lesssim \quad \|f\|_{L^2}
\end{aligned}$$

We can thus give an LP description of the homogeneous Sobolev norms $\| \ \|_{\dot{H}^s(\mathbb{R}^n)}$

$$\|f\|_{\dot{H}^s}^2 \approx \sum_{k \in \mathbb{Z}} 2^{2ks} \|P_k f\|_{L^2}^2 \tag{131}$$

For $k \in \mathbb{Z}^+$, define operator $\Delta_k = P_k$ if $k > 0$, and $\Delta_0 = P_{\leq 0}$. Also for the H^s norms,

$$\|f\|_{H^s}^2 \approx \sum_{k=0}^{\infty} 2^{2ks} \|\Delta_k f\|_{L^2}^2 \tag{132}$$

The Littlewood- Paley decompositions can be used to define new spaces of functions such as Besov spaces.

Definition: The Besov space $B^s_{p,q}(\mathbb{R}^n)$ is the closure of $\mathcal{C}^\infty_0(\mathbb{R}^n)$ relative to the norm:

$$\|f\|_{B^s_{p,q}} = (\sum_{k=0}^{\infty} 2^{ksq}\|\Delta_k f\|^q_{L^p})^{\frac{1}{q}} \tag{133}$$

The corresponding homogeneous Besov norm is defined by,

$$\|f\|_{\dot{B}^s_{p,q}} = (\sum_{k\in\mathbb{Z}} 2^{sqk}\|P_k f\|^q_{L^p})^{\frac{1}{q}}, \tag{134}$$

One similarly define Triebel space $F^s_{p,q}$ by reversing the L^p norm and l^q norm in (133). Thus, for example, the H^s norm is equivalent with the Besov norm $B^s_{2,2}$. Observe that, $H^s \subset B^s_{2,1}$. One reason why the larger space $B^s_{2,1}$ is useful is because of the following

$$\|f\|_{L^\infty} \lesssim \|f\|_{\dot{B}^{n/2}_{2,1}} \tag{135}$$

which follows from the Bernstein inequality LP4. (135) will play a key role in the following section. Another reason to use the Besov norms $B^s_{2,1}$ will become transparent in the next section where we discuss product estimates.

3.5. Product estimates. The LP calculus is particularly useful for nonlinear estimates. Let f, g be two functions on $\mathbb{R}^n$. Consider,

$$P_k(fg) = \sum_{k',k''\in\mathbb{Z}} P_k(P_{k'}fP_{k''}g) \tag{136}$$

Now, since $P_{k'}f$ has Fourier support in the set $D' = 2^{k'-1} \le |\xi| \le 2^{k'+1}$ and $P_{k''}f$ has Fourier support in $D" = 2^{k''-1} \le |\xi| \le 2^{k''+1}$ it follows that $P_{k'}fP_{k''}g$ has Fourier support in $D' + D''$. We only get a nonzero contribution in the sum (136) if $D' + D''$ intersects $2^{k-1} \le |\xi| \le 2^{k+1}$. Therefore, writing $f_k = P_k f$ and $f_{<k} = P_{<k}f$, and $f_J = P_J f$ for any interval $J \subset \mathbb{Z}$ we derive,

Lemma 3.6. *Given functions f, g we have the following decomposition:*

$$\begin{aligned} P_k(f\cdot g) &= HH_k(f,\, g\,) + LL_k(f,\, g\,) + LH_k(f,\, g\,) + HL_k(f,\, g\,) \quad (137)\\ HH_k(f,\, g\,) &= \sum_{k',k''>k+5, |k'-k''|\le 3} P_k\big(f_{k'}\cdot P_{k''}g\big)\\ LL_k(f,\, g\,) &= P_k\big(f_{[k-5,k+5]}\cdot g_{[k-5,k+5]}\big)\\ LH_k(f,\, g\,) &= P_k\big(f_{\le k-5}\cdot g_{[k-3,k+3]}\big)\\ HL_k(f,\, g\,) &= P_k\big(f_{[k-3,k+3]}\cdot g_{\le k-5}\big) \end{aligned}$$

The term $HH_k(f,\, g\,)$ corresponds to high-high interactions. More precisely, each term in the sum defining $HH_k(f,\, g\,)$ has frequency $\sim 2^m$ for some $2^m >> 2^k$. We shall write schematically,

$$HH_k(f,\, g\,) = P_k\big(\sum_{m>k} f_m\cdot g_m\big) \tag{138}$$

The term $LL_k(f,\, g\,)$ consists of a finite number of terms which can be typically ignored. Indeed they can be treated, in any estimates, like either a finite number of HH terms or a finite number of LH and HL terms. We write, schematically,

$$LL_k(f,\, g\,) = 0 \tag{139}$$

Finally the LH_k and HL_k terms consist of low high, respectively high-low, interactions. We shall write schematically,

$$LH_k(f,\, g\,) \quad = \quad P_k\big(f_{<k}\cdot g_k\big) \tag{140}$$

$$HL_k(f,\, g\,) \quad = \quad P_k\big(f_k\cdot g_{<k}\big) \tag{141}$$

Remark. In the correct expression of LH_k given by (137) the terms of the form $f_{\le k-5}\cdot g_{k''}$, $k''\in[k-3,k+3]$, have Fourier supports in the dyadic region $\sim 2^k$. Thus P_k can be safely ignored and we can write,

$$LH_k(f,\, g\,)\sim f_{<k}\cdot g_k.$$

We have thus established, the famous *trichotomy formula*,

$$P_k(f\cdot g) = LH_k(f,g) + HL_k(f,g) + HH_k(f,g) \tag{142}$$

which is the basis of *paradifferential calculus.* In practice whenever we apply formula (142) we have to recall that formulas (139)–(141) are only appproximate; the correct definitions are given by (137). However in any estimates we can safely ignore the additional terms as they are estimated precisely in the same way as the terms we keep.

We shall now make use of the trichotomy formula to prove a product estimate.

Theorem 3.7. *The following estimate holds true for all $s>0$.*

$$\|fg\|_{H^s}\lesssim \|f\|_{L^\infty}\|g\|_{H^s} + \|g\|_{L^\infty}\|f\|_{H^s} \tag{143}$$

Thus for all $s>n/2$,

$$\|fg\|_{H^s}\lesssim \|f\|_{H^s}\|g\|_{H^s} \tag{144}$$

In what follows we give a somewhat simple proof of theorem (3.7) which is very instructive. The proof[6] shows that it is sometimes better not to rely on the full decomposition (137) but rather using decompositions sparingly whenever needed. Indeed, we write,

$$\|fg\|^2_{\dot H^s} \quad\lesssim\quad \sum_k 2^{2ks}\|P_k(fg)\|^2_{L^2}\lesssim \sum_k 2^{2ks}\|P_k(f_{<k}g)\|^2_{L^2} + \sum_k 2^{2ks}\|P_k(f_{\ge k}g)\|^2_{L^2}$$

Now,

$$\begin{aligned}\sum_k 2^{2ks}\|P_k(f_{\ge k}g)\|^2_{L^2} \quad&\lesssim\quad \|g\|^2_{L^\infty}\sum_k 2^{2ks}\ \|f_{\ge k}\|^2_{L^2}\\ &\lesssim\quad \|g\|^2_{L^\infty}\sum_k\sum_{k'\ge k} 2^{2(k-k')s}\|2^{k's}f_{k'}\|^2_{L^2}\\ &=\quad \|g\|^2_{L^\infty}\sum_{k'}\big(\sum_{k\le k'} 2^{2(k-k')s}\big)\|2^{k's}f_{k'}\|^2_{L^2}\\ &\lesssim\quad \|g\|^2_{L^\infty}\|f\|^2_{\dot H^s}\end{aligned}$$

[6] I thank Igor Rodnianski for pointing the argument to me.

To estimate $\sum_k 2^{2ks}\|P_k(f_{\le k}g)\|_{L^2}^2$ we shall decompose further, proceeding as in the decomposition (137). But first observe that the term $\sum_k 2^{2ks}\|P_k(f_{[k-3,k]}g)\|_{L^2}^2$ can be treated precisely as $\sum_k 2^{2ks}\|P_k(f_{>k}g)\|_{L^2}^2$. Indeed we might as well estimated $\sum_k 2^{2ks}\|P_k(f_{>k-3}g)\|_{L^2}^2$ instead. Now,

$$\begin{aligned} P_k(f_{\le k-3}g) &= \sum_{k'} P_k(f_{\le k-3}g_{k'}) = \sum_{k'<k-2} P_k(f_{\le k-3}g_{k'}) + \sum_{k-2\le k'\le k+2} P_k(f_{\le k-3}g_{k'}) \\ &+ \sum_{k'>k+2} P_k(f_{\le k-3}g_{k'}) \end{aligned}$$

Observe that the first and last term are zero, therefore,

$$P_k(f_{\le k-3}g) = \sum_{k-2\le k'\le k+2} P_k(f_{\le k-3}g_{k'}) \approx P_k(f_{\le k-3}g_k).$$

Often, for simplicity, we simply write,

$$P_k(f_{<k}g) \approx f_{<k}\cdot g_k \tag{145}$$

Of course this formula is not quite right, but is morally right. Now,

$$\begin{aligned} \sum_k 2^{2ks}\|P_k(f_{<k}g)\|_{L^2}^2 &= \sum_k 2^{2ks}\|f_{<k}g_k\|_{L^2}^2 \\ &\lesssim \|f\|_{L^\infty}^2 \sum_k 2^{2ks}\|g_k\|_{L^2}^2 = \|f\|_{L^\infty}^2\|g\|_{\dot H^s}^2 \end{aligned}$$

as desired.

Remark. In view of (145) we have the following partial decomposition formula,

$$P_k(fg) = f_{<k}g_k + P_k(f_{\ge k}g) = LH_k(f,g) + P_k(f_{\ge k}g) \tag{146}$$

Contrast this with the full trichotomy decomposition (142).

Similar estimates, easier to prove, hold in Besov spaces. Indeed, for every $s>0$ we have,

$$\|fg\|_{B^s_{2,1}} \lesssim \|f\|_{L^\infty}\|g\|_{B^s_{2,1}} + \|g\|_{L^\infty}\|f\|_{B^s_{2,1}} \tag{147}$$

Exercise. Prove estimate (147).

4. Wente's Inequality

In this section we prove Wente's inequality as an application of Littlewood-Paley theory. In what follows given two functions f, g in $\mathbb{R}^2$ we consider the bilinear expression $(df\wedge dg)^* = \partial_x f\partial_y g - \partial_y f\partial_x g$, where $*$ denotes the trivial Hodge duality in $\mathbb{R}^2$. By abuse of language we drop the dual sign below and write simply $df\wedge dg$.

Theorem 4.1. *On $\mathbb{R}^2$, assume $f,\ g\in H^1(\mathbb{R}^2)$, $\Delta u = (df\wedge dg)$. Then $u\in L^\infty$ is in fact continuous.*

Remark. In fact $df \wedge dg$ If $\wedge$ is replaced by ordinary multiplication, then the best we can get is $df \cdot dg \in L^1$. This is obviously not enough to obtain that $u \in L^\infty$. It turns out however that $df \wedge dg$ has special structure which allows us to derive the desired estimate. In the above theorem, we refer to the canonical solution $u = \Delta^{-1}(df \wedge dg)$ obtained through the canonical solution.

Proof : It is easy to see from finite band property that Δ is a isometric operator from $\dot{B}^s_{p,1}$ to $\dot{B}^{s-2}_{p,1}$. In fact we shall work with $p = 2$, In view of the Sobolev inequality (135), it suffices to show that $df \wedge dg \in \dot{B}^{-1}_{2,1}(\mathbb{R}^2)$. Using the trichotomy formula and the fact that the LP projections P_k commute with d we write,

$$\begin{aligned} I &= df \wedge dg = LH_k + HL_k + HH_k \\ LH_k &= dP_{<k} f \wedge dP_k g \\ HL_k &= dP_k \wedge dP_{<k} g \\ HH_k &= P_k (\sum_{m \geq k} (dP_m f \wedge dP_m g) \end{aligned}$$

By symmetry we only need to deal with LH and HH. The LH term is trivial to estimate, without using the special structure of the wedge product. Using the Bernstein inequality we write,

$$\begin{aligned} 2^{-k} \|LH_k\|_{L^2} &\lesssim 2^{-k} \sum_{l<k} \|dP_l f\|_{L^\infty} \|dP_k(g)\|_{L^2} \\ &\lesssim \sum_{l<k} 2^{l-k} \|DP_l f\|_{L^2} \|DP_k f\|_{L^2} \end{aligned}$$

The proof now follows with the following discrete version of the Young inequality.

Lemma 4.2. *Let $f(k) \in l^1(\mathbb{Z})$ and $g(k), h(k) \in l^2(\mathbb{Z})$. Then,*

$$\sum_{k,l} f(k-l) g(l) h(k) \leq \|f\|_{l^1} \|g\|_{L^2} \|h\|_{l^2}.$$

Using the lemma, we derive,

$$\begin{aligned} \sum_k 2^{-k} \|LH_k\|_{L^2} &\lesssim \big(\sum_l \|DP_l f\|^2_{L^2}\big)^{1/2} (\sum_k \|DP_k f\|^2_{L^2})^{1/2} \\ &\lesssim \|Df\|_{L^2} \|Dg\|_{L^2} \end{aligned}$$

We now consider HH_k. It is here that we need to use the special structure of the wedge product. In fact we shall simply use the identity, $df \wedge dg = d(f \wedge dg)$. Thus,

$$\begin{aligned} HH_k &= \sum_{m \geq k} P_k (dP_m f \wedge dP_m g) \\ &= \sum_{m \geq k} dP_k (P_m f \wedge dP_m g) \end{aligned}$$

Thus, using the finite band property and Bernstein inequality,

$$\begin{aligned} \|HH_k\|_{L^2} &\lesssim 2^{2k} \|P_m f \wedge dP_m g\|_{L^1} \\ &\lesssim 2^{2k} \|P_m f\|_{L^2} \|DP_m g\|_{L^2} \\ &\lesssim 2^{2k-m} \|DP_m f\|_{L^2} \|DP_m g\|_{L^2} \end{aligned}$$

Therefore,

$$2^{-k}\|HH_k\|_{L^2} \quad \lesssim \quad 2^{k-m}\|DP_m f\|_{L^2}\|DP_m g\|_{L^2}$$

Thus, again, using the discrete Young inequality of the lemma above,

$$\sum_k 2^{-k}\|LH_k\|_{L^2} \quad \lesssim \quad \|Df\|_{L^2}\|Dg\|_{L^2}$$

as desired. ■

5. A Sharp Trace Theorem

In this section, we provide another application of LP theory: a stronger version of the the Trace Theorem, in Besov spaces, see [**Kl-Rodn3**]

For simplicity, let $I = [0,1]$ and consider $I \times \mathbb{R}^2$. We will use the mixed norm notation:

$$\|f\|_{L^q_t L^p_x} = \left(\int_0^1 \|f(t,\cdot)\|^q_{L^p_x(\mathbb{R}^2)} dt\right)^{\frac{1}{q}}$$

$$\|f\|_{L^p_x L^q_t} = \left(\int_{\mathbb{R}^2} \|f(\cdot,x)\|^p_{L^q_t(I)} dx\right)^{\frac{1}{p}}$$

with the obvious modifications if $p = \infty$ or $q = \infty$.

We will get the following trace-like estimate:

$$\|\int_I |\partial_t f|^2 dt\|_{B^1_{2,1}} \lesssim \|f\|^2_{H^2(I\times\mathbb{R}^2)} \tag{148}$$

We observe that

$$\|g\|_{B^1_{2,1}} \lesssim \|\nabla g\|_{B^0_{2,1}} + \|g\|_{L^2}$$

Thus, (148) follows from the "sharp bilinear trace" theorem below.

Theorem 5.1. *For any smooth, scalar functions g,h on $I \times \mathbb{R}^2$, we have*

$$\|\int_I \partial_t g \cdot h dt\|_{B^0_{2,1}} \lesssim \|g\|_{H^1(I\times\mathbb{R}^2)} \cdot \|h\|_{H^1(I\times\mathbb{R}^2)} \tag{149}$$

Proof Immediately we see:

$$\begin{aligned}\|\int_I \partial_t g \cdot h dt\|_{B^0_{2,1}} \quad &= \quad \sum_{k\geq 0} \|P_k \int_0^1 \partial_t g \cdot h dt\|_{L^2_x} + \|P_{<0}\int_0^1 \partial_t g \cdot h dt\|_{L^2_x} \\ &\lesssim \quad \sum_{k\geq 0} \|P_k \int_0^1 \partial_t g \cdot h dt\|_{L^2_x}\end{aligned}$$

We will then decompose g and h with respect to x; $g = \sum_k P_k g = \sum_k g_k$, $h = \sum_k P_k h = \sum_k h_k$. Then we can decompose $P_k \int_0^1 (\partial_t g \cdot h) = A_k + B_k + C_k + D_k$, where

$$A_k = P_k \int_0^1 (\partial_t g)_{<k} \cdot h_{\geq k}$$

$$B_k = P_k \int_0^1 (\partial_t g)_{\geq k} \cdot h_{<k}$$

$$C_k = P_k \int_0^1 (\partial_t g)_{<k} \cdot h_{<k}$$

$$D_k = P_k \int_0^1 (\partial_t g)_{\geq k} \cdot h_{\geq k}$$

As in the Trichotomy Formula, C_k is essentially zero (with the exception of finitely many terms which can be subsumed in A_k, B_k, or D_k).

We now briefly sketch how to estimate each of A_k, B_k, D_k, leaving the details to be filled in. Note that P_k trivially commutes with the integrals $\int_0^1 dt$ and any partial derivatives ∂_t.

To estimate A_k, note that we can write (using **LP2**):

$$\|A_k\|_{L_x^2} \lesssim \sum_{k'<k\leq k''} \int_0^1 \|(\partial_t g)_{k'} \cdot h_{k''}\|_{L_x^2} dt$$

We can then use Bernstein inequality **LP4** and property **LP3** on h to pull out the power $2^{k'-k''}$. Writing $2^{k'-k''} \lesssim 2^{(k'-k)/2+(k-k'')/2}$, using **LP1**, and summing over k, we can then get:

$$\sum_{k\geq 0} \|A_k\|_{L_t^\infty L_x^2} \lesssim \|\partial_t g\|_{L_t^\infty L_x^2} \cdot \|\nabla h\|_{L_t^\infty L_x^2}$$

To estimate $D_k = P_k \int_0^1 (\partial_t g)_{\geq k} \cdot h_{\geq k}$, write

$$D_k = D_k^1 + D_k^2 = \sum_{k\leq k'\leq k''} P_k \int_0^1 (\partial_t g)_{k'} \cdot h_{k''} + \sum_{k\leq k'\leq k''} P_k \int_0^1 (\partial_t g)_{k''} \cdot h_{k'}$$

D_k^1 can be estimated straightforwardly, without integration by parts. Use **LP4** and **LP3** to write

$$\|D_k^1\|_{L_x^2} \lesssim 2^{k-k'} \|\partial_t g\|_{L_t^2 L_x^2} \cdot \|\nabla h\|_{L_t^2 L_x^2}$$

Then sum over k and use **LP1** to get:

$$\sum_{k\geq 0} \|D_k^1\|_{L_x^2} \lesssim \|\partial_t g\|_{L_t^2 L_x^2} \cdot \|\nabla h\|_{L_t^2 L_x^2}$$

To estimate D_k^2 we use integration by parts to transfer the ∂_t from the high-frequency $g_{k''}$ to the low-frequency $h_{k'}$. After integrating by parts we treat the result exactly as D_k^1. Thus, we need only estimate the boundary terms: $\|I_k(1) - I_k(0)\|_{L_x^2} \lesssim \|I_k\|_{L_t^\infty L_x^2}$, where

$$I_k = \sum_{k \le k' < k''} P_k(g_{k''} \cdot h_{k'})$$

We use the following lemma to do so:

Lemma 5.2. *For any k,k',k" we have*

$$\|P_k(g_{k'} \cdot h_{k''})\| \lesssim 2^{-\frac{1}{4}(|k'-k|+|k''-k|)} \|g_{k'}\| \|h_{k''}\|$$

Using this lemma, we integrate by parts and bound D_k^2 just as D_k^1 plus the boundary term, and eventually get:

$$\sum_k \|D_k^2\|_{L_x^2} \lesssim \|g\|_{H^1} \cdot \|h\|_{H^1}$$

Now we estimate B_k by similarly decomposing to $B_k = \sum_{k' < k \le k''} P_k \int_0^1 (\partial_t g)_{k''} \cdot h_{k'}$. As above, we integrate by parts and use the lemma to estimate the boundary terms $J_k = \sum_{k' < k \le k''} P_k(g_{k''}) \cdot h_{k'})$. It is then not hard to manipulate and sum over k to get

$$\sum_k \|B_k\|_{L_x^2} \lesssim \|g\|_{H^1} \cdot \|h\|_{H^1}$$

Combining all the estimates for A_k, B_k, and D_k completes the proof of the theorem.

It only remains to prove the above Lemma which helped us estimate the boundary terms. Without going into all the details, this is done by considering the three cases:

$$k' \ge k'' \ge k, \; k' \ge k > k'', \; k > k' \ge k''$$

We note that the third ("low-low") case is impossible. The other two cases are bounded using **LP3** and the the following (simple) calculus inequality:

$$\|f\|_{L_t^\infty L_x^2} \lesssim \|\partial_t f\|_{L_t^2 L_x^2}^{\frac{1}{2}} \cdot \|f\|_{L_t^2 L_x^2}^{\frac{1}{2}} + \|f\|_{L_t^2 L_x^2} \tag{150}$$

Estimating $\|P_k(g_{k'} \cdot h_{k''})\|_{L_t^\infty L_x^2}$ using (150) and **LP3** yields the estimate in the lemma. ∎

Exercise. Fill in the missing steps in the proof of the above theorem.

6. Calderon-Zygmund theory

The following L^2 identity

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

for any $u \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ can be easily established by integration by parts, see below in (154). Thus,

$$\|\partial^2 u\|_{L^2} \lesssim \|\Delta u\|_{L^2} \tag{151}$$

It is natural to ask whether such estimate still holds true for other L^p norms. It turns out that the problem can be reduced to that of study the L^p boundedness properties for a very important class of linear operators called Calderon-Zygmund.

Definition 6.1. A linear operator T acting on $L^2(\mathbb{R}^n)$ is called a *Calderon-Zygmund* operator if:

(1) T is bounded from L^2 to L^2.
(2) There exists a measurable kernel k such that for every $f \in L^2$ with compact support and for $x \notin \operatorname{supp} f$, we have

$$Tf(x) = \int_{\mathbb{R}^n} k(x-y) f(y)\mathrm{d}y,$$

where the integral converges absolutely for all x in the complement of $\operatorname{supp} f$.
(3) There exists constants $C > 1$ and $A > 0$ such that

$$\int_{|x|\geq C|y|} |k(x-y) - k(x)|\, \mathrm{d}x \leq A, \tag{152}$$

uniformly in y. For simplicity one can take $C = 2$.

Proposition 6.2. *Assume that the kernel $k(x)$ verifies, for all $x \neq 0$,*

$$|k(x)| \lesssim |x|^{-n}, \qquad |\partial k(x)| \lesssim |x|^{-n-1} \tag{153}$$

Then k verifies the cancellation condition (152).

Exercise. Prove the proposition.

Example 1. Hilbert transform $Hf(x) = \int e^{ix\cdot\xi} \operatorname{sign}\xi\, \hat{f}(\xi)\mathrm{d}\xi$. By Plancherel it is easy to check that H is a bounded linear operator on L^2. On the other hand we know that the inverse Fourier transform of $\operatorname{sign}\xi$ is proportional to the principal value distribution $\mathrm{pv}(1/x)$. Hence, if $x \notin \operatorname{supp} f$,

$$Hf(x) = c\int_{-\infty}^{+\infty} \frac{1}{x-y} f(y)dy.$$

It is easy to check that the kernel $k(x) = \frac{1}{x}$ verifies condition 3 above.

Example 2. Consider the equation $\Delta u = f$ in $\mathbb{R}^n$, $n \geq 3$, for f, smooth, compactly supported. Recall, see (**??**), that any solution u, vanishing at[7] ∞, can be represented in the form, $u = K_n * f$ where $K_n(x) = c_n|x|^{2-n}$. Thus, if $x \notin \operatorname{supp} f$, it makes sense to differentiate under the integral sign and derive,

$$\partial_i\partial_j u = \partial_i\partial_j K_n * f = \int_{\mathbb{R}^n} \partial_i\partial_j K_n(x-y) f(y)dy.$$

[7]In the case of $n = 2$ any solution whose first derivatives vanish at ∞.

It is easy to check that the kernel $k(x) = \partial_i\partial_j K_n(x)$ verifies condition 3. To show that the operators $R_{ij}f(x) = \int_{\mathbb{R}^n} \partial_i\partial_j K_n(x-y)f(y)dy$ are Calderon-Zygmund operators, it only remains to check the L^2 -boundedness property. This follows easily from the equation $\Delta u = f$. Indeed $u = K_n * f$ is the unique solution of the equation vanishing at ∞. Moreover $|u(x)| \lesssim |x|^{2-n}$, $|\partial u(x)| \lesssim |x|^{1-n}$ and $R_{ij}f = \partial_i\partial_j u(x)$. Thus we can integrate by parts in the expression,

$$\begin{aligned}\int_{\mathbb{R}^n} |f(x)|^2 dx &= \int_{\mathbb{R}^n} \Delta u(x)\Delta u(x)dx = \sum_{i,j=1}^n \int_{\mathbb{R}^n} |\partial_i\partial_j u(x)|^2 dx \\ &= \sum_{i,j=1}^n \int_{\mathbb{R}^n} |R_{ij}f(x)|^2 dx \end{aligned} \tag{154}$$

Hence for each pair $1 \le i, j \le n$,

$$\|R_{ij}f\|_{L^2} \le \|f\|_{L^2}.$$

Thus the operators R_{ij} are Calderon-Zygmund. We shall write schematically $R_{ij} = \partial_i\partial_j(-\Delta)^{-1}$.

Theorem 6.3. *Calderon-Zygmund operators are bounded from L^1 into weak-L^1.*

As a consequence we derive,

Corollary 6.4. *Calderon-Zygmund operators are bounded from L^p into L^p, for any $1 < p < \infty$. They are not bounded, in general, for $p = 1$ and $p = \infty$.*

Proof : The boundedness over L^p for $1 < p < 2$ follows from the weak-L^1 and the L^2 boundedness by Marcinkiewicz interpolation. The cases $p > 2$ follow by duality from the fact that the dual of a Calderon-Zygmund operator, with kernel $k(x)$, is again a Calderon-Zygmund operator, with kernel $k(-x)$. More precisely, if f, g have disjoint supports,

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

where

$$T^*g(y) = \int_{\mathbb{R}^n} k(-y+x)g(x)dx, \qquad \forall y \notin \ \mathrm{supp} g.$$

On the other hand $\|T^*f\|_{L^2} = \|Tf\|_{L^2} \lesssim \|f\|_{L^2}$. Hence T^* is indeed a CZ operator. Now, using the duality between L^p and $L^{p'}$, $1/p + 1/p' = 1$ and the fact that T^* is $L^{p'}$ bounded for $p' \le 2$,

$$\begin{aligned}\|Tf\|_{L^p} &= \sup_{\|g\|_{L^{p'}} \le 1} \left|\int_{\mathbb{R}^n} Tf(x)g(x)dx\right| = \sup_{\|g\|_{L^{p'}} \le 1} \left|\int_{\mathbb{R}^n} f(x)\, T^*g(x)dx\right| \\ &= \sup_{\|g\|_{L^{p'}} \le 1} \|f\|_{L^p} \cdot \|T^*g\|_{L^{p'}} \lesssim \|f\|_{L^p}.\end{aligned}$$

∎

We shall prove the main theorem 6.3 in the next two subsections.

6.5. Calderon-Zygmund decompositions.

During our study of the Laplace operator in the Introduction, we found that for $f \in \mathcal{C}_0^\infty$, $\Delta^{-1}(f)$ would decay rapidly away from the support of f provided $\int f(x)dx = 0$. This fact is physically important: it explains why we must have our hands "in contact" with an item in order to move it, even though the same electromagnetic force is well-known to move objects at much greater distances when there is a concentration of positive or negative charge. We also find that a related special behavior with respect to oscillation is quite important to the analysis of CZO's (of which Δ^{-1} is not an example, but its close relatives the Riesz potentials are). We therefore devote the following section to a way of decomposing a general function into one part which is bounded and other parts which oscillate and are physically localized, and this decomposition will allow us to prove theorem 6.3.

Definition 6.6. We define a *dyadic cube* in $\mathbb{R}^n$ to be a cube Q of the form

$$Q = [2^k a_1, 2^k(a_1+1)] \times \cdots \times [2^k a_n, 2^k(a_n+1)],$$

where $k, a_1, \ldots, a_n \in \mathbb{Z}$. We then say that $\operatorname{size}(Q) = 2^k$. If Q is a dyadic cubes then its *parent* is the only dyadic cube Q^* such that $Q \subset Q^*$ and $\operatorname{size}(Q^*) = 2\operatorname{size}(Q)$ and we say that Q is a *child* of Q^*.

Lemma 6.7 (Whitney decomposition)**.** *Any proper open set Ω in $\mathbb{R}^n$ can be covered by a family $\mathcal{Q} = \{Q\}$ of disjoint dyadic cubes*

$$\Omega = \cup_{Q \in \mathcal{Q}} Q,$$

where each cube $Q \in \mathcal{Q}$ satisfies the property

$$\textit{size}\,(Q) \approx \textit{dist}\,(Q, \partial\Omega). \tag{155}$$

Proof : For each $x \in \Omega$ denote by Q_x the largest dyadic cube containing x with the property: $\operatorname{dist}(Q_x, \partial\Omega) > \operatorname{size}(Q_x)$. If Q^* denotes the parent of Q_x then $\operatorname{dist}(Q^*, \partial\Omega) \le \operatorname{size}(Q^*)$. By the triangular inequality it follows that

$$\operatorname{dist}(Q_x, \delta\Omega) \le \sqrt{n}\operatorname{size}(Q_x) + \operatorname{dist}(Q^*, \delta\Omega) \le \left(\sqrt{n}+2\right)\operatorname{size}(Q_x).$$

Hence, Q_x verifies (155). If $y \in Q_x$ then, by the maximality property of Q_x and Q_y, we necessarily have $Q_y = Q_x$. Hence, the family $\mathcal{Q} = \{Q_x\}_{x\in\Omega}$ is formed of disjoint cubes and covers Ω. ■

Proposition 6.8 (Calderon-Zygmund decomposition)**.** *Let $f \in L^1(\mathbb{R}^n)$ and $\alpha > 0$. Then it is possible to find a countable family of disjoint dyadic cubes $\mathcal{Q} = \{Q\}$ and a decomposition $f = g + \sum_{Q\in\mathcal{Q}} b_Q$, such that:*

$$\|g\|_{L^\infty} \lesssim \alpha, \tag{156a}$$

$$\textit{supp}\; b_Q \subseteq Q, \tag{156b}$$

$$\int b_Q(x)dx = 0, \tag{156c}$$

$$\|b_Q\|_{L^1} \lesssim \alpha|Q|, \tag{156d}$$

$$\sum_Q |Q| \lesssim \frac{1}{\alpha}\|f\|_{L^1}. \tag{156e}$$

Remark Note that in the above α, g, b_Q and f all have the same units, so that these estimates on the sizes and supports of g and b_Q are the only ones possible that are still dimensionally correct.

Proof : Let $\mathcal{Q}$ be the Whitney decomposition of the open set $\Omega = \{\mathcal{M}f(x) > \alpha\}$ as indicated in Lemma (6.7). For each Q, define $f_Q = |Q|^{-1}\int_Q f(x)\mathrm{d}x$. Let

$$g(x) = \begin{cases} f(x), & \text{if } x \notin \Omega, \\ f_Q, & \text{if } x \in Q, \end{cases}$$

and $b_Q(x) = \chi_Q(x)(f(x) - f_Q)$ with χ_Q the characteristic function of the cube Q. Of course we have $f = g + \sum_Q b_Q$. The important property, which follows from (155), is that each cube Q is contained inside a ball B which is not entirely contained in Ω and with $|Q| \approx |B|$. Let $x \in B \setminus \Omega$, we have

$$|f_Q| \leq \frac{1}{|Q|}\int_Q |f(y)|\,\mathrm{d}y \lesssim \frac{1}{|B|}\int_B |f(y)|\,\mathrm{d}y \leq \mathcal{M}f(x) \leq \alpha. \tag{157}$$

We check now that this decomposition has the desired properties. For almost every x outside Ω, by Lebesgue's differentiation theorem, Corollary 2.7, we have $|g(x)| \leq \mathcal{M}f(x) \leq \alpha$. When $x \in \Omega$ it follows from (157) that $g(x) \lesssim \alpha$. Hence (156a) is satisfied. Properties (156b) and (156c) are immediate consequences of the definition of h_Q. Property (156d) is implied by (157). Finally, (156e) is nothing but the weak L^1 property for $\mathcal{M}f$ proved in Theorem 2.4. ∎

6.9. Proof of Theorem 6.3. Consider $f \in L^1$ and $\alpha > 0$. Let $f = g + \sum_Q b_Q = g + b$ be the Calderon-Zygmund decomposition of f according to Theorem 6.8. Since

$$\{|Tf(x)| > \alpha\} \subseteq \{|Tg(x)| > \alpha/2\} \cup (\{|Tb(x)| > \alpha/2\})$$

and in view of (156e) it is enough to prove separately that

$$|\{|Tg(x)| > \alpha/2\}| \lesssim \frac{1}{\alpha}\|f\|_{L^1}, \tag{158}$$

$$|\{|Tb(x)| > \alpha/2\}| \lesssim \frac{1}{\alpha}\|f\|_{L^1} \tag{159}$$

Estimate (158) follows from Chebyschev's inequality, the boundedness of T on L^2 and the uniform bound on g,

$$\begin{aligned} |\{|Tg(x)| > \alpha/2\}| &\lesssim \frac{1}{\alpha^2}\|Tg\|_{L^2}^2 \lesssim \frac{1}{\alpha^2}\|g\|_{L^2}^2 \lesssim \frac{1}{\alpha}\|g\|_{L^1} \leq \\ &\leq \frac{1}{\alpha}\left(\|f\|_{L^1} + \sum_Q \|b_Q\|_{L^1}\right) \lesssim \frac{1}{\alpha}\|f\|_{L^1} + \sum_Q |Q| \lesssim \frac{1}{\alpha}\|f\|_{L^1}. \end{aligned}$$

It remains to derive (159). Since the family $\mathcal{Q}$ is countable we denote its members by Q_j, $j \in \mathbb{N}$. For each Q_j let $y_{(j)}$ be its center and take $\hat{Q}_j$ to be the cube with the same center but with the sides expanded by $2n^{1/2}$, such that for all x in the complement of $\hat{Q}_j$,

$$|x - y_{(j)}| \geq 2\max_{y \in Q_j}|y - y_{(j)}|$$

Let $\Omega = \cup_j \hat{Q}_j$ and F its complement. We denote $b_j = b_{Q_j}$. Since $\int b_j dy = 0$ we write, for $x \in F$,

$$T(b_j)(x) = \int_{Q_j} \big(k(x-y) - k(x-y_{(j)})\big)\, b_j(y)\mathrm{d}y,$$

or, since the cubes Q_j are disjoint,

$$T(b_j)(x) = \int_{Q_j} \big(k(x-y) - k(x-y_{(j)})\big)\, b(y)\mathrm{d}y,$$

Thus, in view of (152),

$$\begin{aligned}
\int_F |T(b)(x)|dx &\leq \sum_j \int_F |T(b)(x)|dx \lesssim \sum_j \int_{x\in\mathbb{R}^n\setminus\hat{Q}_j} \int_{y\in Q_j} |k(x-y) - k(x-y_{(j)})|\,|b(y)|,\\
&= \sum_j \int_{y\in Q_j} |b_j(y)| \int_{x\in\mathbb{R}^n\setminus\hat{Q}_j} |k(x-y) - k(x-y_{(j)})|\\
&\leq \sum_j \int_{y\in Q_j} |b(y)| \int_{x\in\mathbb{R}^n\setminus\{\hat{Q}_j - y_{(j)}\}} |k(x-(y-y_{j)}) - k(x)|\\
&\lesssim \sum_j \int_{y\in Q_j} |b(y)| \int_{|x|\geq 2|(y-y_{j)}|} |k(x-(y-y_{j)}) - k(x)|\\
&\lesssim A\sum_j \int_{y\in Q_j} |b(y)| \lesssim \|f\|_{L^1}
\end{aligned}$$

Therefore,

$$|\{x\in F : |Tb(x)| > \alpha/2\}| \lesssim \alpha^{-1}\|f\|_{L^1}$$

On the other hand, the measure of the complement of F, i.e. $\Omega = \cup\hat{Q}_j$ is also controlled by,

$$|\Omega| \leq \sum_j |\hat{Q}_j| \lesssim \sum_j Q_j \lesssim \alpha^{-1}\|f\|_{L^1}.$$

Hence,

$$|\{x\in\mathbb{R}^n : |Tb(x)| > \alpha/2\}| \lesssim \alpha^{-1}\|f\|_{L^1}$$

as desired.

6.10. Michlin-Hörmander theorem. An important class of CZ operators can be defined by means of Fourier multiplier operators. Recall that these are defined by Fourier transform,

$$\widehat{Tf}(\xi) = m(\xi)\widehat{f}(\xi), \tag{160}$$

where m is a bounded function, called the multiplier. We can view these operators as convolution operators, $Tf = k * f$, where $\widehat{k} = m$. It is natural to ask when a Fourier multiplier operator gives rise to a CZ operator. Since we know that a CZO will grant extra decay to a localized function of mean zero, we would expect that the multiplier m should be fairly away from the origin. This is precisely the content of the following theorem.

Theorem 6.11. *Let $l > n/2$. Suppose m is a Fourier multiplier of class C^l on $\widehat{\mathbb{R}}^n \setminus 0$, such that*

$$\left|\partial_\xi^\alpha m(\xi)\right| \lesssim |\xi|^{-|\alpha|}, \qquad \forall \xi \in \widehat{\mathbb{R}}^n \setminus 0$$

for every multiindex α with $|\alpha| \leq l$. Then the operator defined by (160) *is a Calderon-Zygmund operator.*

Proof : Consider the same dyadic partition of unity as that used in the LP projections,

$$1 = \sum_{\lambda \in 2^{\mathbb{Z}}} \chi_\lambda(\xi) \qquad \text{for} \quad \xi \in \mathbb{R}^n \setminus 0,$$

generated by $\chi \in C_0^\infty$ with $\operatorname{supp}\chi \subseteq \{1/2 \leq |\xi| \leq 2\}$, and $\chi_\lambda(\xi) = \chi(\xi/\lambda)$.

Decompose m into dyadic pieces, $m = \sum_\lambda m_\lambda$, where $m_\lambda = \chi_\lambda m$. Since $|\partial^\gamma m(\xi)| \lesssim |\xi|^{-|\gamma|}$ and all derivatives of $\chi(\xi)$ are bounded,

$$\left|\partial_\xi^\alpha m_\lambda(\xi)\right| \leq \sum_{|\beta|+|\gamma| \leq |\alpha|} \left|\partial^\beta \chi_\lambda |\xi|^{-\gamma}\right| \lesssim \sum_{|\beta|+|\gamma| \leq |\alpha|} \lambda^{-|\beta|} \lambda^{-|\gamma|} \approx \lambda^{-|\alpha|}. \tag{161}$$

Let k_λ be the inverse Fourier transform of m_λ. Since m_λ has compact support k_λ is a smooth function. Moreover, for any integer N we have[8]

$$|k_\lambda(x)| \lesssim |x|^{-N} \|\partial^N m_\lambda\|_{L^1} \lesssim |x|^{-N} \lambda^{n-N}.$$

Now take $N > n$ and sum over $\lambda \in 2^{\mathbb{Z}}$. Observe that $\sum_\lambda k_\lambda$ converges to a well defined measurable function k on $\mathbb{R}^n \setminus 0$, and it easy to see that k satisfies property 2 of Definition 6.1.

The boundedness of T on L^2 follows immediately from the boundedness of m on $\widehat{\mathbb{R}}^n$.

For $0 \leq j \leq l$, by Plancherel's theorem and (161) we obtain

$$\int |x|^{2j} \left|k_\lambda(x)\right|^2 \mathrm{d}x \simeq \sum_{|\alpha|=j} \int \left|\partial_\xi^\alpha m_\lambda(\xi)\right|^2 \mathrm{d}\xi \lesssim \lambda^{n-2j}.$$

Let $R > 0$, using the case $j = 0$ we find that

$$\int_{|x| \leq R} |k_\lambda(x)| \,\mathrm{d}x \lesssim \left(\int |k_\lambda(x)|^2 \,\mathrm{d}x\right)^{1/2} R^{n/2} \lesssim (\lambda R)^{n/2}, \tag{162}$$

while using the case $j = l$ we find that

$$\int_{|x| \geq R} |k_\lambda(x)| \,\mathrm{d}x \lesssim \left(\int |x|^{2l} |k_\lambda(x)|^2 \,\mathrm{d}x\right)^{1/2} \left(\int_{|x|>R} \frac{\mathrm{d}x}{|x|^{2l}}\right)^{1/2} \lesssim (\lambda R)^{n/2-l}. \tag{163}$$

[8]Recall that, by integration by parts, we have $\left|\mathcal{F}^{-1} f(x)\right| \leq |x|^{-N} \|\partial_\xi^N f\|_{L^1}$,

If we choose $R = 1/\lambda$, summing (162) and (163) we obtain $\|k_\lambda\|_{L^1} \lesssim 1$ uniformly in λ. We can apply the same procedure to ∂k_λ, which has symbol $\xi m_\lambda \approx \lambda m_\lambda$, to prove that $\|\partial k_\lambda\|_{L^1} \lesssim \lambda$. Hence,

$$\begin{aligned} \int_{|x|\gg|y|} |k_\lambda(x-y) - k_\lambda(x)|\,\mathrm{d}x &\le \int\int_0^{|y|} |\partial k_\lambda(x - ty/|y|)|\,\mathrm{d}t\mathrm{d}x && (164) \\ &= |y| \cdot \|\partial k_\lambda\|_{L^1} \lesssim \lambda|y|, && (165) \end{aligned}$$

but also, by (163),

$$\int_{|x|\gg|y|} |k_\lambda(x-y) - k_\lambda(x)|\,\mathrm{d}x \le 2\int_{|x|\ge|y|} |k_\lambda(x)|\,\mathrm{d}x \lesssim (\lambda|y|)^{n/2-l}\,. \tag{166}$$

We sum over λ using (164) when $\lambda|y| \le 1$ and (166) when $\lambda|y| > 1$, and obtain[9]

$$\int_{|x|\gg|y|} |k(x-y) - k(x)|\,\mathrm{d}x \lesssim |y| \sum_{\lambda\le|y|^{-1}} \lambda + |y|^{n/2-l} \sum_{\lambda>|y|^{-1}} \lambda^{n/2-l} \lesssim 1.$$

as desired. ■

Exercise. Let $\phi \in \mathcal{C}_0^\infty(\mathbb{C})$ and let f be the solution to the inhomogeneous Cauchy-Riemann equations $\frac{\partial f}{\partial \bar z} = \phi$ which decays at infinity. Show that for $1 < p < \infty$ we have the estimate

$$||\partial f||_{L^p} \lesssim ||\phi||_{L^p}$$

6.12. Square function estimates. We recall property LP6 for the square function, $Sf = \left(\sum_k |P_k f|^2\right)^{1/2}$,

Theorem 6.13 (Littlewood-Paley)**.** *We have,*

$$\|f\|_{L^p} \lesssim \|Sf\|_{L^p} \lesssim \|f\|_{L^p} \tag{167}$$

for all $1 < p < \infty$*.*

We give two proofs of this estimate.

Proof [first proof]: First we show using duality arguments that the first inequality in (167) follows from the second one. Indeed using Plancherel's theorem, the fact that $P_k P_{k'} = 0$ unless $k \sim k'$, and Cauchy-Schwartz inequality we obtain

$$\begin{aligned} \int f(x)g(x)\mathrm{d}x &\simeq \int \sum_{k\approx k'} P_k f(x) P_{k'} g(x)\mathrm{d}x \\ &\lesssim \int \left(\sum_k |P_k f(x)|^2\right)^{1/2} \left(\sum_{k'} |P_{k'} g(x)|^2\right)^{1/2} \mathrm{d}x \le \\ &\lesssim \|Sf\|_{L^p} \|Sg\|_{L^{p'}} \lesssim \|Sf\|_{L^p} \|g\|_{L^{p'}}. \end{aligned}$$

The left inequality in (167) now follows by taking the sup over all g with $\|g\|_{L^{p'}} = 1$.

[9] Here we used the following summation properties, in dyadic notation, for geometric series, $\sum_{\lambda\le L} \lambda^\alpha \simeq L^\alpha$ and $\sum_{\lambda\ge L} \lambda^{-\alpha} \simeq L^{-\alpha}$ for $\alpha > 0$.

To prove the right inequality in (167) we need to introduce the Rademacher functions $r_k(t)$ defined on $\mathbb{R}$ as follows: for every $k \geq 0$, $k \in \mathbb{Z}$ and $t \in \mathbb{R}$ set $r_k(t) = r_0(2^k t)$, where $r_0(t)$ is the periodic function, $r_0(t+1) = r_0(t)$, such that $r_0(t) = 1$ for $0 \leq t < 1/2$ and $r_0(t) = -1$ for $1/2 \leq t < 1$. These Rademacher functions form an orthonormal sequence in $L^2[0,1]$ and they form a sequence of independent identically distributed random variables. The basic property that we need is that the L^p norm of a linear combination of Rademacher function is equivalent to the l^2 norm of its coefficients.

Lemma 6.14. *Given a sequence of real numbers $\{a_k\}$ satisfying $\sum_{k=0}^{\infty} a_k^2 < \infty$, define*

$$F(t) = \sum_{k=0}^{\infty} a_k r_k(t).$$

Then $F \in L^2([0,1])$ with $\|F\|_{L^2} = (\sum_{k=0}^{\infty} a_k^2)^{1/2}$. In addition, $F \in L^p([0,1])$ for $1 < p < \infty$, and there exist constants A_p so that

$$A_p^{-1}\|F\|_{L^p} \leq \|F\|_{L^2} \leq A_p\|F\|_{L^p}.$$

For a proof of this lemma see Stein, [**?**, Appendix D].

Define the operator T_t so that

$$T_t f = \sum_{k=0}^{\infty} r_k(t) P_k f$$

Clearly T_t is the Fourier multiplier operator with symbol $m_t(\xi) = \sum_k r_k(t)\chi(2^{-k}\xi)$, where χ is the smooth cut-off function used to define the LP projections. For $\xi \neq 0$, at most three of the terms in the sum defining $m_t(\xi)$ can be non-zero. We can then easily verify that m_t verifies the condition of Thm. 6.11. That is, that

$$|\partial_\xi^\alpha m_t(\xi)| \leq C_\alpha |\xi|^{-|\alpha|},$$

with constants C_α independent of t. Thus, by Calderon-Zygmund theory (specifically Corollary 6.4), we have:

$$\|T_t f\|_{L^p} \lesssim \|f\|_{L^p}$$

And so,

$$\left(\int_0^1 \|T_t f\|_{L^p}^p dt\right)^{1/p} \lesssim \|f\|_{L^p}$$

In addition, we can use Lemma 6.14 to see that:

$$\begin{aligned}\int_0^1 \|T_t f\|_{L^p}^p dt &= \int_0^1 \int_{\mathbb{R}} |\sum_k r_k(t)(P_k f)(x)|^p dx dt \\ &\gtrsim \int_{\mathbb{R}} \left(\sum_k |(P_k f)(x)|^2\right)^{p/2} dx\end{aligned}$$

And so combining our results we get:

$$\|Sf\|_{L^p} \lesssim \|f\|_{L^p}$$

(Note that this argument proves the theorem only in the one-dimensional case, $n = 1$. It can, however, be extended to $\mathbb{R}^n$ as in Stein, *Singular Integrals*, Ch. IV, Section 5.)

■

Proof [second proof]: We recall the definition for the vector-valued function,

$$\mathbf{S}f(x) = \big(P_k f(x)\big)_{k\in\mathbb{Z}}.$$

Clearly, if $f \in \mathcal{S}(\mathbb{R}^n)$, for every $x \in \mathbb{R}^n$, $\mathbf{S}f(x) \in l^2$ and $Sf(x) = |\mathbf{S}f(x)|$ denotes the l^2 norm of $\mathbf{S}f(x)$. We claim that

$$\mathbf{S}f(x) = \int \mathbf{K}(x-y)f(y)dy$$

is a an l^2 -valued Calderon-Zygmund operator with the l^2-valued kernel defined by,

$$\mathbf{K}(x) = \big(K_k(x)\big)_{k\in\mathbb{Z}}, \qquad K_k(x) = 2^{nk}\hat{\chi}(2^k x)$$

Denote $|\mathbf{K}(x)| = \big(\sum_k |K_k(x)|^2\big)^{1/2}$, $|\partial\mathbf{K}(x)| = \big(\sum_k |\partial K_k(x)|^2\big)^{1/2}$. We easily check that the l^2- valued version of the condition (153) is verified,

$$|\mathbf{K}(x)| \lesssim |x|^{-n} \qquad |\partial\mathbf{K}(x)| \lesssim |x|^{-(n+1)}, \qquad \text{for} \quad x \neq 0. \tag{168}$$

On the other hand,

$$\|\mathbf{S}f\|_{L^2} := \|Sf\|_{L^2} \lesssim \|f\|_{L^2}.$$

Thus $\mathbf{S}$ is indeed an l^2 valued C-Z operator and therefore, in view of a straightforward extension of Theorem 6.3 and its corollary, we infer that,

$$\|\mathbf{S}f\|_{L^p} := \||\mathbf{S}f|\|_{L^p} = \|Sf\|_{L^p} \lesssim \|f\|_{L^p}$$

In view of the beginning of the first proof of our theorem we infer that also,

$$\|f\|_{L^p} \lesssim \|Sf\|_{L^p}.$$

■

Remark that, according to theorem 6.13, $|\sum_k P_k f| \approx \big(\sum_k |P_k f|^2\big)^{1/2}$. A more general principle asserts that if a sequence of functions $f_1, f_2, \dots f_k \dots$ oscillate at different rates, that is any two phases are different, then $|\sum_k f_k| \approx \big(\sum_k |f_k|^2\big)^{1/2}$.

The following version of the property LP6, and theorem 6.13, also holds true for LP projections $\tilde{P}_k \sim P_k$. More precisely,

$$\|\big(\sum_k |\tilde{P}_k f|^2\big)^{1/2}\|_{L^p} \lesssim \|f\|_{L^p}, \qquad 1 < p < \infty. \tag{169}$$

This can be proved in the same manner as the inequality $\|Sf\|_{L^p} \lesssim \|f\|_{L^p}$ by introducing the l^2 valued operator, $\tilde{\mathbf{S}}f = (\tilde{P}_k f)_{k\in\mathbb{Z}}$, and proceeding exactly as in the second proof of theorem 6.13. Given an l^2 valued vector function $\mathbf{g} = (g_k)_{k\in\mathbb{Z}}$ observe that

$$< \tilde{\mathbf{S}}f, \mathbf{g} >= \int_{\mathbb{R}^n} \tilde{\mathbf{S}}f(x)\cdot\bar{\mathbf{g}}(x)dx = \int_{\mathbb{R}^n}\sum_k \tilde{P}_k f(x)\overline{g_k}(x)dx = \int_{\mathbb{R}^n} f(x)\overline{\sum_k \tilde{P}_k g_k(x)}dx$$

Thus,

$$\tilde{\mathbf{S}}^*\mathbf{g} = \sum_k \tilde{P}_k g_k \tag{170}$$

and therefore the estimate dual to (169) has the form, $\|\tilde{\mathbf{S}}^*\mathbf{g}\|_{L^{p'}} \lesssim \|\mathbf{g}\|_{L^{p'}}$, for $1/p + 1/p' = 1$. In other words,

$$\|\sum_k \tilde{P}_k g_k\|_{L^p} \lesssim \|\big(\sum_k |g_k|^2\big)^{1/2}\|_{L^p}, \qquad 1 < p < \infty. \tag{171}$$

The following is an easy consequence of theorem 6.13.

Corollary 6.15. *For* $2 \le p < \infty$ *we have*

$$\|f\|_{L^p}^2 \lesssim \sum_{k\in\mathbb{Z}} \|P_k f\|_{L^p}^2. \tag{172}$$

For $1 < p \le 2$ *we have*

$$\sum_{k\in\mathbb{Z}} \|P_k f\|_{L^p}^2 \lesssim \|f\|_{L^p}^2. \tag{173}$$

Proof : Recall that $Sf(x)^2 = \sum_{k\in\mathbb{Z}} |P_k f|^2$. If $p/2 \ge 1$, in view of LP6 and Minkowski inequality, we have

$$\|f\|_{L^p}^2 \lesssim \|Sf\|_{L^p}^2 = \|\sum_k |P_k f|^2\|_{L^{p/2}} \le \sum_k \|\,|P_k f|^2\,\|_{L^{p/2}} = \sum_k \|P_k f\|_{L^p}^2.$$

If $p/2 \le 1$, we make use instead of the reverse Minkowski inequality,

$$\|f\|_{L^p}^2 \gtrsim \|\sum_k |P_k f|^2\|_{L^{p/2}} \ge \sum_k \|\,|P_k f|^2\,\|_{L^{p/2}} = \sum_k \|P_k f\|_{L^p}^2.$$

The reverse Minkowski inequality we have used here states that for $0 < q \le 1$ and a sequence of positive functions $(f_k)_{k\in\mathbb{Z}}$

$$\|\sum_k |f_k|\|_{L^q} \ge \sum_k \|f_k\|_{L^q}. \tag{174}$$

We briefly sketch a proof of (174); it can be found in many books (e.g. Garling, *Inequalities* or DiBenedetto, *Real Analysis*, from which we take this particular proof).

One way is to first prove a reverse Hölder inequality: For $0 < p < 1$, $q < 0$, $\frac{1}{p} + \frac{1}{q} = 1$, $f \in L^p$, $g \in L^q$, we have $\int |fg| \ge \|f\|_{L^p}\|g\|_{L^q}$. This can be easily shown by writing $\|f\|_{L^p} = \left(\int \frac{|fg|^p}{|g|^p}\right)^{1/p}$ and applying the usual Hölder inequality with the exponents $\tilde{p} = 1/p > 1$ and $\tilde{q} = 1/(1-p) > 1$.

With this in hand, the reverse Minkowski inequality in two terms ($\||f|+|g|\|_{L^q} \ge \|f\|_{L^q} + \|g\|_{L^q}$ for $0<q\le 1$) follows (writing $\frac{1}{q'} = 1 - \frac{1}{q}$):

$$\begin{aligned}
\||f|+|g|\|_{L^q}^q &= \int (|f|+|g|)^{q-1}(|f|+|g|) \\
&\ge \Big(\int (|f|+|g|)^{(q-1)q'}\Big)^{1/q'} (\|f\|_{L^q} + \|g\|_{L^q}) \\
&\ge \||f|+|g|\|_{L^q}^{q-1}(\|f\|_{L^q} + \|g\|_{L^q})
\end{aligned}$$

∎

6.16. $W^{s,p}$**- Sobolev spaces.** We recall that we have defined the $W^{s,p}$ norm of a function by,

$$\|f\|_{W^{s,p}} = \sum_{j=0}^{s} \|\partial^j f\|_{L^p}$$

We claim the following

Lemma 6.17. *For any $j \ge 0$, $1<p<\infty$ we have,*

$$\|\partial^j f\|_{L^p} \approx \|\big(\sum_k |2^{jk} P_k f|^2\big)^{1/2}\|_{L^p}$$

Proof : We first write,

$$\|\partial^j f\|_{L^p} \lesssim \|\sum_k \partial^j P_k f\|_{L^p}$$

As in the proof of the property LP5, we can express $\nabla^j P_k f = 2^{jk}\tilde{P}_k P_k f$ for some P_k similar to P_k. Hence, using the estimate (171)

$$\|\partial^j f\|_{L^p} \lesssim \|\sum_k 2^{jk}\tilde{P}_k P_k f\|_{L^p} \lesssim \|\big(\sum_k |2^{jk} P_k f|^2\big)^{1/2}\|_{L^p}.$$

On the other hand, we can also write $2^{jk} P_k f = \tilde{P}_k \partial^j f$ for some other similar LP projection. Then, in view of (169),

$$\|\big(\sum_k |2^{jk} P_k f|^2\big)^{1/2}\|_{L^p} \lesssim \|\big(\sum_k |\tilde{P}_k \partial^j f|^2\big)^{1/2}\|_{L^p} \lesssim \|\partial^j f\|_{L^p}$$

∎

Using the lemma we can now find an equivalent definition using LP projections:

Proposition 6.18. *For any $1<p<\infty$ and any $s \in \mathbb{N}$ we have,*

$$\|f\|_{W^{s,p}} \approx \|\sum_k (1+2^k)^s P_k f\|_{L^p}. \tag{175}$$

Moreover, for the homogeneous $W^{s,p}$ norm $\|f\|_{\dot{W}^{s,p}} = \|\partial^s f\|_{L^p}$,

$$\|f\|_{\dot{W}^{s,p}} \approx \|\sum_k 2^{ks} P_k f\|_{L^p}. \tag{176}$$

Observe that the expressions on the right hand side of (175) and (176) make sense for every value $s \in \mathbb{R}$. We can thus extend the definitions of $W^{s,p}$, and $\dot{W}^{s,p}$ spaces to all real values s.

Additional characterizations of the homogeneous Sobolev norms $\| \, \|_{\dot{W}^{s,p}}$ can be given using the following,

Proposition 6.19. *For $2 \le p < \infty$ and any s we have,*

$$\left(\sum_k 2^{kps} \|P_k f\|_{L^p}^p\right)^{1/p} \lesssim \|f\|_{\dot{W}^{s,p}} \lesssim \left(\sum_k 2^{2ks} \|P_k f\|_{L^p}^2\right)^{1/2}. \tag{177}$$

For $1 < p \le 2$ and $s \in \mathbf{R}$ we have

$$\left(\sum_k 2^{2ks} \|P_k f\|_{L^p}^2\right)^{1/2} \lesssim \|f\|_{\dot{W}^{s,p}} \lesssim \left(\sum_k 2^{kps} \|P_k f\|_{L^p}^p\right)^{1/p}. \tag{178}$$

Proof : If $p/2 \ge 1$, by Theorem 6.13 and Minkowski inequality we have

$$\|f\|_{L^p}^2 \lesssim \| \sum_k |P_k f|^2 \|_{L^{p/2}} \le \sum_k \| |P_k f|^2 \|_{L^{p/2}} = \sum_k \|P_k f\|_{L^p}^2.$$

If $p/2 \le 1$, by Theorem 6.13 and the reverse Minkowski inequality we have

$$\|f\|_{L^p}^2 \gtrsim \| \sum_k |P_k f|^2 \|_{L^{p/2}} \ge \sum_k \| |P_k f|^2 \|_{L^{p/2}} = \sum_k \|P_k f\|_{L^p}^2.$$

The remaining details should be clear to fill in. ■

7. Problems

Problem 1.[Distributions in $\mathbb{R}$]

Let $f(z)$ be a an analytic function in the domain $D_+ = \{z \in \mathbb{C} : 0 < \Im(z) < \epsilon\}$ such that $|f(z)| \lesssim |\Im(z)|^{-N}$ for all $z \in D$. Show that there exists a distribution $f_+ = f(\cdot + i0)$ such that for every $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$,

$$\lim_{y \to 0, y > 0} \int_{\mathbb{R}} f(x+iy)\phi(x)dx \quad = \quad < f_+, \phi >,$$

Similarly, for analytic functions defined on $D_- = \{z \in \mathbb{C}/) - \epsilon < \Im(z) < 0\}$ we can define a distribution $f_- = f(\cdot - i0)$,

$$\lim_{y \to 0, y < 0} \int_{\mathbb{R}} f(x+iy)\phi(x)dx \quad = \quad < f_-, \phi >$$

This defines, in particular when $f = \frac{1}{z} = \frac{1}{x+iy}$, the distributions $(x+i0)^{-1}$ and $(x-i0)^{-1}$. Prove the formulas,

$$(x+i0)^{-1} - (x-i0)^{-1} \quad = \quad -2\pi i \delta_0(x).$$

Show also that,

$$(x+i0)^{-1} = x^{-1} - i\pi\delta_0(x)$$

where $\frac{1}{x}$ is the principal value distribution defined in the text.

Problem 2.[Fundamental solutions] Consider the operator $Lu = \Delta u + u$ in $\mathbb{R}^3$. Find all solutions of $Lu = 0$ with spherical symmetry. Show that

$$K(x) = -\frac{\cos|x|}{4\pi|x|}$$

is a fundamental solution for L.

Problem 3.[Initial value problem] Consider the initial value problems for the following, four evolution equations in $\mathbb{R} \times \mathbb{R}^n$,

$$\partial_t u = \Delta u, \qquad u(0,x) = f(x) \tag{179}$$

$$\partial_t u = i\Delta u, \qquad u(0,x) = f(x) \tag{180}$$

$$\partial_t^2 u = \Delta u, \qquad u(0,x) = f(x), \quad \partial_t u(0,x) = g(x) \tag{181}$$

$$\partial_t^2 u = -\Delta u, \qquad u(0,x) = f(x), \quad \partial_t u(0,x) = g(x) \tag{182}$$

In each of these cases write down solutions using the Fourier transform method. In other words take the Fourier transform of each equation, set

$$\hat{u}(t,\xi) = \int e^{-ix\cdot\xi} u(t,x)dx$$

and solve the resulting differential equation in t. Compare the results for the last two equations. Show that (181) has solutions for all $f, g \in \mathcal{S}(\mathcal{R}^n)$ while (182) does not. Show however that if we only prescribe $u(0,x) = f$ (this is the Dirichlet problem for the Laplacian $\partial_t^2 + \Delta$ in $\mathbb{R}^{n+1}$), then the problem has a unique solution u, which decays to zero as $|t| + |x| \to \infty$, for all functions $f \in \mathcal{S}(\mathbb{R}^n)$. In all cases express[10] the resulting solutions as integral operators applied to the initial data(in physical space).

Problem 4.[Extension operator] Let H be the half space $x_n > 0$ in $\mathbb{R}^n$ and $1 \le p \le \infty$. Show that there exists an extension operator, that is a bounded linear operator $E : W^{1,p}(H) \to W^{1,p}(\mathbb{R}^n)$ such that for all $u \in W^{1,p}(H)$ we have $Eu = u$ a.e. in H and

$$\|Eu\|_{W^{1,p}(\mathbb{R}^n)} \lesssim \|u\|_{W^{1,p}(H)}.$$

Extend the result to any $s \in \mathbb{N}$. Can you extend the result to arbitrary domains $U \subset \mathbb{R}^n$? What about domains with smooth boundaries ?

Problem 5.[Distributions and Fourier Analysis on the Circle] A smooth function on the circle $\mathbb{R}/\mathbb{Z}$ is a smooth function on $\mathbb{R}$ which is 1-periodic

$$f(x+k) = f(x), \quad k \in \mathbb{Z}$$

The circle has a discrete space of frequencies $m \in \widehat{(\mathbb{R}/\mathbb{Z})} = \mathbb{Z}$ corresponding to the functions $x \mapsto e^{2\pi imx}$. The discreteness of the frequency space is intimately related

[10] You will have to perform the inverse Fourier tarnsform, $u(t,x) = \mathcal{F}^{-1}\hat{u}(t,\xi)$. For the wave equation this is more difficult, in general, but you can do it for dimension $n = 3$.

to the compactness of the circle. A Schwartz function on the circle is just a smooth function; a Schwartz function on $\mathbb{Z}$ is one which decays faster than any polynomial at infinity.

a. We define the Fourier transform of a periodic function $\hat{f}(m) = \int_0^1 f(x)e^{-2\pi imx}$. Prove the Fourier inversion formula

$$f(x) = \sum_{m\in\mathbb{Z}} \hat{f}(m)e^{2\pi imx}$$

for smooth functions on the circle. Deduce the Plancharel formula $< f, g >=< \hat{f}, \hat{g} >$.

b. We define a distribution u on the circle to be an element of the dual of $\mathcal{C}^k(\mathbb{R}/\mathbb{Z})$ for some k, i.e. $< u, \phi >\leq C||\phi||_{\mathcal{C}^k}$ for some k, C and all $\phi \in \mathcal{C}^\infty(\mathbb{R}/\mathbb{Z})$. The circle has a smooth structure, so it is possible to formulate the notion of a fundamental solution for a differential operator (the group structure on the circle allows convolution to make sense as well) – however it is not always possible to find such a solution. Show that there is no fundamental solution u to the operator $\frac{d}{dx}$. In other words, there is no distribution u for which $\frac{du}{dx} = \delta(x)$ in the sense that

$$< \frac{du}{dx}, \phi >\equiv - < u, \frac{d\phi}{dx} >= \phi(0), \quad \phi \in \mathcal{C}^\infty(\mathbb{R}/\mathbb{Z})$$

There are many ways to prove this. Can you see this in both physical and frequency space? What if we replace the vector field $\frac{d}{dx}$ by another nonvanishing vector field $\tilde{D} = \psi\frac{d}{dx}$ for some nonvanishing, smooth function $\psi \in \mathcal{C}^\infty(\mathbb{R}/\mathbb{Z})$?

Problem 6.[Trace theorems] Let $\mathbb{R}^{n-1}$ be a hyperplane in $\mathbb{R}^n$, for example $x_n = 0$. For any $f \in \mathcal{S}(\mathbb{R}^n)$ let Rf denote the restriction of f to $\mathbb{R}^{n-1}$.

i. Prove that, for any $s > \frac{1}{2}$,

$$\|Rf\|_{L^2(\mathbb{R}^{n-1})} \lesssim \|f\|_{H^s(\mathbb{R}^n)} \tag{183}$$

ii. Show that the result is not true for $s \leq 1/2$. Show however that the following sharp trace theorem holds for all $s > 0$,

$$\|Rf\|_{H^s(\mathbb{R}^{n-1})} \lesssim \|f\|_{H^{s+1/2}(\mathbb{R}^n)} \tag{184}$$

iii. Show that f is a function with Fourier support in the ball $|\xi| \lesssim 2^k$ for some integer k then, for all $1 \leq p \leq \infty$ and $s > 1/p$,

$$\|f\|_{L^p(\mathbb{R}^{n-1})} \lesssim 2^{k/p}\|f\|_{W^{s,p}(\mathbb{R}^n)}$$

Can you deduce from here a trace result, in L^p norms, generalizing that of (183) ? What about (184) ?

iv. Let H be the half space $x_n > 0$. According to the above considerations we can talk about the trace of a function in $W^{1,p}(H)$ to the hyperplane $x_n = 0$(Prove this !). Show that a function $f \in W^{1,p}(H)$ belongs[11] to $W_0^{1,p}(H)$ if and only if its trace to $x_n = 0$ is zero.

[11]recall that $W_0^{1,p}(H)$ is the closure of $\mathcal{C}_0^\infty(H)$ in $W^{1,p}(H)$

Problem 7.[Littlewood-Paley] Consider the spaces $\Lambda_\gamma = C^{0,\gamma}(\mathbb{R}^n)$ with norm

$$\|f\|_{\Lambda_\gamma} = \|f\|_{L^\infty(\mathbb{R}^n)} + \sup_{x\neq y\in\mathbb{R}^n} \frac{|u(x)-u(y)|}{|x-y|^\gamma}$$

i. Show, using the Littlewood-Paley projections P_k, that

$$\|f\|_{\Lambda^\gamma} \approx \|P_{\leq 0} f\|_{L^\infty} + \sup_{k>0} 2^{k\gamma} \|P_k f\|_{L^p}$$

ii. Define the Zygmund class Λ_* of functions with norm,

$$\|f\|_{\Lambda^*} = \|f\|_{L^\infty} + \sup_{x\in\mathbb{R}^n,\ 0\leq h\leq 1} \frac{|f(x+h)+f(x-h)-2f(x)|}{h}$$

Show that

$$\|f\|_{\Lambda_*} \approx \|P_{\leq 0} f\|_{L^\infty} + \sup_{k>0} 2^k \|P_k\|_{L^p}.$$

iii. Prove the product estimate in Besov spaces $B^s = H^{s,1}$, $s > 0$.

$$\|fg\|_{B^s} \lesssim \|f\|_{L^\infty}\|g\|_{B^s} + \|g\|_{L^\infty}\|f\|_{B^s}$$

Problem 8. Read on your own the section on Calderon-Zygmund operators. Indicate how the theory can be extended to operators valued in a given Hilbert space, such as l^2.

8. Restriction Theorems

It is well known that when $f \in L^1(\mathbb{R}^n)$ then its Fourier transform $\hat{f}$ is a bounded and continuous function, thus the restriction of $\hat{f}$ to any hypersurface is perfectly well defined. On the other hand, if $f \in L^2(\mathbb{R}^n)$ then $\hat{f}$ may be any function in L^2, hence defined only almost everywhere and completely arbitrary on sets of measure zero like hypersurfaces.

Can one make sense of the restriction of $\hat{f}$ to a smooth hypersurface S when f belongs to some L^p with $1 < p < 2$? This is a basic question in modern Fourier analysis, which, as we shall see, turns out to be intimately tied to regularity properties of solutions to wave equations.

If we take S to be a hyperplane, we immediately see that the answer is negative. Indeed, let $f(x_1, x') = u(x_1)v(x')$, $\hat{f}(\xi_1, \xi') = \hat{u}(\xi_1)\hat{v}(\xi')$, with $x_1, \xi_1 \in \mathbb{R}$ and $x', \xi' \in \mathbb{R}^{n-1}$. The restriction of $\hat{f}$ to the hyperplane $\xi_1 = 0$ is well defined only when $\hat{u}(0) = \int u(x)\mathrm{d}x$ is well defined. For any $p > 1$ it is always possible to find $u \in L^p(\mathbb{R})$ such that $\int u\mathrm{d}x$ doesn't make sense. We deduce that the restriction of the Fourier transform on hyperplanes cannot be defined when $p > 1$.

The answer is different if we consider hypersurfaces which have non vanishing curvature. For simplicity we consider the model case of the sphere.

8.1. The Stein-Tomas theorem. The following type of result was first proved by Stein [], then extended by Tomas [] and given its final form again by Stein [].

Theorem 8.2 (Stein-Tomas)**.** *Let $\mathbb{S} = \mathbb{S}^{n-1}$ be the standard unit sphere in $\mathbb{R}^n$ and $d\sigma$ its standard volume element. Let $f \in L^p(\mathbb{R}^n)$ with*

$$1 \le p \le \frac{2(n+1)}{n+3}.$$

Then $\mathcal{R}f = \hat{f}\Big|_S \in L^2(\mathbb{S})$ and

$$\|\mathcal{R}f\|_{L^2(\mathbb{S})} \lesssim \|f\|_{L^p(\mathbb{R}^n)}.$$

This theorem has an equivalent dual formulation. Define the Stein operator to be the dual of the Fourier restriction operator $\mathcal{R}f = \hat{f}\Big|_{\mathbb{S}}$,

$$\mathcal{S}g(x) = \mathcal{R}^* g(x) = \int_{\mathbb{S}} e^{ix\cdot\xi} g(\xi) \mathrm{d}\sigma_\xi \simeq (g\mathrm{d}\sigma)^\vee(x),$$

where now g is a function defined on the sphere.

Theorem 8.3. *Let $f \in L^2(\mathbb{S})$ and*

$$\frac{2(n+1)}{n-1} \le p \le \infty.$$

Then $\mathcal{S}f \in L^p(\mathbb{R}^n)$ and

$$\|\mathcal{S}f\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^2(\mathbb{S})}. \tag{185}$$

Remark 8.4. It suffices to prove Theorem 8.3 for $p = p_* = 2(n+1)/(n-1)$. Indeed for $p > p_*$, by Sobolev inequality we have

$$\|\mathcal{S}f\|_{L^p} \lesssim \|D^s \mathcal{S}f\|_{L^{p_*}}$$

for $s = n(1/p_* - 1/p) > 0$, where $(D^s u)^\wedge(\xi) = |\xi|^s \hat{u}(\xi)$. But here

$$D^s\mathcal{S}f = \mathcal{S}(|\cdot|^s f) = \mathcal{S}f.$$

Thus, if we can prove the theorem when $p = p_*$ then

$$\|\mathcal{S}f\|_{L^p} \lesssim \|\mathcal{S}f\|_{L^{p_*}} \lesssim \|f\|_{L^2(\mathbb{S})}$$

Remark 8.5. The result remains true if we replace $\mathrm{d}\sigma$ by $\mathrm{d}\mu = \psi\mathrm{d}\sigma$, with $\psi \in C_0^\infty(\mathbb{R}^n)$, since the theorem implies

$$\|(f\mathrm{d}\mu)^\vee\|_{L^p} \lesssim \|f\psi\|_{L^2(\mathbb{S})} \lesssim \|f\|_{L^2(\mathbb{S})}.$$

Moreover, using a partition of unity, it suffices to prove Theorem 8.3 just for $\mathcal{S}f = (f\mathrm{d}\mu)^\vee$, with $\mathrm{d}\mu = \psi\mathrm{d}\sigma$ and $\psi \in C_0^\infty(\mathbb{R}^n)$ supported in a small neighborhood of a point on the sphere. Though obvious, it is a very important fact that we can localize the restriction estimate as we shall see in the future.

8.6. Knapp counterexample. The result of theorem 8.3 is false for any $p < p_*$ in virtue of the following counterexample ([**?**]).

Define, for some small $\delta > 0$, the region in phase space

$$D = \left\{ \xi \in \mathbb{R}^n : |\xi_1 - 1| < \delta^2,\, |\xi'| < \delta \right\}.$$

Let now $f = \chi_{\mathbb{S}\cap D}$ be the characteristic function of the cap $\mathbb{S} \cap D$, then

$$\|f\|_{L^2(\mathbb{S})} = |\mathbb{S} \cap D|^{1/2} \sim \delta^{(n-1)/2}.$$

We can write

$$\mathcal{S}f(x) = e^{ix_1} \int_{\mathbb{S}\cap D} e^{i\phi(x,\xi)} \mathrm{d}\sigma_\xi,$$

with phase $\phi(x, \xi) = x_1(\xi_1 - 1) + x' \cdot \xi'$. It then possible to fix a region in physical space,

$$R = \left\{ x \in \mathbb{R}^n : |x_1| < \frac{\pi}{6}\delta^{-2}, |x'| < \frac{\pi}{6}\delta^{-1} \right\},$$

such that for $x \in R$ and $\xi \in D$ we have $|\phi(x, \xi)| \leq \pi/3$, hence, when $x \in R$,

$$|\mathcal{S}f(x)| \geq \mathrm{Re}(e^{-ix_1}\mathcal{S}f(x)) = \int_{\mathbb{S}\cap D} \cos(\phi(x,\xi))\mathrm{d}\sigma_\xi \geq \frac{1}{2}|\mathbb{S} \cap D|.$$

This implies that

$$\frac{\|\mathcal{S}f\|_{L^p}}{\|f\|_{L^2}} \gtrsim |\mathbb{S} \cap D|^{1/2}|R|^{1/p} \sim \delta^{\frac{n-1}{2} - \frac{n+1}{p}}.$$

For small values of δ, an estimate like (185) will necessarily require $\frac{n-1}{2} - \frac{n+1}{p} \geq 0$, which is possible only if $p \geq p_* = 2(n + 1)/(n - 1)$.

This example suggests that there is some sort of parabolic scaling property in the structure of the operator $\mathcal{S}$ which comes from the nonvanishing curvature of the sphere.

8.7. The importance of curvature. The restriction theorem and its dual counterpart remain true if we replace the standard sphere $\mathbb{S}^{n-1}$ by a compact hypersurface $H \subset \mathbb{R}^n$ with non-vanishing Gauss curvature. The importance of non-vanishing Gauss curvature is illustrated by the following result.

Lemma 8.8. *Let $H \subset \mathbb{R}^n$ be a compact hypersurface with non-vanishing Gauss curvature (i.e. with all its principal curvatures different from zero) and volume element $d\sigma$. Then, for any smooth function ψ, we have,*

$$|(\psi d\sigma)^\vee(x)| \lesssim (1 + |x|)^{-\frac{n-1}{2}} \tag{186}$$

If exactly one principal curvature vanishes then we have instead,

$$|(\psi d\sigma)^\vee(x)| \lesssim (1 + |x|)^{-\frac{n-2}{2}}$$

Proof The general proof is based on the method of stationary phase, see Stein's Harmonic Analysis book. For the particular case of the standard sphere $H = \mathbb{S}^{n-1}$ and odd n the proof can be done by a direct computation in polar coordinates. ■

Exercise Prove the lemma for $\mathbb{S}^2 \subset \mathbb{R}^3$.

Remark 8.9. Another interesting observation links these restiction theorems to partial differential equations. Indeed if $u = \mathrm{d}\sigma^{\vee} * f$, then u is a solution of the linear elliptic equation

$$\Delta u + u = 0,$$

as we can be easily seen taking the Fourier transform,

$$\mathcal{F}(u + \Delta u)(\xi) \simeq (1 - |\xi|^2)\delta(1 - |\xi|)\hat{f}(\xi) = 0.$$

where δ is the Dirac distribution.

8.10. TT^* **principle.** The following simple functional analysis result plays an important role in restriction and Strichartz type estimates. Let B be a Banach space and denote by B' its dual. Let H be an Hilbert space with inner product denoted by $\langle \cdot, \cdot \rangle$. Consider a linear operator $T : H \to B'$. Since we can identify H with its dual, we can consider T to be the adjoint of the operator $T^* : B \to H$ defined by

$$\langle h, T^*(x) \rangle = Th(x).$$

Actually, T^* is the adjoint of T when B is reflexive, but for our purposes we shall keep calling T^* the adjoint of T.

The TT^* principle states that the boundedness of T is equivalent to the boundedness of TT^*. More precisely we have:

Proposition 8.11. *The following statements are equivalent:*

(i) $T : H \to B'$ *is bounded and* $\|T\| = M$*;*
(ii) $T^* : B \to H$ *is bounded and* $\|T^*\| = M$*;*
(iii) $TT^* : B \to B'$ *is bounded and* $\|TT^*\| = M^2$*;*
(iv) *the bilinear form* $(x, y) \mapsto \langle T^*x, T^*y \rangle$ *is bounded on* $B \times B$ *with norm* M^2*.*

The proof is a standard exercise in functional analysis.

8.12. TT^* **formulation of the restriction theorem.** The TT^* formulation for the Stein operator corresponds to a convolution with the (inverse) Fourier transform of the measure on the sphere. Formally, we have,

$$SS^*f(x) = SRf(x) = \int_{\mathbb{S}} e^{ix\cdot\xi}\hat{f}(\xi)\mathrm{d}\sigma_\xi = \int_{\mathbb{R}^n}\int_{\mathbb{S}} e^{i(x-y)\cdot\xi}\mathrm{d}\sigma_\xi\, f(y)\mathrm{d}y = \mathrm{d}\sigma^{\vee} * f(x).$$

We are thus led to the following equivalent form of the restriction theorem,

$$\|\mathrm{d}\sigma^{\vee} * f\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^{p'}(\mathbb{R}^n)}, \tag{187}$$

for $p \geq p^*$.

One can give three distinct proofs of Theorem 8.3. We shall sketch the first proof based on analytic interpolation. This is essentially the original proof of Stein and Tomas. The second proof, based on introducing a time parameter and treating $\mathcal{S}f$ as an evolution operator allows us to regard the restriction theorem as part of a more general framework which includes Strichartz estimates for various linear

PDE with constant coefficients. Finally the third approach, which only applies for specific exponents, will allow us to to connect with bilinear estimates.

8.13. First proof: analytic interpolation. According to Remark 8.12 and Remark 8.4 it suffices to prove that $Uf = \mathrm{d}\sigma^\vee * f$ verifies

$$\|Uf\|_{L^{p_*}(\mathbb{R}^n)} \lesssim \|f\|_{L^{p'_*}(\mathbf{R}^n)}, \tag{188}$$

where $p_* = 2(n+1)/(n-1)$ and $p'_* = 2(n+1)/(n+3)$.

In general, to obtain $L^{p'} - L^p$ estimates directly is usually very complicated and we don't know any direct proof except in cases where p is a nice exponent like $p = 4, 6$ (which happens only for $n = 2$ or $n = 3$). We would feel more comfortable with $L^2 - L^2$ type estimates, where Plancherel's theorem is a powerful tool, or with $L^1 - L^\infty$ type estimates, since pointwise decay estimates of oscillatory integrals can be obtained from stationary phase methods. This suggests to use some interpolation theory for L^p spaces. But, an $L^2 - L^2$ estimate for the operator U is ruled out by the Knapp counterexample and a $L^\infty - L^1$ one is too trivial and doesn't answer to our question. It is here that the Stein interpolation theorem, Thm. 1.8, shows its power, since it allows us to obtain the $L^{p'} - L^p$ estimate for U from $L^2 - L^2$ and $L^\infty - L^1$ estimates for other (reasonable) operators different from U.

We will accomplish this by constructing a family of convolution operators $U_z f = \mu_z^\vee * f$, with μ_z being distributions depending analytically in z. The parameter z will essentially reflect the degree of homogeneity of the distribution μ_z. For this reason it is natural to place our target at $z = -1$, requiring $U_{-1} = U$ or $\mu_{-1} = \mathrm{d}\sigma$, since $\mathrm{d}\sigma$ can be written as the pullback of a delta distribution (which is homogeneous of degree -1) on the sphere: $\mathrm{d}\sigma \simeq \delta(1 - |\xi|)\mathrm{d}\xi$.

An $L^2 - L^2$ estimate for U_z will follow if μ_z coincides with a bounded function, indeed, by Plancherel's theorem, we have

$$\|U_z f\|_{L^2} \simeq \|(U_z f)^\wedge\|_{L^2} \simeq \|\mu_z \cdot \hat{f}\|_{L^2} \lesssim \|\mu_z\|_{L^\infty} \|f\|_{L^2}. \tag{189}$$

To have $\mu_z(\xi)$ bounded we must require that $\mu_z(\xi)$ is essentially homogeneous of degree 0, hence when z lies on the line $\mathrm{Re}(z) = 0$.

An $L^1 - L^\infty$ estimate for U_z will follow instead when $\mu_z^\vee$ coincides with a bounded function, since we directly have

$$\|U_z f\|_{L^\infty} \lesssim \|\mu_z^\vee\|_{L^\infty} \|f\|_{L^1}. \tag{190}$$

To obtain (188) from the analytic interpolation of (189) and (190), we would like the latter to happen on the line $\mathrm{Re}(z) = a$, where a is chosen so that

$$-1 = \theta a + (1-\theta)0, \quad \frac{1}{p_*} = \frac{\theta}{\infty} + \frac{1-\theta}{2}, \quad \frac{1}{p'_*} = \frac{\theta}{1} + \frac{1-\theta}{2},$$

and this happens precisely when $\mathrm{Re}(z) = a = -(n+1)/2$.

This argument leads to the precise version of the Stein analytic interpolation theorem that we are going to use.

Proposition 8.14. *Let U_z be an analytic family of linear operators such that:*

(i) $U_{-1} = U$;
(ii) $\|U_z f\|_{L^2} \lesssim \|f\|_{L^2}$, *uniformly on the line* $Re(z) = 0$;
(iii) $\|U_z f\|_{L^\infty} \lesssim \|f\|_{L^1}$, *uniformly on the line* $Re(z) = -(n+1)/2$.

Then it follows that

$$\|Uf\|_{L^{p_*}} \lesssim \|f\|_{L^{p'_*}}.$$

The above discussion showed that, when we write $U_z f$ as the convolution $\mu_z^\vee * f$, then the hypothesis of the proposition are fulfilled whenever μ_z is an analytic family of distribution such that

(i') $\mu_{-1} = \mathrm{d}\sigma$;
(ii') $\mu_z(\xi)$ coincides with a bounded function, with a uniform bound on the line $\mathrm{Re}(z) = 0$;
(iii') $\mu_z^\vee(x)$ coincides with a bounded function, with a uniform bound on the line $\mathrm{Re}(z) = -(n+1)/2$.

It thus remains to define the distributions μ_z and verify these properties.

Inspired by the identity $\delta = \chi_+^{-1}$ and $\mathrm{d}\sigma_\xi \simeq \delta(1 - |\xi|)$, we define our family of distributions as

$$\mu_z(\xi) = e^{z^2} \chi_+^z (1 - |\xi|) \psi(|\xi|), \tag{191}$$

where $\psi \in C_0^\infty(\mathbf{R})$ is a cut-off function supported in a small neighborhood of 1, say $[1/2, 3/2]$, and $\psi(1) = 1$.

We recall that the homogeneous distributions χ_+^z, when $\mathrm{Re}(z) > -1$, coincide with the functions:

$$\chi_+^z(t) = \begin{cases} t^z/\Gamma(z+1) & \text{if } t \geq 0, \\ 0 & \text{if } t < 0, \end{cases}$$

where the Gamma function is defined by $\Gamma(z+1) = \int_0^\infty t^z e^{-t} \mathrm{d}t$. From the identity $\Gamma(z+1) = z\Gamma(z)$, it follows that

$$\frac{\mathrm{d}}{\mathrm{dt}} \chi_+^z(t) = \chi_+^{z-1}(t). \tag{192}$$

Using this formula, χ_+^z can be analytically continued for all $z \in \mathbb{C}$ by performing repeated integrations by parts. To do this we first observe that for $\mathrm{Re}(z) > -1$ and $\phi \in C_0^\infty$ we have

$$\int \chi_+^z(t)\phi(t)\mathrm{d}t = -\int \chi_+^{z+1}(t)\phi'(t)\mathrm{d}t = \ldots = (-1)^m \int \chi_+^{z+m}(t)\phi^{(m)}(t)\mathrm{d}t.$$

Thus integrating by parts sufficiently many times we can make sense of $\int \chi_+^z \phi \mathrm{d}t$ when $\mathrm{Re}(z) > -1 - m$ for any m, and hence for all z. To see that $\chi_+^{-1} = \delta$ it takes just an integration by parts, indeed

$$\int \chi_+^{-1} \phi \mathrm{d}t = -\frac{1}{\Gamma(1)} \int_0^\infty \phi'(t)\mathrm{d}t = \phi(0).$$

For more information about χ_+^z and distribution theory one can consult the books by Gel'fand and Shilov [**Ge-S**] or Hormander [**?**].

The factor e^{z^2} in the definition of μ_z is chosen in order to garantee a uniform boundedness of our operators for large $\Im(z)$, indeed e^{z^2} decreases exponentially as $\Im(z) \to \infty$, uniformly on the strip $-(n+1)/2 \leq \mathrm{Re}(z) \leq 0$. This permits to allow the various constants in the following inequalities to have a polynomial growth in terms of $b = \Im(z)$.

Clearly $\mu_{-1} \simeq \delta(1-|\xi|)\psi(|\xi|) \simeq \mathrm{d}\sigma$. This verifies (i').

Condition (ii') is immediately verified, since χ_+^{-z} is always a bounded function when $\mathrm{Re}(z) = 0$. Condition (iii') will follow from stationary phase arguments, more generally we have:

Proposition 8.15.

$$|\mu_z^\vee(x)| \lesssim (1+|x|)^{-Re(z)-1-\frac{n-1}{2}}. \tag{193}$$

8.16. Second proof: evolution operators approach. In this section we make the following assumption on f:

$$f \in C^\infty(\mathbb{S}), \qquad \mathrm{supp} f \subset \{\xi_1 > 1/2\}. \tag{194}$$

With this assumption we can relabel $x_1 = t$ as a time parameter and rewrite $\mathcal{S}f$ as

$$\begin{aligned} \mathcal{S}f(t,x') &= \int_{|\xi'|<\sqrt{3}/2} e^{it\sqrt{1-|\xi'|^2}} e^{ix'\cdot\xi'} f(\sqrt{1-|\xi'|^2}, \xi') \frac{\mathrm{d}\xi'}{\sqrt{1-|\xi'|^2}} \\ &= \int e^{it\sqrt{1-|\xi'|^2}} e^{ix'\cdot\xi'} \beta(|\xi'|) g(\xi') \mathrm{d}\xi'. \end{aligned}$$

with $\beta \in C_0^\infty$ supported in $|\xi'| < 1$ and $g(\xi') = f(\sqrt{1-|\xi'|^2}, \xi')/\sqrt{1-|\xi'|^2}$. Observe that

$$\int |g(\xi')|^2 \mathrm{d}\xi' = \int_{\mathbb{S}} \frac{|f(\xi)|^2}{|\xi_1|^2} \mathrm{d}\sigma_\xi \simeq \|f\|^2_{L^2(\mathbb{S})}$$

by the assumption on the support of f.

Theorem 8.17. *Let $\beta \in C_0^\infty(\mathbf{R}^{n-1})$ be supported in the unit ball $\{\xi \in \mathbf{R}^{n-1} : |\xi| < 1\}$ and consider the operator*

$$Tg(t,x) = \int_{\mathbf{R}^{n-1}} e^{it\sqrt{1-|\xi|^2}} e^{ix\cdot\xi} \beta(\xi) g(\xi) d\xi, \quad t \in \mathbf{R},\ x \in \mathbf{R}^{n-1}.$$

Let q, r be Lebesgue exponents verifing the conditions:

$$0 \leq \frac{2}{q} \leq \min\{1, \gamma(r)\}, \tag{195}$$

$$\left(\frac{2}{q}, \gamma(r)\right) \neq (1,1), \tag{196}$$

where $\gamma(r) = (n-1)(1/2 - 1/r)$. Then the following estimate holds true for all $g \in C_0^\infty(\mathbf{R}^{n-1})$,

$$\|Tg\|_{L^q_t L^r_x(\mathbf{R}\times\mathbf{R}^{n-1})} \lesssim \|g\|_{L^2(\mathbf{R}^{n-1})}. \tag{197}$$

where we use the mixed norm notation defined in section 7.

By Remark 8.5, Theorem 8.3 follows from the special case $q = r = 2\frac{n+1}{n-1}$.

Remark 8.18. We can run again the Knapp example to prove the necessity of condition (195), when $q \geq 2$. Indeed let $D \subset \mathbb{R}^{n-1}$ be the disk defined by $|\xi| \leq \delta$, for sufficiently small $\delta > 0$, and take $g = \chi_D$ to be the characteristic function of D. We write,

$$Tg(t,x) \quad = \quad e^{it}\int_D e^{it(\sqrt{1-|\xi|^2}-1)}e^{ix\cdot\xi}\beta(\xi)d\xi$$

and observe that for $|t| \leq \delta^{-2}$ and $|x| \leq \delta^{-1}$ we have, with a fixed constant $c > 0$, $|Tg(t,x)| \geq c$. Indeed this follows easily from $\xi| \leq \delta$ and $|\sqrt{1-|\xi|^2}-1| \lesssim \delta^2$. Therefore, if (197) holds true, we must have, for all sufficiently small $\delta > 0$,

$$c\delta^{-\frac{2}{q}}\delta^{-\frac{n-1}{r}} \lesssim \|Tg\|_{L^q_tL^r_x} \lesssim \|\chi_D\|_{L^2} \lesssim \delta^{-\frac{n-1}{2}}$$

from which (195), $q \geq 2$ follows.

Remark 8.19. The end-point restriction (196) can be removed when $n \neq 3$, due to a well known result by Keel and Tao [**K-T**] ("Endpoint Strichartz Inequalities"). The other restriction $q \geq 2$, implicit in (195) will be discussed in the next chapter.

We start by calculating T^* and TT^*.

$$\begin{aligned} < T^*F, g > =< F, Tg >= & \iint F\overline{Tg}\mathrm{d}t\mathrm{d}x = \\ = & \iint F(t,x)\int e^{-it\sqrt{1-|\xi|^2}}e^{-ix\cdot\xi}\overline{\beta(\xi)g(\xi)}\mathrm{d}\xi\mathrm{d}t\mathrm{d}x = \\ = & \int \overline{g(\xi)\beta(\xi)}\left(\iint e^{-it\sqrt{1-|\xi|^2}}e^{-ix\cdot\xi}F(t,x)\mathrm{d}t\mathrm{d}x\right)\mathrm{d}\xi. \end{aligned}$$

Hence

$$T^*F(\xi) = \overline{\beta(\xi)}\iint e^{-it\sqrt{1-|\xi|^2}}e^{-ix\cdot\xi}F(t,x)\mathrm{d}t\mathrm{d}x,$$

and

$$\begin{aligned} TT^*F(t,x) = & \int e^{it\sqrt{1-|\xi|^2}}e^{ix\cdot\xi}\beta(\xi)T^*F(\xi)\mathrm{d}\xi \\ = & \iint e^{i(t-s)\sqrt{1-|\xi|^2}}e^{ix\cdot\xi}|\beta(\xi)|^2\hat{F}(s,\xi)\mathrm{d}\xi\mathrm{d}s, \end{aligned}$$

where $\hat{F}(s,\xi) = \int e^{-ix\cdot\xi}F(s,x)\mathrm{d}x$. If we introduce the family of operators

$$U(t)f(x) = \int e^{it\sqrt{1-|\xi|^2}}e^{ix\cdot\xi}|\beta(\xi)|^2\hat{f}(\xi)\mathrm{d}\xi,$$

we can write TT^* as a convolution operator,

$$TT^*F(t,\cdot) = \int U(t-s)F(s,\cdot)\mathrm{d}s. \tag{198}$$

By Proposition 8.11, to show that T is a bounded operator from $L^q_tL^r_x(\mathbf{R}^n)$ to $L^2(\mathbf{R}^{n-1})$ it suffices to prove that TT^* is a bounded operator from $L^{q'}_tL^{r'}_x(\mathbf{R}^n)$ to $L^q_tL^r_x(\mathbf{R}^n)$.

We shall first prove an estimate for $U(t)$.

Proposition 8.20. *Let $2 \leq r \leq \infty$ and $\gamma(r) = (n-1)(1/2 - 1/r)$. Then $U(t)$ verifies the estimate*

$$\|U(t)f\|_{L^r(\mathbf{R}^{n-1})} \lesssim (1+|t|)^{-\gamma(r)} \|f\|_{L^{r'}(\mathbf{R}^{n-1})}. \tag{199}$$

Proof Once we have proved the two extreme cases $r = 2$ and $r = \infty$,

$$\|U(t)f\|_{L^2(\mathbf{R}^{n-1})} \lesssim \|f\|_{L^2(\mathbf{R}^{n-1})} \tag{200}$$

$$\|U(t)f\|_{L^\infty(\mathbf{R}^{n-1})} \lesssim (1+|t|)^{-(n-1)/2} \|f\|_{L^1(\mathbf{R}^{n-1})} \tag{201}$$

then the estimate follows from the standard Riesz interpolation theorem.

We obtain (200) immediately using Plancherel formula, since

$$(U(t)f)^\wedge(\xi) \simeq e^{it\sqrt{1-|\xi|^2}} |\beta(\xi)|^2 \hat{f}(\xi).$$

To prove (201) we write

$$U(t)f(x) = \int K_t(x-y) f(y) \mathrm{d}y,$$

where

$$\begin{aligned} K_t(x) &= \int e^{ix\cdot\xi} e^{it\sqrt{1-|\xi|^2}} |\beta(\xi)|^2 \mathrm{d}\xi \\ &\simeq \iint e^{ix\cdot\xi} e^{it\tau} \delta(1-\tau^2-|\xi|^2) \sqrt{1-|\xi|^2} |\beta(\xi)|^2 \mathrm{d}\tau \mathrm{d}\xi \\ &\simeq \iint e^{i(t,x)\cdot(\tau,\xi)} \delta(1-|(\tau,\xi)|) \beta_1(\tau,\xi) \mathrm{d}\tau \mathrm{d}\xi, \qquad \left(\beta_1(\tau,\xi) = \tau|\beta(\xi)|^2\right), \\ &= \left(\beta_1 \mathrm{d}\sigma_{n-1}\right)^\vee (t,x). \end{aligned}$$

Hence K_t is just the Fourier transform of a measure supported on the sphere $\mathbb{S}^{n-1}$, for which we have the decay estimate

$$|K_t(x)| \lesssim (1+|t|+|x|)^{-(n-1)/2},$$

which implies (201). ■

We next apply Proposition 8.20 to (198),

$$\|TT^*F(t,\cdot)\|_{L^r_x} \lesssim \int \frac{1}{(1+|t-s|)^{\gamma(r)}} \|F(s,\cdot)\|_{L^{r'}_x} \mathrm{d}s. \tag{202}$$

Finally, we are in a position to apply the Hardy-Littlewood-Sobolev inequality and, if $0 < \gamma(r) < 1$, we obtain

$$\|TT^*F\|_{L^q_t L^r_x} \lesssim \|F\|_{L^{q'}_t L^{r'}_x},$$

when $-\gamma(r) + 1 + 1/q = 1/q'$, hence $\gamma(r) = 2/q$. Therefore we proved Theorem 8.17 in the case $0 < \gamma(r) = 2/q < 1$.

On the other hand if $q = 2$ and $\gamma(r) > 1$ we have from (202),

$$\|TT^*F\|_{L^2_t L^r_x} \lesssim \|F\|_{L^2_t L^{r'}_x},$$

by an application of the standard Hausdorff-Young inequality.

Finally, if $2/q < 1$ and $\gamma(r) > 2/q$ the result follows from the case $\gamma(r) = 2/q$ using Sobolev inequalities.

8.21. Third proof: bilinear forms ($n = 2$ and $n = 3$). We present now another method to prove the restriction theorem for the sphere that works for the special cases $n = 2$, $p = 6$ or $n = 3$, $p = 4$. The idea is that when p is an even integer, the restriction theorem can be viewed as an L^2 estimate for a multilinear form, which, through the Fourier transform, has a convolution structure that provides some smoothing effects. The proofs given below are at the root of the so called bilinear trilinear estimates, which play a fundamental role in the modern theory of nonlinear wave and dispersive equations.

Let us see the case $n = 3$ first. We consider the Stein operator $\mathcal{S}f = (f\mathrm{d}\sigma)^\vee$, and use the fact that $(\mathcal{S}f \cdot \mathcal{S}f)^\wedge \simeq (f\mathrm{d}\sigma) * (f\mathrm{d}\sigma)$. Let $B(f,g) = \mathcal{S}f \cdot Sg$, then an L^4 estimate for $\mathcal{S}f$ corresponds to an L^2 estimate for $B(f,f)$. We have

$$\hat{B}(f,g)(\xi) \simeq (f\mathrm{d}\sigma) * (g\mathrm{d}\sigma)(\xi) = \int_{\mathbf{R}^3} \delta(1 - |\xi - \eta|)\delta(1 - |\eta|) f(\xi - \eta) g(\eta) \mathrm{d}\eta,$$

and applying Cauchy-Schwarz with respect to the measure $\delta(1-|\xi-\eta|)\delta(1-|\eta|)\mathrm{d}\eta$ we find

$$|\hat{B}(f,g)(\xi)|^2 \leq \hat{B}(1,1)(\xi)\hat{B}\left(|f|^2, |g|^2\right)(\xi).$$

Integrating with respect to ξ, we obtain

$$\|B(f,g)\|^2_{L^2(\mathbf{R}^3)} \lesssim A \|f\|^2_{L^2(\mathbb{S}^2)} \|g\|^2_{L^2(\mathbb{S}^2)}, \tag{203}$$

with

$$A = \sup_\xi |\hat{B}(1,1)(\xi)| = \sup_\xi \int \delta(1 - |\xi - \eta|)\delta(1 - |\eta|)\mathrm{d}\eta. \tag{204}$$

Thus, to prove the theorem in this case it suffices to check that A is finite. It is useful to carry out the explicit calculation of $A(\xi) = \hat{B}(1,1)(\xi)$. For any dimension $n \geq 2$ we have:

Lemma 8.22.

$$A(\xi) = \int_{\mathbf{R}^n} \delta(1 - |\xi - \eta|)\delta(1 - |\eta|)\, d\eta \simeq \frac{1}{|\xi|}\left(4 - |\xi|^2\right)_+^{\frac{n-3}{2}}. \tag{205}$$

Proof

$$A(\xi) = \int \delta(1 - |\xi - \eta|)\delta(1 - |\eta|)\mathrm{d}\eta \simeq \int_{|\eta|=1} \delta(1 - |\xi - \eta|^2)\mathrm{d}\sigma_\eta =$$

$$= \int_{|\eta|=1} \delta(|\xi|^2 - 2\xi \cdot \eta)\mathrm{d}\sigma_\eta \simeq \frac{1}{|\xi|}\int_{|\eta|=1} \delta\left(\frac{|\xi|}{2} - \frac{\xi}{|\xi|}\cdot\eta\right)\mathrm{d}\sigma_\eta.$$

Because of the rotational symmetry, we may assume that $\xi = (|\xi|, 0, \dots, 0)$, so that

$$A(\xi) \simeq \frac{1}{|\xi|}\int_0^\pi \delta\left(\frac{|\xi|}{2} - \cos\theta\right)(\sin\theta)^{n-2}\mathrm{d}\theta =$$

$$= \frac{1}{|\xi|}\int_{-1}^1 \delta\left(\frac{|\xi|}{2} - u\right)(1-u^2)^{\frac{n-3}{2}}\mathrm{d}u = \frac{1}{|\xi|}\left(1 - \frac{|\xi|^2}{4}\right)^{\frac{n-3}{2}},$$

when $|\xi|/2 \in [-1,1]$. ■

When $n=3$, $A(\xi) \simeq 1/|\xi|$ is singular only at $\xi = 0$, but we can avoid this difficulty by assuming that f and g are supported in a small neighborhood of a point in S^2 (recall that without loss of generality we can localize the estimate on a small cap on the sphere). Then the supremum in (204) can be taken over just all $\xi \in \operatorname{supp}(f) + \operatorname{supp}(g)$, which is a set bounded away from 0. Hence we may restrict to $|\xi| \geq C > 0$ in (204) and the singularity disappears leaving $A < \infty$.

From the L^2 estimate (203) of the bilinear form $B(f,g)$, it follows the L^4 estimate for the Stein operator $\mathcal{S}f$:

$$\|\mathcal{S}f\|^2_{L^4(\mathbf{R}^3)} = \|B(f,f)\|_{L^2} \simeq A^{1/2}\|f\|^2_{L^2(\mathbb{S}^2)},$$

with the assumption that f is supported in a small cap on the sphere.

In the case $n=2$ what we want is an L^6 estimate for $\mathcal{S}f$. Since $6 = 3\times 2$ we can try to repeat the same calculation using this time a trilinear form, $T(f,g,h) = \mathcal{S}f \cdot Sg \cdot Sh$, and the fact that $\|\mathcal{S}f\|^3_{L^6} = \|T(f,f,f)\|_{L^2}$. We have

$$\begin{aligned}\hat{T}(f,g,h)(\xi) &\simeq (f\mathrm{d}\sigma) * (g\mathrm{d}\sigma) * (h\mathrm{d}\sigma)(\xi) = \\ &= \iint_{\mathbb{R}^2\times\mathbb{R}^2} \delta(1-|\xi-\eta-\zeta|)\delta(1-|\eta|)\delta(1-|\zeta|)f(\xi-\eta)g(\eta)h(\zeta)\mathrm{d}\eta\mathrm{d}\zeta,\end{aligned}$$

and applying Cauchy-Schwarz with respect to the measure $\delta(1-|\xi-\eta|)\delta(1-|\eta|)\delta(1-|\zeta|)\mathrm{d}\eta\mathrm{d}\zeta$ we find

$$|\hat{T}(f,g,h)(\xi)|^2 \leq \hat{T}(1,1,1)(\xi)\hat{T}\left(|f|^2,|g|^2,|h|^2\right)(\xi).$$

Integrating with respect to ξ, we obtain

$$\|T(f,g,h)\|^2_{L^2(\mathbb{R}^2)} \lesssim A\|f\|^2_{L^2(\mathbb{S}^1)}\|g\|^2_{L^2(\mathbb{S}^1)}\|h\|^2_{L^2(\mathbb{S}^1)}, \tag{206}$$

with

$$A = \sup_\xi |\hat{T}(1,1,1)(\xi)| = \sup_\xi \iint \delta(1-|\xi-\eta|)\delta(1-|\eta|)\delta(1-|\zeta|)\mathrm{d}\eta\mathrm{d}\zeta. \tag{207}$$

The convolution structure allows us to restrict ξ to the set $\operatorname{supp}f + \operatorname{supp}g + \operatorname{supp}h$, and, if we make the hypothesis of f,g,h supported in a small cap of the sphere, we can assume $1 \leq |\xi| \leq 3$. Using Lemma 8.22 we can evaluate $T(1,1,1)$ and show that A is bounded,

$$\begin{aligned} T(1,1,1)(\xi) &= \int B(1,1)(\xi-\zeta)\,\delta(1-|\zeta|)\mathrm{d}\zeta \sim \\ &\sim \int_{|\xi-\zeta|<2} \frac{\delta(1-|\zeta|)}{(4-|\xi-\zeta|^2)^{1/2}}\mathrm{d}\zeta = \int_{\substack{\zeta\in\mathbb{S}^1\\ |\xi-\zeta|<2}} \frac{\mathrm{d}\sigma_\zeta}{(3-2\xi\cdot\zeta+|\xi|^2)^{1/2}} \simeq \\ &\simeq \int_{a(\xi)}^1 \frac{\mathrm{d}a}{(3-|\xi|^2+2|\xi|a)^{1/2}(1-a^2)^{1/2}} \sim \int_{a(\xi)}^1 \frac{\mathrm{d}a}{(a-a(\xi))^{1/2}(1-a)^{1/2}} \simeq 1,\end{aligned}$$

where $a(\xi) = -\frac{3-|\xi|^2}{2|\xi|}$. From the L^2 estimate (206) of the trilinear form $T(f,g,h)$, it follows the L^6 estimate for the Stein operator $\mathcal{S}f$:

$$\|\mathcal{S}f\|^3_{L^6(\mathbb{R}^2)} = \|T(f,f,f)\|_{L^2} \simeq A^{1/2}\|f\|^3_{L^2(\mathbb{S}^2)}.$$

We can also try to repeat the bilinear argument for $n = 2$. As before, for $B(f, g) = \mathcal{S}f \cdot Sg$ we have

$$|\hat{B}(f,g)(\xi)|^2 \leq \hat{B}(1,1)(\xi)\hat{B}\left(|f|^2, |g|^2\right)(\xi).$$

Integrate with respect to ξ, and use Lemma 8.22 to evaluate $\hat{B}(1,1)$,

$$\|B(f,g)\|_{L^2(\mathbb{R}^2)}^2 \lesssim \iint \frac{\delta(1-|\xi-\eta|)\delta(1-|\eta|)}{|\xi|(4-|\xi|^2)^{1/2}}|f(\xi-\eta)|^2|g(\eta)|^2 \mathrm{d}\eta \mathrm{d}\xi.$$

Change variable, $\xi \to \zeta = \xi - \eta$, and observe that when $|\eta| = |\zeta| = 1$ we have

$$|\xi| = |\eta + \zeta| \simeq (1 + \eta \cdot \zeta)^{1/2},$$
$$(4 - |\xi|^2)^{1/2} = (4 - |\eta + \zeta|^2)^{1/2} \simeq (1 - \eta \cdot \zeta)^{1/2},$$

hence

$$\|B(f,g)\|_{L^2(\mathbb{R}^2)}^2 \lesssim \iint_{\mathbb{S}^1 \times \mathbb{S}^1} \frac{|f(\zeta)|^2|g(\eta)|^2}{\left(1-(\eta\cdot\zeta)^2\right)^{1/2}} \mathrm{d}\sigma_\eta \mathrm{d}\sigma_\zeta. \tag{208}$$

This is an interesting formula. Observe that if the supports of f and g on $\mathbb{S}^1$ are projectionally disjoint, i.e. don't contain points in the same direction, then the quantity $1 - (\eta \cdot \zeta)^2$ is bounded below by a positive constant and in this case we obtain the bilinear restriction estimate

$$\|B(f,g)\|_{L^2(\mathbb{R}^2)} \lesssim \|f\|_{L^2(\mathbb{S}^1)}\|g\|_{L^2(\mathbb{S}^1)}.$$

We can consider also other types of bilinear forms which have a special structure that cancel the singularity in the denominator. Take for example $Q(f, g) = \partial_1 \mathcal{S}f \partial_2 Sg - \partial_2 \mathcal{S}f \partial_1 Sg$, then taking the Fourier transform and proceeding as before we see that

$$\begin{aligned}\|Q(f,g)\|_{L^2(\mathbb{R}^2)}^2 &\lesssim \iint_{\mathbb{S}^1\times\mathbb{S}^1} \frac{|\eta_1\zeta_2 - \eta_2\zeta_1|^2}{\left(1-(\eta\cdot\zeta)^2\right)^{1/2}}|f(\zeta)|^2|g(\eta)|^2 \mathrm{d}\sigma_\eta \mathrm{d}\sigma_\zeta \\ &\lesssim \|f\|_{L^2(\mathbb{S}^1)}^2\|g\|_{L^2(\mathbb{S}^1)}^2,\end{aligned}$$

since we have the identity $|\eta_1\zeta_2 - \eta_2\zeta_1|^2 = 1 - (\eta \cdot \zeta)^2 \leq 1$.

Part 2

Partial Differential Equations

CHAPTER 5

General Equations

It is tempting to define PDE as the subject which is concerned with all partial differential equations, just as Algebraic Geometry, say, deals with all polynomial equations. According to this view, the goal of the subject is to find a general theory of all, or very general classes of PDE's. Though this point of view is quite out of fashion, it has nevertheless important merits which I hope to illustrate below. To see the full power of the general theory we need to, at least, write down general equations, yet will make sure to explain the main ideas in simplified cases. We consider equations, or systems of equations, in $\mathbb{R}^d$ with respect to the variables $x = (x^1, x^2, \dots x^d)$. As before we denote by $\partial_i = \frac{\partial}{\partial x^i}$ the partial derivatives relative to the coordinate x^i and by $\partial^\alpha = \partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_d^{\alpha_d}$ the mixed partial derivatives corresponding to a multi-index $\alpha = (\alpha_1, \alpha_2, \dots \alpha_d) \in \mathbb{N}^d$. We denote by ∂^k the vector of all partial derivatives ∂^α with $|\alpha| = \alpha_1 + \cdots + \alpha_d = k$. Finally we denote by $\Lambda^k u = (u, \partial u, \dots \partial^k u)$ the set of all partial derivatives of order less or equal to k. In most interesting examples k is one or two.

Example. To make these notations more transparent consider the case of $\mathbb{R}^2$ and coordinates x^1, x^2. For the multi-index $\alpha = (2, 0)$ we have $\partial^\alpha u = \frac{\partial}{\partial x^1}\frac{\partial}{\partial x^1} u = \partial_1^2 u$ while for $\alpha = (1, 1)$ we have $\partial^\alpha u = \frac{\partial}{\partial x^1}\frac{\partial}{\partial x^2} u = \partial_1\partial_2 u$. Also

$$\partial^2 u = (\frac{\partial}{\partial x^1}\frac{\partial}{\partial x^1} u, \frac{\partial}{\partial x^1}\frac{\partial}{\partial x^2} u, \frac{\partial}{\partial x^2}\frac{\partial}{\partial x^2} u) = (\partial_1^2\, u, \partial_1\partial_2\, u, \partial_2^2\, u)$$

and $\Lambda^2 u = (u, \partial_1 u, \partial_2 u, \partial_1^2\, u, \partial_1\partial_2\, u, \partial_2^2\, u)$.

With this notation the Laplace operator in $\mathbb{R}^d$ has the form $\Delta = \partial_1^2 + \partial_2^2 + \dots \partial_d^2$ while the D'Alembertian in the Minkowski space $\mathbb{R}^{d+1}$ has the form $\square = -\partial_t^2 + \partial_1^2 + \dots + \partial_d^2$. To make sense of an equation in which there appear partial derivatives of order up to k we need to work with functions which are k-time differentiable at every point. It is convenient to work with the class C^k of functions which are continuous and whose all partial derivatives $\partial^\alpha u$ of order $|\alpha| \le k$ are continuous.

Definition. *A general partial differential equation in $\mathbb{R}^d$ of order k is of the form,*

$$F\big(x, \Lambda^k u(x)\big) = 0, \tag{209}$$

where F is a specified function. We also consider $N \times N$ systems[1] *in which case F and u are column N-vectors. A function u of class C^k is said to be a classical*[2] *solution* (209) *if it verifies the equation as all points x in a specified domain of $\mathbb{R}^d$.*

[1] That is *determined systems* of N equations for N unknowns.

[2] We call it classical to distinguish from generalized solutions to be discussed in the following sections.

Consider first the one dimensional situation $d = 1$ in which case (209) becomes an ordinary differential equation (ODE), or system of ODE. To simplify further take $k = 1$ and $N = 1$, that is the case of an ordinary differential equation of order $k = 1$. Then (209) is simply, $F(x, u(x), \partial_x u(x)) = 0$ where F is a given function of the three variables x, u and $p = \partial_x u$ such as, for example, $F(x, u, p) = x \cdot p + u^3 - sinx$. To solve the equation (209) in this case is to find a function a C^1 function $u(x)$ such that

$$x \cdot \partial_x u(x) + u^3 = sinx. \tag{210}$$

Now consider the case of a second order ODE, i.e. $d = N = 1$ and $k = 2$. Then (209) becomes, $F(x, u(x), \partial_x u(x), \partial_x^2 u(x)) = 0$, where F now depends on the four variables $x, u, p = \partial_x u, q = \partial_x^2 u$. As an example take $F = q^2 + V'(u)$, for some given function $V = V(u)$, in which case (209) becomes the nonlinear harmonic oscillator equation,

$$\partial_x^2 u(x) + V'(u(x)) = 0 \tag{211}$$

Passing to a system of ODE, with $d = 1$, $k = 1$ and $N = 2$ we will need a vector function $F = (F_1, F_2)$ with both F_1 and F_2 depending on the five variables $x, u_1, u_2, p_1 = \partial_x u_1, p_2 = \partial_x u_2$. Then (209) becomes,

$$\begin{aligned} F_1\big(x, u_1(x), u_2(x), \partial_x u_1(x), \partial_x u_2(x)\big) &= 0 \\ F_2\big(x, u_1(x), u_2(x), \partial_x u_1(x), \partial_x u_2(x)\big) &= 0 \end{aligned}$$

The case of PDE gets a bit more complicated because of the large number of variables involved in the definition of F. Thus for first order ($k = 1$) scalar equations (N=1) in two space dimensions ($d = 2$) we need functions F depending on the two spatial variables x^1, x^2 as well as u, $p_1 = \partial_1 u$ and $p_2 = \partial_2 u$. For a given function of five variables $F = F(x, u, p)$, a general first order PDE in two space dimensions takes the form,

$$F(x, u(x), \partial_1 u(x), \partial_2 u(x)) = 0. \tag{212}$$

As a particular example take $F = p_1^2 + p_2^2 - 1$. The corresponding equation is,

$$(\partial_1 u(x))^2 + (\partial_2 u(x))^2 = 1 \tag{213}$$

which plays an important role in geometric optics. A classical solution of the equation is a C^1 function $u = u(x^1, x^2)$ which verifies (213) at all points of a domain $D \subset \mathbb{R}^2$. A similar example is that given by the Eikonal equation (51).

Remark 1. We have excluded from our definition over-determined (i.e. the number of equations exceeds that of unknowns) or underdetermined systems (i.e. the number of equations is less than that of unknowns) despite their obvious interest to Geometry and Physics. The Einstein vacuum equations (24), for example, look underdetermined at first glance. They become determined once we fix a particular coordinate condition, such as the *wave coordinate* condition alluded to in the introduction. Gauge field theories, such as Yang-Mills, have a similar structure. On the other hand the equations defined by the De Rham complex on an open set of $\mathbb{R}^n$ form an overdetermined system. For example, given a one form $\omega = \omega_i dxi$, the system $df = \omega$ is overdetermined and can only be solved, locally, if the exterior derivative of ω vanishes i.e. $d\omega = 0$.

Remark 2. All higher order scalar equations or systems can in fact be re-expressed as first order systems, i.e. $k = 1$, by simply introducing all higher order derivatives of u as unknowns together with the obvious compatibility relations between partial derivatives. As an example consider equation (211) and set $v = \partial_x u$. We can then rewrite the equation as a first order system with $N = 2$, namely $\partial_x v + V'(u) = 0, \partial_x u - v = 0$.

An equation, or system, is called *quasi-linear* if it is linear with respect to the highest order derivatives. A quasilinear system of order one ($k = 1$) in $\mathbb{R}^d$ can be written in the form,

$$\sum_{i=1}^{d} A^i(x, u(x))\partial_i u = F(x, u(x)) \tag{214}$$

Here u and F are column $N-$vectors and the coefficients $A^1, A^2, \dots A^d$ are $N \times N$ matrix valued functions.

The minimal surface equation is an example of a second order ($k = 2$) quasilinear scalar equation ($N = 1$) in two space dimensions. Indeed, using the coordinates x^1, x^2, instead of x, y, we can manipulate (6) with the help of Leibnitz formula and rewrite in the form,

$$\sum_{i,j=1,2} h^{ij}(\partial u)\partial_i \partial_j u = 0, \tag{215}$$

with $h^{11}(\partial u) = 1 + (\partial_2 u)^2$, $h^{22}(\partial u) = 1 + (\partial_1 u)^2, h^{12}(\partial u) = h^{21}(\partial u) = -\partial_1 u \cdot \partial_2 u$, which is manifestly a second order quasi-linear equation.

In the particular case when the top order coefficients of a quasilinear equation, i.e. those corresponding to the highest order derivatives, depend only on the space variables $x \in \mathbb{R}^d$, the equation, or system, is called semi-linear. For example, equation (20) derived in connection to the uniformization theorem, is semi-linear.

A linear equation, or system, of order k can be written in the form,

$$\sum_{|\alpha| \le k} A_\alpha(x)\partial^\alpha u(x) = F(x). \tag{216}$$

Observe that the differential operator on the left hand side is indeed linear in the sense discussed in our introduction. If in addition the coefficients A_α are constant in x, the system is called linear with constant coefficients. The five basic equations (1)–(5) discussed in the introduction are all linear with constant coefficients. Typically, these are the only equations which can be solved explicitly.

We thus have our first useful, indeed very useful, classification of PDE's into fully nonlinear, quasi-linear, semi-linear and linear. A fully nonlinear equation is nonlinear relative to the highest derivatives. The typical example is the Monge Ampere equation. For simplicity consider the case of functions of 2 variables $u(x^1, x^2)$ in $\mathbb{R}^2$ with hessian $\partial^2 u = (\partial_i \partial_j u)_{i,j=1,2}$. Clearly the determinant $\det(\partial^2 u) = (\partial_1^2 u) \cdot (\partial_2^2 u) - (\partial_1 \partial_2 u)^2$, is quadratic with respect to the second derivatives of

u. Thus the Monge -Ampère equation,

$$\det(\partial^2 u) = f(x, u, \partial u), \tag{217}$$

with f a given function defined on $\mathbb{R}^2\times\mathbb{R}\times\mathbb{R}^2$, is fully nonlinear. This equation plays an important role in Geometry, in relation to the isometric embedding problem as well as to the problem of finding surfaces in $\mathbb{R}^2$ with prescribed Gauss curvature. A variant of the Monge Ampère equation, for complex valued functions, plays a central role in complex geometry in connection to Calabi -Yau manifolds. Calabi-Yau manifolds, on the other hand, are central mathematical objects in String Theory.

Remark. Most of the basic equations of Physics, such as the Einstein equations, are quasilinear. Fully nonlinear equations appear however in connection to the theory of characteristics of linear PDE, which we discuss at length below, or in geometry.

1. First order scalar equations

It does not make sense to give a systematic treatment of this classical topic since there are many PDE books which do an excellent job, such as [**E**] or [**J**]. In what follows I will only attempt to give the main ideas behind the theory. It turns out that scalar ($N = 1$) first order ($k = 1$) PDE in d space dimensions can be reduced to systems of first order ODE.

As a simple illustration of this important fact consider the following equation in two space dimensions,

$$a^1(x^1, x^2)\partial_1 u(x^1, x^2) + a^2(x^1, x^2)\partial_2 u(x^1, x^2) = f(x^1, x^2) \tag{218}$$

where a^1, a^2, f are given real functions in the variables $x = (x^1, x^2) \in \mathbb{R}^2$. We associate to (218) the first order 2×2 system

$$\frac{dx^1}{ds}(s) = a^1(x^1(s), x^2(s)), \qquad \frac{dx^2}{ds} = a^2(x^1(s), x^2(s)) \tag{219}$$

To simplify matters we assume $f = 0$. Observe that any solution $u = u(x^1, x^2)$ of (218), with $f = 0$, is constant along any solution $x(s) = (x^1(s), x^2(s))$, i.e.

$$\frac{d}{ds}u\big(x^1(s), x^2(s)\big) = 0.$$

Thus, in principle, the knowledge of solutions to (219), which are called *characteristic curves* for (218), allows us to find all solutions to (218). I say in principle because, in general, the nonlinear system (219) is not so easy to solve. Yet ODE are simpler to deal with and the fundamental theorem of ODE, which we will discuss later in this section, allows us to solve (219), at least locally for a small interval in s. The constancy of u along characteristic curves allows us to obtain, even when we cannot find explicit solutions, important qualitative information. For example, suppose that the coefficients a^1, a^2 are smooth (or real analytic) and that the initial data is smooth (or real analytic) everywhere on $\mathcal{H}$ except at some point $x_0 \in \mathcal{H}$ where it is discontinuous. Then, clearly, as long as the trajectories of (219) are well defined and distinct, the solution u remains smooth (or real analytic) at all points except along the characteristic curve Γ which initiates at x_0, i.e. along the solution to (219) which verifies the initial condition $x(0) = x_0$. The discontinuity at x_0 propagates precisely along Γ. We see here the simplest manifestation of a general principle, which we shall state later, that singularities of solutions to PDE propagate along characteristics.

One can generalize equation (218) to allow the coefficients a_1, a_2 and f to depend not only on $x = (x^1, x^2)$ but also on u,

$$a^1(x, u(x))\partial_1 u(x) + a^2(x, u(x))\partial_2 u(x) = f(x, u(x)) \tag{220}$$

The associated *characteristic system* becomes,

$$\frac{dx^1}{ds}(s) = a^1\big(x(s), u(s, x(s))\big), \qquad \frac{dx^2}{ds} = a^2\big(x(s), u(s, x(s))\big) \tag{221}$$

Then, as in the previous case,

$$\frac{d}{ds}u(x^1(s), x^2(s)) = f(x^1(s), x^2(s), u(x^1(s), x^2(s))) \tag{222}$$

Unlike the previous case however (221) is undetermined; we need now to consider the enlarged ODE system (221)-(222). where the unknowns are $x^1(s), x^2(s), u(s) = u(x^1(s), x^2(s))$. As a special example of (220) consider the scalar equation in two space dimensions,

$$\partial_t u + u\partial_x u = 0, \qquad u(0, x) = u_0(x) \tag{223}$$

called the Burger equation. Since $a^1 = 1, a^2 = u$ we can set $x^1(s) = s$, $x^2(s) = x(s)$ in (221) and thus derive its *characteristic equation* in the form,

$$\frac{dx}{ds}(s) = u(s, x(s)). \tag{224}$$

Observe that, for any given solution u of (223) and any characteristic curve $(s, x(s))$ we have $\frac{d}{ds}u(s, x(s)) = 0$. Thus, in principle, the knowledge of solutions to (224) would allow us to determine the solutions to (223). This, however, seems circular since u itself appears in (224). To see how this difficulty can be circumvented consider the initial value problem for (223), i.e. look for solutions u which verify $u(0, x) = u_0(x)$. Consider an associated characteristic curve $x(s)$ such that, initially, $x(0) = x_0$. Then, since u is constant along the curve, we must have $u(s, x(s)) = u_0(x_0)$. Hence, going back to (224), we infer that $\frac{dx}{ds} = u_0(x_0)$ and thus $x(s) = x_0 + su_0(x_0)$. We thus deduce that,

$$u(s, x_0 + su_0(x_0)) = u_0(x_0) \tag{225}$$

which gives us, implicitly, the form of the solution u. We see once more, from (225), that if the initial data is smooth (or real analytic) everywhere except at a point x_0, of the line $t = 0$, then the corresponding solution is also smooth (or real analytic) everywhere, in a small neighborhood V of x_0, except along the characteristic curve which initiates at x_0. The smallness of V is necessary here because new singularities can form in the large. Observe indeed that u has to be constant along the lines $x + su_0(x)$ whose slopes depend on $u_0(x)$. At a point when these lines cross, we would obtain different values of u which is impossible unless u *becomes singular at that point. In fact one can show that the first derivative u_x becomes infinite at the first singular point*, i.e. the singular point with the smallest value of $|t|$. This blow-up phenomenon occur for any smooth, non-constant, initial data u_0.

Remark. There is an important difference between the linear equation (218) and quasi-linear equation (220). The characteristics of the first depend only on the coefficients $a^1(x), a^2(x)$ while the characteristics of the second depend, explicitely, on a particular solution u of the equation. In both cases, *singularities can only propagate along the characteristic curves of the equation.* For nonlinear equations, however, new singularities can form in the large, independent of the smoothness of the data.

The above procedure extends to fully nonlinear scalar equations in $\mathbb{R}^d$ of the form,

$$\partial_t u + H(x, \partial u) = 0, \qquad u(0,x) = u_0(x) \tag{226}$$

with $H = H(x,p)$ a given function of the variables $x = (x^1, x^2, \dots, x^d)$ and $p = (p_1, p_2, \dots p_d)$, called the *Hamiltonian* of the system, and $\partial u = (\partial_1 u, \partial_2 u, \dots, \partial_d u)$. We associate to (226) the ODE system, with $i = 1, 2 \dots, d$,

$$\frac{dx^i}{dt} = \frac{\partial}{\partial p_i} H(x(t), p(t)), \qquad \frac{dp_i}{dt} = -\frac{\partial}{\partial x^i} H(x(t), p(t)). \tag{227}$$

The equation (226) is called a Hamilton-Jacobi equation while (227) is known as a Hamiltonian system of ODE. The relationship between them is a little more involved than in the previous cases discussed above. To simplify the calculations below we assume $d = 1$, so that $H = H(x,p)$ is only a function of two variables. Let u be a solution of (226). Differentiating (226) in x and applying the chain rule we derive,

$$\partial_t \partial_x u + \partial_p H(x, \partial_x u) \partial_x^2 u = -\partial_x H(x, \partial_x u) \tag{228}$$

Now take $x(t)$ a solution of the equation $\frac{dx}{dt} = \partial_p H(x(t), \partial_x u(x(t))$ and set $p(t) := \partial_x u(t, x(t))$. Then, by using first the chain rule and then equation (228) we derive,

$$\begin{aligned}\frac{dp}{dt} &= \partial_x \partial_t u(t, x(t)) + \partial_x^2 u(t, x(t)) \partial_p H(x(t), p(t)) \\ &= -\partial_x H(x(t), \partial_x u(t, x(t))) = -\partial_x H(x(t), p(t))\end{aligned}$$

Hence $x(t), p(t)$ verify the Hamilton equation

$$\frac{dx}{dt} = \partial_p H(x(t), p(t)), \qquad \frac{dp}{dt} = -\partial_x H(x(t), p(t)).$$

On the other hand, $\frac{d}{dt} u(t, x(t)) = \partial_t u(t, x(t)) + \partial_x u(t, x(t)) \partial_p H(x(t), p(t))$, and, using equation (226), $\partial_t u(t, x(t)) = -H(x(t), \partial_x u(t, x(t)) = -H(x(t), p(t))$. Thus,

$$\frac{d}{dt} u(t, x(t)) = -H(x(t), p(t)) + p(t) \partial_p H(x(t), p(t)),$$

from which we see, in principle, how to construct u based only on the knowledge of the solutions $x(t), p(t)$, called the *bicharacteristic curves* of the nonlinear PDE. Once more singularities can only propagate along bichararcteristics. As in the case of the Burger equation singularities will occur, for essentially, all smooth data; thus a classical, i.e. continuously differentiable, solution can only be constructed locally in time. Both Hamilton-Jacobi equation and hamiltonian systems play a fundamental role in Classical Mechanics as well as in the theory of propagation of singularities in linear PDE. The deep connection between hamiltonian systems and first oder Hamilton-Jacobi equations have played an important role in the introduction of the Schrödinger equation in quantum mechanics.

2. Initial Value Problem for ODE

To go further with our general presentation we need to discuss the initial value problem. For simplicity let us start with a first order ODE

$$\partial_x u(x) = f(x, u(x)) \tag{229}$$

subject to the initial condition

$$u(x_0) = u_0 \tag{230}$$

The reader may assume, for simplicity, that (229) is a scalar equation and that f is a nice function of x and u, such as $f(x,u) = u^3 - u + 1 + sinx$. Observe that the knowledge of the initial data u_0 allows us to determine $\partial_x u(x_0)$. Differentiating the equation (229) with respect to x and applying the chain rule, we derive,

$$\partial_x^2 u(x) = \partial_x f(x, u(x)) + \partial_u f(x, u(x))\partial_x u(x) = \cos x + 3u^2(x)\partial_x u(x) - \partial_x u(x)$$

Hence, $\partial_x^2 u(x_0) = \partial_x f(x_0, u_0) + \partial_u f(x_0, u_0)\partial_x u_0$ and since $\partial_x u(x_0)$ has already been determined we infer that $\partial_x^2 u(x_0)$ can be explicitely calculated from the initial data u_0. The calculation also involves the function f as well as its first partial derivatives. Taking higher derivatives of the equation (229) we can recursively determine $\partial_x^3 u(x_0)$, as well as all other higher derivatives of u at x_0. One can than, in principle, determine $u(x)$ with the help of the Taylor series $u(x) = \sum_{k\ge 0} \frac{1}{k!}\partial_x^k u(x_0)(x-x_0)^k = u(x_0)+\partial_x u(x_0)(x-x_0)+\frac{1}{2!}\partial_x^2(x_0)(x-x_0)^2+\dots$. We say in principle because there is no guarantee that the series converge. There is however a very important theorem, called the *Cauchy-Kowalewski theorem*, which asserts that, if the function f is real analytic, which is certainly the case for our $f(x,u) = u^3 - u + 1 + sinx$, then there exists a neighborhood J of x_0 where the Taylor series converge to a real analytic solution u of the equation. One can the easily show that the solution such obtained is the unique solution to (229) subject to the initial condition (230).

The same result may not hold true if we consider a more general equation of the form,

$$a(x, u(x))\partial_x u = f(x, u(x)), \qquad u(x_0) = u_0 \tag{231}$$

Indeed the recursive argument outlined above breaks down in the case of the scalar equation $(x-x_0)\partial_x u = f(x,u)$ for the simple reason that we cannot even determine $\partial_x u(x_0)$ from the initial condition $u(x_0) = u_0$. A similar problem occurs for the equation $(u - u_0)\partial_x u = f(x,u)$. An obvious condition which allows us to extend our previous recursive argument to (231) is that $a(x_0,u_0) \neq 0$. Otherwise we say that the initial value problem (231) is *characteristic*. If both a and f are also real analytic the Cauchy-Kowalewski theorem applies and we obtain a unique, real analytic, solution of (231) in a small neighborhood of x_0. In the case of a $N \times N$ system,

$$A(x, u(x))\partial_x u = F(x, u(x)), \qquad u(x_0) = u_0 \tag{232}$$

$A = A(x,u)$ is $N \times N$ matrix and the *non-characteristic condition* becomes

$$\det A(x_0, u_0) \neq 0. \tag{233}$$

It turns out, and this is extremely important, that while the non-degeneracy condition (233) is essential to obtain a unique solution of the equation, the analyticity condition is not at all important, in the case of ODE. It can be replaced by a simple *local Lipschitz condition* for A and F, i.e. it suffices to assume, for example, that only their first partial derivatives exist and that they are merely locally bounded. This is always the case if the first derivatives of A, F are continuous.

The following local existence and uniqueness (LEU) theorem is called the fundamental theorem of ODE.

Theorem[Fundamental theorem for ODE] *If the matrix $A(x_0, u_0)$ is invertible and if A, F are continuous and have locally bounded first derivatives then there exists a time interval $x_0 \in J \subset \mathbb{R}$ and a unique solution*[3] *u defined on J verifying the initial conditions $u(x_0) = u_0$.*

Proof The proof of the theorem is based on the *Picard iteration method*. The idea is to construct a sequence of approximate solutions $u_{(n)}(x)$ which converge to the desired solution. Without loss of generality we can assume A to be the identity matrix[4]. One starts by setting $u_{(0)}(x) = u_0$ and then defines recursively,

$$\partial_x u_{(n)}(x) = F(x, u_{(n-1)}(x)), \qquad u_{(n-1)}(x_0) = u_0 \tag{234}$$

Observe that at every stage we only need to solve a very simple linear problem, which makes Picard iteration easy to implement numerically. As we shall see below, variations of this method are also used for solving nonlinear PDE.

...... To fill in the proof..... ∎

Remark. The local existence theorem is sharp, in general. Indeed we have seen that the invertibility condition for $A(x_0, u_0)$ is necessary. Also, in general, the interval of existence J may not be extended to the whole real line. As an example consider the nonlinear equation $\partial_x u = u^2$ with initial data $u = u_0$ at $x = 0$, for which the solution $u = \frac{u_0}{1 - x u_0}$ becomes infinite in finite time, i.e. it *blows-up*.

Once the LEU result is established one can define the main goals of the mathematical theory of ODE to be:

(1) Find criteria for global existence. In case of blow-up describe the limiting behavior.
(2) In case of global existence describe the asymptotic behavior of solutions and family of solutions.

Though is impossible to develop a general theory, answering both goals (in practice one is forced to restrict to special classes of equations motivated by applications), the general LEU theorem mentioned above gives a powerful unifying theme. It

---

[3] Since we are not assuming analyticity for A, F the solution may not be analytic, but it has continuous first derivatives.

[4] since A is invertible we can multiply both sides of the equation by the inverse matrix A^{-1}

would be very helpful, really wonderful, if a similar situation were to hold for general PDE.

3. Initial value problem for PDE

By analogy to the one dimensional situation it is natural to consider, instead of points, hyper-surfaces $\mathcal{H} \subset \mathbb{R}^d$ on which to specify initial conditions for u. For a general equation of order k, i.e. involving k derivatives, we would need to specify the values of u and its first $k-1$ normal derivatives[5] to $\mathcal{H}$. For example in the case of the second order wave equation (3) we need to specify the initial data for u and $\partial_t u$. along the hypersurface $t = 0$. Without getting into details at this point we can give the following general definition.

Definition. *We say that an initial value problem, for a k-order quasilinear system, in which we specify, as data, the first $k-1$ normal derivatives of a solution u along $\mathcal{H}$, is non-characteristic at a point x_0 of $\mathcal{H}$, if we can formally determine all other higher partial derivatives of u at x_0, uniquely, in terms of the data.*

To understand the definition, which may seem too general at this point, consider the much simpler case $k = 1$, $N = 1$. In this case we only need to specify the restriction $u|_{\mathcal{H}} = u_0$ of u to $\mathcal{H}$. Our initial value problem takes the form,

$$\sum_{i=1}^{d} a^i(x, u(x))\partial_i u(x) = f(x, u(x)), \qquad u|_{\mathcal{H}} = u_0 \tag{235}$$

with a^i, f real valued functions of $x \in \mathbb{R}^d$ and $u \in \mathbb{R}$. To simplify further take $d = 2$, i.e. we have the equation in $x = (x^1, x^2)$,

$$a^1(x, u(x))\partial_1 u(x) + a^2(x, u(x))\partial_2 u(x) = f(x, u(x)) \tag{236}$$

we have encountered earlier in (220). Consider a curve $\mathcal{H}$ in $\mathbb{R}^2$, parametrized by $x^1 = x^1(s), x^2 = x^2(s)$ whose tangent vector $V(s) = (\frac{dx^1}{ds}, \frac{dx^2}{ds})$ is non-degenerate, i.e. $|V(s)| = (|\frac{dx^1}{ds}|^2 + |\frac{dx^2}{ds}|^2)^{1/2} \neq 0$. It has a well defined unit normal $N(s) = (n_1(s), n_2(s))$, which verifies the conditions,

$$N(s) \cdot V(s) = 0, \qquad N(s) \cdot N(s) = 1$$

Observe that the coefficients a^1, a^2 in (236) can be completely determined, along $\mathcal{H}$, from the knowledge of the initial condition $u_0 = u_0(s)$. Consider the first derivatives $(\partial_1 u, \partial_2 u)$ evaluated along $\mathcal{H}$, i.e. $U(s) = \big(\partial_1 u(x(s)), \partial_2 u(x(s))\big)$. At every point along $\mathcal{H}$ our equation reads,

$$A(s) \cdot U(s) = f(s), \tag{237}$$

where $A(s) = \big(a^1(x(s), u_0(s)), a^2(x(s), u_0(s)\big)$ and $f(s) = f\big(x(s), u_0(s)\big)$ are completely determined by the data $u_0(s)$. Differentiating $u(x(s)) = u_0(s)$ with respect to s we infer that,

$$U(s) \cdot V(s) = U_0(s), \qquad U_0(s) = \frac{d}{ds} u_0(s).$$

[5]These are derivatives in the direction of the normal to $\mathcal{H}$.

To fully determine $U(s)$ it remains to determine its projection on the normal vector $N(s)$, i.e. $U(s) \cdot N(s)$. Indeed, since $V(x)$ and $N(x)$ span $\mathbb{R}^2$, at all points $x = (x^1(s), x^2(s))$ along our curve, we have

$$U(s) = (U \cdot V)(s) \frac{V(s)}{|V(s)|^2} + (U \cdot N)(s)N(s) \tag{238}$$

Therefore, from the equation (237),

$$f(s) = A(s) \cdot U(s) = (U(s) \cdot V(s)) \frac{A(s) \cdot V(s)}{|V(s)|^2} + (U(s) \cdot N(s))\, A(s) \cdot N(s)$$

from which we can determine $U(s) \cdot N(s)$ provided that,

$$A(s) \cdot N(s) \neq 0. \tag{239}$$

If, on the other hand, $A(s) \cdot N(s) = 0$ then, since $V(s) \cdot N(s) = 0$, we infer that the vectors $A(s)$ and $V(s) = \frac{dx}{ds}$ must be proportional, i.e. $\frac{dx}{ds} = \lambda(s)A(s)$. One can then reparametrize the curve $\mathcal{H}$, i.e. introduce another parameter $s' = s'(s)$ with $\frac{ds'}{ds} = \lambda(s)$, such that relative to the new parameter we have $\lambda = 1$. This leads to the equation,

$$\frac{dx^1}{ds} = a^1\big(x(s), u(x(s))\big), \qquad \frac{dx^2}{ds} = a^2\big(x(s), u(x(s))\big)$$

which is precisely the characteristic system (221). Thus,

Along a characteristic curve, the equation (236) *is degenerate, that is we cannot determine the first order derivatives of u uniquely in terms of the data u_0. On the other hand the non-degenerate condition,*

$$A(s_0) \cdot N(s_0) \neq 0, \quad \textit{i.e.} \quad a^1(x_0, u(x_0))n_1(x_0) + a_2(x_0, u(x_0))n_2(x_0) \neq 0 \tag{240}$$

at some point $x_0 = x(s_0) \in \mathcal{H}$, allows us to determine all higher derivatives of u at x_0, uniquely in terms of the data u_0.

Indeed, if the condition $A(s_0) \cdot N(s_0) \neq 0$ is satisfied at $x_0 = x(s_0) \in \mathcal{H}$ we have seen already how to determine the first derivatives $\partial_1 u, \partial_2 u$ at that point. Once we have these it is not difficult to determine all higher derivatives of u. For example, observe, by differentiating equation (236) with respect to x^1, that the function $v = \partial_1 u$ verifies an equation of the form,

$$a^1(x, u(x))\partial_1 v(x) + a^2(x, u(x))\partial_2 v = g(x, u(x), v(x))$$

with a function g which can be easily determined from the coefficients a and f. We can now proceed as before and determine the first derivatives of v i.e. $\partial_1^2 u, \partial_2\partial_1 u$. Thus, recursively, we can determine all partial derivatives of u of any order.

We can easily extend the discussion above to the higher dimensional case (235). Given a hypersurface $\mathcal{H}$ in $\mathbb{R}^d$, with unit normal $N = (n_1, n_2, \ldots n_d)$, we find that $\mathcal{H}$ is non-characteristic at x_0 for the initial value problem (235) if,

$$\sum_{i=1}^{n} a^i(x_0, u_0(x_0))n_i(x_0) \neq 0 \tag{241}$$

With a little more work we can extend our discussion to general higher order quasi-linear equations, or systems and get a simple, sufficient condition, for a Cauchy problem to be non-characteristic. Particularly important for us are second order ($k = 2$) scalar equations ($N = 1$). To keep things simple consider the case of a second order, semi-linear equation in $\mathbb{R}^d$,

$$\sum_{i,j=1}^{d} a^{ij}(x)\partial_i\partial_j u = f(x, u(x), \partial u(x)) \tag{242}$$

and a hypersurface $\mathcal{H}$ in $\mathbb{R}^d$ defined by the equation $\psi(x) = 0$ with non-vanishing gradient $\partial\psi$. Define the unit normal at a point $x_0 \in \mathcal{H}$ to be $N = \frac{\partial\psi}{|\partial\psi|}$, or in components $n_i = \frac{\partial_i\psi}{|\partial\psi|}$. As initial conditions for (242) we prescribe u and its normal derivative $Nu(x) = n_1(x)\partial_1 u(x) + n_2(x)\partial_2 u(x) + \ldots n_d(x)\partial_d u(x)$ on $\mathcal{H}$,

$$u(x) = u_0(x), \quad Nu(x) = u_1(x), \qquad x \in \mathcal{H} \tag{243}$$

We need to find a condition on $\mathcal{H}$ such that we can determine all higher derivatives of a solution u, at $x_0 \in \mathcal{H}$, from the initial data u_0, u_1. We can proceed exactly in the same manner as before, and find that all second order derivatives of u can be determined at a point $x_0 \in \mathcal{H}$, provided that,

$$\sum_{i,j=1}^{d} a^{ij}(x_0)n_i(x_0)n_j(x_0) \neq 0 \tag{244}$$

It is indeed easy to see that the only second order derivative of u, which is not automatically determined from u_0, u_1, is of the form $N^2u(x_0) = N(N(u))(x_0)$. This latter can be determined from the equation (242), provided that (244) is verified. One does this by decomposing all partial derivatives of u into tangential and normal components, as we have done in (238). One can then show, recursively, that all higher derivatives of u can also be determined. Thus, (244) is exactly the *non-characteristic* condition we were looking for.

If, on the other hand, $\sum_{i,j=1}^{d} a^{ij}(x)n_i(x)n_j(x) = 0$ at all points we call $\mathcal{H}$ *a characteristic hypersurface* for the equation (242). Since $n_i = \frac{\partial\psi}{|\partial_i\psi|}$ we find that $\mathcal{H}$ is characteristic if and only if,

$$\sum_{i,j=1}^{d} a^{ij}(x)\partial_i\psi(x)\partial_j\psi(x) = 0 \tag{245}$$

Remark Observe that only the left hand side[6] of (242), called [7], is relevant in determining the characteristic surfaces of the equation.

Example 1. Assume that the coefficients a of (242) verify,

$$\sum_{i,j=1}^{d} a^{ij}(x)\xi_i\xi_j > 0, \quad \forall\, \xi \in \mathbb{R}^d, \quad \forall x \in \mathbb{R}^d \tag{246}$$

Then no surface in $\mathbb{R}^d$ can be characteristic. This is the case, in particular, for the equation $\Delta u = f$. Consider also the minimal surfaces equation written in the form

[6] containing second order derivatives

[7] principal part

(215). It is easy to check that, the quadratic form associated to the symmetric matrix $h^{ij}(\partial u)$ is positive definite independent of ∂u. Indeed,

$$h^{ij}(\partial u)\xi_i\xi_j = (1+|\partial u|^2)^{-1/2}\big(|\xi|^2 - (1+|\partial u|^2)^{-1}(\xi\cdot\partial u)^2\big) > 0$$

Thus, even though (215) is not linear, we see that all surfaces in $\mathbb{R}^2$ are non-characteristic.

Example 2. Consider the wave equation $\Box u = f$ in $\mathbb{R}^{1+d}$. All hypersurfaces of the form $\psi(t,x) = 0$ for which,

$$(\partial_t\psi)^2 = \sum_{i=1}^{d}(\partial_i\psi)^2, \tag{247}$$

are characteristic. This is the same eikonal equation which has appeared before in (51). Observe that it splits into two Hamilton-Jacobi equations, see (226),

$$\partial_t\psi = \pm\Big(\ \sum_{i=1}^{d}(\partial_i\psi)^2\ \Big)^{1/2} \tag{248}$$

The bicharacteristic curves of the associated Hamiltonians are called bicharacteristic curves of the wave equation. As particular solutions of (439) we find, $\psi_+(t,x) = (t-t_0) + |x-x_0|$ and $\psi_-(t,x) = (t-t_0) - |x-x_0|$ whose level surfaces $\psi_\pm = 0$ correspond to forward and backward light cones with vertex at $p = (t_0, x_0)$. These represent, physically, the union of *all light rays emanating from a point source at* p. The light rays are given by the equation $(t-t_0)\omega = (x-x_0)$, for $\omega \in \mathbb{R}^3$ with $|\omega| = 1$, and are precisely the (t,x) components of the bicharacteristic curves of the Hamilton-Jacobi equations (248).

More general, consider the linear wave equation,

$$g^{\alpha\beta}\partial_\alpha\partial_\beta\phi = 0. \tag{249}$$

where $g^{\alpha\beta}$ is the inverse of a general Lorentz metric[8] $g_{\alpha\beta}$. The characteristic surfaces of (249) are given by the eikonal equation,

$$g^{\alpha\beta}\partial_\alpha\psi\partial_\beta\psi = 0. \tag{250}$$

They are also called *null hypersurfaces* for the metric g.

Remark. In the particular case when $g_{01} = \ldots = g_{0n} = 0$, $g_{00} < 0$ and g_{ij} positive definite, (249) takes the form,

$$-a^{00}(t,x)\partial_t^2 u + \sum_{i,j} a^{ij}(t,x)\partial_i\partial_j u = 0, \tag{251}$$

where with $a_{00} = -g_{00}^{-1}$ and $a^{ij}g_{jk} = \delta^i_k$. Thus the characteristic equation has the form $-a^{00}(t,x)(\partial_t\psi)^2 + a^{ij}(x)\partial_i\psi\partial_j\psi = 0$ or,

$$\partial_t\psi = \pm\big((a^{00})^{-1}\sum_{i,j} a^{ij}(x)\partial_i\psi\partial_j\psi\big)^{1/2}. \tag{252}$$

[8] i.e. $g_{\alpha\beta} = g_{\beta\alpha}$ and the associated quadratic form, $g_{\alpha\beta}(p)X^\alpha X^\beta$ is non-degenerate at every point p and has signature $(-1, 1, \ldots, 1)$.

Thus, through any point $p \in \mathbb{R}^{1+n}$ pass two distinct characteristic surfaces. The same is true for the general case.

The bicharacteristics of the corresponding hamiltonian systems are called *bicharacteristic curves* of (251).

Remark. In the case of the first order scalar equations (218) we have seen how the knowledge of characteristics can be used to find, implicitly, the general solutions. We have shown, in particular, that singularities propagate only along characteristics. In the case of second order equations the situation is more complicated. the characteristics are typically[9] not sufficient to solve the equations, but they continue to provide important information, such as propagation of singularities. For example, in the case of the wave equation $\square u = 0$ with smooth initial data u_0, u_1 everywhere except at a point $p = (t_0, x_0)$, the solution u has singularities present at all points of the light cone $-(t-t_0)^2 + |x-x_0|^2 = 0$ with vertex at p. A more refined version of this fact shows that the singularities propagate along bicharacteristics. The general principle here is that *singularities propagate along characteristic hypersurfaces of a PDE.* Since this is a very important principle it pays to give it a more precise formulation which extends to general boundary conditions, such as the Dirichlet condition for (**??**).

Propagation of singularities[10]. *If the boundary conditions, or the coefficients of a linear PDE with smooth (or real analytic) coefficients are singular at some point p, and smooth (or real analytic) away from p in some small neighborhood V, then a solution of the equation may only be singular in V along a characteristic hypersurface passing through p. If there are no such characteristic hypersurfaces, any solution of the equation must be smooth (or real analytic) in $V \setminus \{p\}$.*

Remark 1. The principle as stated is far too general, it can be proved only if specific assumptions are made on the symbol of the operator. It should be viewed however as something one might expect for a reasonable equation.

Remark 2. The principle can be extended, under specific minimum regularity assumptions on solutions, to the nonlinear case. It is however invalid in the large. Indeed, as we have shown in in the case of the Burger equation, solutions to nonlinear evolution equations, can develop new singularities independent of the smoothness of the initial conditions. Global versions of the principle can be formulated for linear equations, based on the bicharacteristics of the equation, see remark 3 below.

Remark 3. According to the principle it follows that any solution of the equation $\Delta u = f$, verifying the boundary condition $u|_{\partial D} = u_0$, with a boundary value u_0 which is merely continuous, has to be smooth everywhere in the interior of D

[9] Characteristics enter however in the explicit form of the fundamental solution for the standard wave equation. This was made particularly obvious in the derivation starting with the ansatz (50). The also play a major role to construct approximate solutions for wave equations with variable coefficients, such as (251)

[10] A more precise version of the principle relates propagation of singularities to bicharacteristics curves.

provided that f itself is smooth there. Moreover the solution is real analytic, if f is real analytic.

Remark 4. More precise versions of this principle, which plays a fundamental role in the general theory, can be given for linear equations. In the case of the general wave equation (251), for example, one can show that singularities propagate along bicharacteristics. These are the bicharacteristic curves associated to the Hamilton-Jacobi equation (252).

4. Cauchy-Kowalevsky Theorem

In the case of ODE we have seen that a non-characteristic initial value problem admits always local in time solutions. Is there also a higher dimensional analogue of this fact ? The answer is yes provided that we restrict ourselves to an extension of the Cauchy -Kowalewsky theorem. More precisely one can consider general quasilinear equations, or systems, with real analytic coefficients, real analytic hypersurfaces $\mathcal{H}$, and real analytic initial data on $\mathcal{H}$.

Theorem[Cauchy-Kowalevsky (CK)] *If all the real analyticity conditions made above are satisfied and if $\mathcal{H}$ is non-characteristic at x_0*[11]*, there exists locally, in a neighborhood of x_0, a unique real analytic solution $u(x)$ verifying the system and the corresponding initial conditions.*

The CK theorem validates the most straightforward attempts to find solutions by formal expansions $u(x) = \sum_\alpha C_\alpha (x - x_0)^\alpha$ with constants C_α which can be determined recursively, by simply algebraic formulas, from the equation and initial conditions on $\mathcal{H}$, using only the non-characteristic condition and the analyticity assumptions. Indeed the theorem insures that the naive expansion obtained in this way converges in a small neighborhood of $x_0 \in \mathcal{H}$.

Proof See [**E**] or [**J**] ∎

In the special case of linear equations (216) an important companion theorem, due to Holmgren, asserts that the analytic solution given by the CK theorem is unique in the class of all smooth solutions and smooth non-characteristic hypersurfaces $\mathcal{H}$.

Theorem 4.1 (Holmgren uniqueness theorem)**.** *Consider the initial value problem for a linear equations of the type* (216) *with analytic coefficients. If the the hypersurface $\mathcal{H}$ is also analytic and non-characteristic at $x_0 \in \mathcal{H}$, then the corresponding Cauchy problem is unique in the class of smooth solutions, in a small neighborhood of x_0.*

Proof See [**J**] ∎

[11] In the case of second order equations of type (242) this is precisely condition (244).

Remark. The remarkable thing about Holmgren's theorem is that it proves uniqueness even in cases where existence of solutions cannot be guaranteed. Thus, as we shall see below, the Cauchy problem for the wave equation with data on the hyperplane $x^1 = 0$ does not, in general, have solutions, yet Holmgren's theorem asserts that if a solution exists it must be unique.

At first glance it may seem that the CK theorem is a perfect analogue of the fundamental theorem for ODE's. It turns out, however, that the analyticity conditions required by the CK theorem are much too restrictive and thus the apparent generality of the result is misleading. A first limitation becomes immediately obvious when we consider the wave equation $\Box u = 0$ whose fundamental feature of finite speed of propagation[12] is impossible to make sense in the class of real analytic solutions. A related problem, first pointed out by Hadamard, concerns the impossibility of solving the Cauchy problem, in many important cases, for arbitrary smooth, non analytic, data. Consider, for example, the Laplace equation $\Delta u = 0$ in $\mathbb{R}^d$. As we have established above, any hyper-surface $\mathcal{H}$ is non-characteristic, yet the Cauchy problem $u|_{\mathcal{H}} = u_0$, $N(u)|_{\mathcal{H}} = u_1$, for arbitrary smooth initial conditions u_0, u_1 may admit no local solutions, in a neighborhood of any point of $\mathcal{H}$. Indeed take $\mathcal{H}$ to be the hyperplane $x_1 = 0$ and assume that the Cauchy problem can be solved, for a given, non analytic, smooth data in an domain which includes a closed ball B centered at the origin. The corresponding solution can also be interpreted as the solution to the Dirichlet problem in B, with the values of u prescribed on the boundary ∂B. But this, according to our heuristic principle[13], must be real analytic everywhere in the interior of B, contradicting our initial data assumptions.

On the other hand the Cauchy problem, for the wave equation $\Box u = 0$ in $\mathbb{R}^{d+1}$, has a unique solution for any smooth initial data u_0, u_1, prescribed on a *space-like hyper-surface*, that is a hypersurface $\psi(t, x) = 0$ whose normal vector, at every point $p = (t_0, x_0)$, is directed inside the interior of the future or past directed light cone passing through that point. Formally this means,

$$|\partial_t \psi(p)| > \big(\sum_{i=1}^{d} |\partial_i \psi(p)|^2\big)^{1/2}. \tag{253}$$

The condition is clearly satisfied by the hypersurfaces of $t = t_0$, but any other hypersurface close to it is also spacelike. On the other hand the IVP is *ill posed*, i.e. not well posed, for a time-like hypersurface, i.e a hypersurface for which,

$$|\partial_t \psi(p)| < \big(\sum_{i=1}^{d} |\partial_i \psi(p)|^2\big)^{1/2}. \tag{254}$$

In this case we cannot, for general non real analytic initial conditions, find a solution of the IVP. An example of a time-like hypersurface is given by the hyperplane $x^1 = 0$.

[12] Roughly this means that if a solution u is compactly supported at some value of t it must be compactly supported at all later times. Analytic functions cannot be compactly supported without vanishing identically.

[13] which can be easily made rigorous in this case

Definition. *A given problem for a PDE is said to be well posed if both existence and uniqueness of solutions can be established for arbitrary data which belong to a specified large space of functions, which includes the class of smooth functions*[14]. *Moreover the solutions must depend continuously on the data.*

The continuous dependence on the data is very important. Indeed solutions to the IVP for a PDE would be of little use if very small changes of the initial conditions will result, instantaneously, in very large changes in the corresponding solutions. It is only in the class of smooth solutions that the theory of PDE becomes really interesting, relevant and challenging. It means that we have to give up hope for a all encompassing result and look instead for special classes of equations which have common features, or really just on special important equations. It is in that sense that the generality of the CK theorem is really an illusion. The true study of partial differential equations only begins when we give up on analyticity.

Exercise 1. Use the energy method to prove uniqueness of the initial value problem for $\Box\phi = 0$ for smooth initial data prescribed on a space-like hypersurface. Can you prove existence, say in $\mathbb{R}^{1+3}$?

Exercise 2 Show, using the Fourier transform, that the IVP for data prescribed on the time-like hypersurface $x^1 = 0$ is ill posed.

5. Standard classification

The different behavior of the Laplace and Wave equations mentioned above illustrates the fundamental difference between ODE and PDE and the illusory generality of the CK theorem. Given that the Laplace and wave equation are so important in geometric and physical application one is interested to find the broadest classes of equations with which they share their main properties. The equations modeled by the Laplace equation are called *elliptic* while those modeled by the wave equation are called *hyperbolic*. The other two important models are the the heat, see (179), and Schrödinger equation, see (180). The general classes of equations with which they resemble are called *parabolic* and, respectively, *dispersive*.

Elliptic equations are the most robust and easiest to characterize, they admit no characteristic hypersurfaces.

Definition 1: *A linear, or quasi-linear, $N \times N$ system with no characteristic hyper-surfaces is called elliptic.*

Clearly the equations of type (242) whose coefficients a^{ij} verify condition (246) are elliptic. The minimal surface equation (6) is also elliptic. It is also easy to verify that the Cauchy-Riemann system (15) is elliptic. As it was pointed out by Hadamard, the initial value problem is not well posed for elliptic equations. The natural way of parametrizing the set of solutions to an elliptic PDE is to prescribe conditions for

[14] Here we are necessarily vague. A precise space can be specified in each specific case.

u, and some of its derivatives[15] , at the boundary of a domain $D \subset \mathbb{R}^n$. These are called *boundary value problems (BVP).* A typical example is the Dirichlet boundary condition $u|_{\partial D} = u_0$ for the Laplace equations $\Delta u = 0$ in a domain $D \subset \mathbb{R}^n$. One can show that, under mild regularity assumptions on the domain D and continuous boundary value u_0, this problem admits a unique solution, depending continuously on u_0. We say that the Dirichlet problem for the Laplace equation is well posed. Another well posed problem for the Laplace equation is given by the Neumann boundary condition $N(u)|_{\partial D} = f$, with N the exterior unit normal to the boundary. The problem is well posed for all continuous functions f defined on ∂D with zero mean average. A typical problem of general theory is to classify all well posed BVP for a given elliptic system.

The following is a general fact, concerning classical solutions of general elliptic equations.

Classical solutions of elliptic equations with smooth (or real analytic) coefficients in a regular domain D are smooth (or real analytic), in the interior of D, independent of how smooth are the boundary conditions[16].

This follows, heuristically, from the propagation of singularities principle, discussed earlier, and absence of characteristic surfaces. Parabolic equations share this property. Hyperbolic equations, on the other hand, have a radically different behavior. In this case we expect that singularities of the initial data propagate along characteristic hypersurfaces.

Hyperbolic equations are, essentially, those for which the initial value problem is well posed. In that sense, they provide the natural framework for which one can prove a result similar to the local existence theorem for ODE. More precisely, for each sufficiently regular set of initial conditions there corresponds a unique solution; we can thus think of the Cauchy problem as a natural way of parametrizing the set of all solutions to the equations.

The definition of hyperbolicity depends, however, on the particular initial hypersurface we are considering. Thus, in the case of the wave equation $\Box u = 0$, the standard initial value problem

$$u(0, x) = u_0(x), \quad \partial_t u(0, x) = u_1$$

is well posed. This means that for any smooth initial data u_0, u_1 we can find a unique solution of the equation which depends continuously on u_0, u_1. As we have mentioned earlier, the IVP for $\Box u = 0$ remains well posed if we replace the initial hypersurface $t = 0$ by any space-like hypersurface $\psi(t, x) = 0$, see (253). It fails however to be well posed for timelike hypersurfaces such as $x^1 = 0$, see (254). In that case there may not exist any solution for a prescribed, non-analytic, Cauchy data.

[15]roughly half the order of the equation

[16]Provided that the boundary condition under consideration is well posed. Moreover this heuristic principle holds, in general, only for classical solutions of a nonlinear equation. There are in fact examples of well posed boundary value problems, for nonlinear elliptic systems, with no classical solutions.

It is more difficult to give find algebraic conditions of hyperbolicity. In principle hyperbolic equations differ from the elliptic ones, roughly, by the presence of a maximum number of characteristic hypersurfaces passing through any given point. Rather then attempting a general definition is more useful to give some examples of classes of hyperbolic PDEs.

One of the most useful class of hyperbolic equations is given by second order wave equations of the form

$$\square_a \phi = f, \qquad \square_a := -a^{00}\partial_t^2 + \sum_{i,j=1}^{d} a^{ij}\partial_i\partial_j \tag{255}$$

with coefficients a^{00}, a^{ij} and f which may depend on (t,x) as well as $\phi(t,x)$ and $\partial\phi(t,x)$. We need also to assume that $a^{00} > 0$ and a^{ij} verify the ellipticity condition,

$$\sum_{i,j=1}^{d} a^{ij}\xi_i\xi_j > 0, \quad \xi \in \mathbb{R}^d \tag{256}$$

The IVP for this type of equations is well posed, for any hypersurface $\psi(t,x) = 0$, such as $t = t_0$, for which,

$$-a^{00}(\partial_t\psi)^2 + \sum_{i,j=1}^{d} a^{ij}\partial_i\psi\partial_j\psi < 0, \quad \xi \in \mathbb{R}^d \tag{257}$$

A very useful generalization of (255) consist of the class of system of wave equation, diagonal with respect to the second derivatives, i.e. ,

$$\square_a \phi^I = F^I(\phi, \partial\phi), \qquad I = 1, 2, \ldots N, \tag{258}$$

where $\phi = (\phi^1, \phi^2, \ldots \phi^N)$. One can check, see (25), that the Einstein equations, in wave coordinates, can be written, almost, in this form.

Remark In reality the system (25), in wave coordinates, looks slightly different. The operator $\square_a$ has to be replaced by an operator of the form $\square_g = g^{\alpha\beta}\partial_\alpha\partial_\beta$ where $g^{\alpha\beta}$ is the inverse of a general Lorentz metric $g_{\alpha\beta}$, see (249).

Exercise Show by a local change of variables that an equation of the form $\square_g\phi = F(\phi, \partial_\phi)$ can be reduced, locally, to an equation of the form, $\square_a\phi = F(\phi, \partial_\phi)$ with a different lower order term F.

Another important class, which includes most of the important known examples of first order hyperbolic equations, such as the Maxwell equations, are of the form,

$$A^0(t,x,u)\partial_t u + \sum_{i=1}^{d} A_i(t,x,u)\partial_i u = F(t,x,u), \qquad u|\mathcal{H} = u_0 \tag{259}$$

where all the coefficients $A^0, A^1, \ldots A^d$ are symmetric $N \times N$ matrices and $\mathcal{H}$ is given by $\psi(t,x) = 0$. Such a system is well posed provided that the matrix,

$$A^0(t,x,u)\partial_t\psi(t,x) + \sum_{i=1}^{d} A_i(t,x,u)\partial_i\psi(t,x) \tag{260}$$

is positive definite. A system (259) verifying these conditions is called *symmetric hyperbolic*. In the particular case when $\psi = t$ the condition (260) becomes

$$(A^0\xi, \xi) \geq c|\xi|^2 \qquad \forall \xi \in \mathbb{R}^N. \tag{261}$$

Remark. It turns out that the second order wave equation[17] (255), verifying (257) can be written as a first order symmetric hyperbolic system. This can be simply done by introducing the new variables $v_0 = \partial_t u$, $v_1 = \partial_1 u, \ldots, v_d = \partial_d u$ and the obvious compatibility relations, such as $\partial_i v_j = \partial_j v_i$.

The following is a fundamental result in the theory of general hyperbolic equations It is called the local existence and uniqueness [LEU] for symmetric hyperbolic systems:

Theorem[LEU-Hyperbolic] *The initial value problem* (259)*, is locally well posed, for symmetric hyperbolic systems, with sufficiently smooth A, F, $\mathcal{H}$ and sufficiently smooth initial conditions u_0. In other words, if the above conditions are satisfied, then for any point $p \in \mathcal{H}$ there exist a sufficiently small neighborhood $\mathcal{D} \subset \mathbb{R}^{1+d}$ of p and a unique, continuously differentiable, solution $u : \mathcal{D} \to \mathbb{R}^N$.*

Remark 1. The issue of how smooth the initial data is allowed to be is an important question, still under investigation, for nonlinear equations.

Remark 2. The local character of the theorem is essential, the result cannot be globally true, in general for nonlinear systems. Indeed, as we have seen, the evolution problem (223) for the Burger equation, which fits trivially into the framework of symmetric hyperbolic systems, leads, after a sufficiently large time, to singular solutions. This happens independent of how smooth the initial data u_0 is. A precise version of the theorem above gives a lower bound on how large $\mathcal{D}$ can be.

Remark 3. The proof of the theorem is based on a variation of the Picard iteration method we have encountered earlier for ODE. One starts by taking $u_{(0)} = u_0$ in a neighborhood of $\mathcal{H}$ and then define recursively,

$$A^0(t, x, u_{(n-1)})\partial_t u_{(n)} + \sum_{i=1}^{d} A_i(t, x, u_{(n-1)})\partial_i u_{(n)} = F(t, x, u_{(n-1)}), \; u_{(n)}|\mathcal{H} = u_0 \tag{262}$$

Observe that at every stage of the iteration we have to solve a linear equation. Linearization is an extremely important tool in studying nonlinear PDE. We can rarely understand much about their behavior except near important special solutions, in which case linearization is an essential step. Thus, almost invariably, hard problems in non-linear PDE reduce to understanding specific problems in linear PDE.

Remark 4. To implement the Picard iteration method we need to get precise estimates on the u_n iterate in terms of the u_{n-1} iterate. This step requires energy type a-priori estimates.

[17] as well as the diagonal system (258)

Remark 5. Theorem[LEU-hyperbolic] has wide applications to various hyperbolic systems of physical interests. It applies, in particular, to prove a local existence result for the Einstein equations in wave coordinates, see (25).

Another important, characteristic[18], property of hyperbolic equations is *finite speed of propagation.* Consider the simple case of the wave equation (**??**). In this case the initial value problem can be solved explicitly by the Kirchoff formula (48). The formula allows us to conclude that if the initial data, at $t = 0$, is supported in a ball $B_a(x_0)$ of radius $a >$ centered at $x_0 \in \mathbb{R}^3$ then at time $t > 0$ the solution u is supported in the ball $B_{a+t}(x_0)$. In general finite speed of propagation can be best formulated in terms of domains of dependence and influence of hyperbolic equations. Given a point $p \in \mathbb{R}^{1+d}$, outside the initial hypersurface $\mathcal{H}$, we define $\mathcal{D}(p) \subset \mathcal{H}$ as the complement of the set of points $q \in \mathcal{H}$ with the property that any change of the initial conditions made in a small neighborhood V of q does not influence the value of solutions at p. More precisely if u, v are two solutions of the equation whose initial data differ only in V, must also coincide at p. The property of finite speed of propagation simply means that, for any point p, $\mathcal{D}(p)$ is compact in $\mathcal{H}$. A related notion is that of *domain of influence.* Given a set $D \subset \mathcal{H}$ the domain of influence of D is the smallest set $\mathcal{J}(D) \subset \mathbb{R}^{1+d}$ with the property that any two solutions u, v of the equation whose initial conditions coincide in the complement of D, must also coincide at all points in the complement of $\mathcal{J}(D)$. In the case of $\square u = 0$, if at $t = 0$, u and $\partial_t u$ are zero outside the unit ball B, $|x| \leq 1$, then, u is identically zero in the region $|x| > 1 + |t|$. Thus $\mathcal{J}(B)$ must be a subset of $\{(t, x)/\, |x| \leq 1 + |t|\}$ and it can be shown that in fact $\mathcal{J}(B) = \{(t, x)/\, |x| \leq 1 + |t|\}$. Observe also that the boundary of $\mathcal{J}(B)$ is formed by the union of two smooth characteristic hypersurfaces of the wave equation, $|x| = t + 1$ for $t \geq 0$ and $|x| = -t + 1$ for $t \leq 0$. This is a general fact, which illustrates once more the importance of characteristics.

The boundaries of domains of dependence of classical solutions to hyperbolic PDE are characteristic hypersurfaces, typically piecewise smooth.

In the case of linear hyperbolic equations, the characteristics of the equations, which (as we have seen) are solutions to nonlinear Hamilton-Jacobi equations, can be used to construct approximate solutions[19] to the equations, called *parametrices*, from which one can read the relevant information concerning propagation of singularities. One can then show that these singularities propagate along the bicharacteristics of the associated Hamiltonian. These techniques, however do not apply to general symmetric hyperbolic systems but rather to equations with well defined characteristics, such as those called *strictly hyperbolic.*

Finally a few words for parabolic equations and Schrödinger type equations[20]. A large class of useful equations of this type is given by,

$$\partial_t u - Lu = f, \tag{263}$$

[18]Elliptic, parabolic and dispersive equations do not have this property.

[19] For constant coefficient equations these approximate solutions are in fact exact solutions.

[20]General classes of dispersive equations are a bit harder to describe.

and, respectively

$$i\partial_t u + Lu = f \tag{264}$$

where L is the elliptic operator $L = \sum_{i,j=1}^{d} a^{ij}\partial_i\partial_j$ verifying the ellipticity condition (256). One looks for solutions $u = u(t, x)$, defined for $t \geq t_0$, with the prescribed initial condition,

$$u(t_0, x) = u_0(x) \tag{265}$$

on the hypersurface $t = t_0$. Strictly speaking this hypersurface is characteristic, since the order of the equation is $k = 2$ and we cannot determine $\partial_t^2 u$ at $t = t_0$ directly from the equation. Yet this is not a serious problem; we can still determine $\partial_t^2 u$ formally by differentiating the equation with respect to ∂_t. Thus, the initial value problem (263), (resp. (264)) and (265) is well posed, but in a slightly different sense than for hyperbolic equations. For example the heat equation $-\partial_t u + \Delta u$ is only well posed for positive t and ill posed for negative t. The heat equation may also not have unique solutions for the IVP unless we make assumptions about how fast the initial data is allowed to grow at infinity. One can also show that the only characteristics of the equation (263) are all of the form $t = t_0$ and therefore parabolic equations are quite similar to elliptic equations. For, example, one can show, consistent with our propagation of singularities principle, that if the coefficients a^{ij} and f are smooth (or real analytic), then, even if the initial data u_0 may not be smooth, the solution u must be smooth (or real analytic in x) for $t > t_0$. The heat equation smoothes out initial conditions. It is for this reason that the heat equation is useful in many applications. One often encounters diagonal systems of parabolic equations, of the form

$$\partial_t u^I - Lu^I = f^I(u, \partial u), \quad u = (u^1, u^2, \dots u^N)$$

with L as above. The system of equations (23), connected with the Ricci flow, is of this form.

Dispersive PDE, of which the Schrödinger equation (180) is a fundamental example, are evolution equations which, in many respects, behave analogously to hyperbolic PDEs (for instance, the initial value problem tends to be locally well-posed both forward and backward in time). However, solutions to dispersive PDEs do not propagate along characteristic surfaces, but instead move at speeds that are determined by their spatial frequency; in general, high-frequency waves tend to propagate at much greater speeds than low-frequency waves, eventually leading to a *dispersion* of the solution into increasingly large areas of space. In fact the speed of propagation of solutions is typically infinite. This behavior differs also from that of parabolic equations, which tend to *dissipate* the high-frequency components of a solution (sending them to zero) rather than dispersing them. In physics, dispersive equations arise in quantum mechanics, as the *non-relativistic limit* $c \to \infty$ of relativistic equations, and also as approximations to model certain types of fluid behavior. For instance, the *Korteweg de Vries equation*,

$$\partial_t u + \partial_x^3 u = 6u\partial_x u \tag{266}$$

is a dispersive PDE which models the behavior of small-amplitude waves in a shallow canal.

Remark: Elliptic and hyperbolic equations are the most robust, useful, classes of PDE. Other important classes, such as *parabolic* and *dispersive*, can be interpreted as lying at the boundaries[21] of these two classes. A neat classification of all equations into, elliptic, hyperbolic, parabolic and dispersive is unfortunately not possible, even for second order linear equations in two space dimensions.

5.1. Special topics for linear equations. General theory has been most successful in regard to linear equations (216). This is particularly true for linear equations with constant coefficients, for which Fourier analysis provides an extremely powerful tool, and for general, linear, elliptic equations. The theory of general linear elliptic and parabolic equations is also in good shape. We also have a reasonably good theory for variable coefficients hyperbolic equations[22], though less complete as in the elliptic case.

In connection with general linear equations we point out the existence of scalar, linear, operators $\mathcal{P}$ and smooth functions f for which the equation $\mathcal{P}[u] = f$ may have *no solutions*, in any domain $\Omega \subset \mathbb{R}^n$. The Cauchy-Kowalewski theorem gives a criterion for local solvability when f and the coefficients of $\mathcal{P}$ are real analytic, but it is a remarkable phenomenon that when one allows f to be smooth rather than real analytic, then serious obstructions to local solvability appear. For instance, the *Lewy operator*

$$\mathcal{P}[u](t,z) = \frac{\partial u}{\partial \bar{z}}(t,z) - iz\frac{\partial u}{\partial t}(t,z)$$

defined on complex-valued functions $u : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$, has the property that the equation $\mathcal{P}[u] = f$ is locally solvable for real analytic f, but not for "most" smooth f. The Lewy operator is intimately connected to the tangential Cauchy-Riemann complex on the Heisenberg group in $\mathbb{C}^2$, it shows in fact that such complexes fail, in general, to be exact. In was discovered in the study of the restriction of the 2-dimensional analogue of the Cauchy Riemann operator $\bar{\partial}$, see (16), to a quadric in $\mathbb{C}^2$. The example has provided the starting point the theory of local *local solvability* whose goal is to characterize linear equations which have the property of *local solvability.* Today it remains an important, even though less active, area of research in PDE.

As illustrated by the Lewy example, the theory of CR manifolds and the associated tangential Cauchy-Riemann complex is another extremely rich source of examples of interesting linear PDEs which do not fit in the standard classification. The theory has its origin in the study of restrictions of the Cauchy-Riemann equations, in higher dimensions, to real hypersurfaces, in particular boundaries pseudo-convex domains. This study has led to the Laplacean $\Box_b$, which was not elliptic but rather hypoelliptic. A linear operator $\mathcal{P}$ is said to be hypoelliptic in a domain D if any

[21] parabolic equations are singular, formal, limits of elliptic equations. Dispersive equations can be regarded also as singular limits of hyperbolic equations.

[22] Symmetric hyperbolic systems are suitable in the study of well-posedness and finite speed of propagation, but not so useful for the more refined question of propagation of singularities. For this goal one uses instead strictly hyperbolic systems or various definitions of hyperbolic systems of higher multiplicity.

solution to $\mathcal{P}[u] = f$, with f infinitely differentiable, i.e. smooth, in D, is infinitely differentiable id D.

Questions of *unique continuation* of solutions are also investigated by the general theory. Unique continuation results concern *ill posed problems* where general existence may fail, yet uniqueness survives. A typical example is Holmgren's theorem mentioned above. It asserts, in the particular case of the wave equation, that, even though the Cauchy problem for time-like hyper-surfaces is ill posed, if a solution exists it must necessarily be unique. More precisely, assume that a solution u of (**??**) is such that u and $\partial_z u$ vanish along the hyperplane $z = 0$. Then u must vanish identically in the whole space. Another fundamental example is that of *analytic continuation*: two complex-analytic functions on a connected domain D which agree on a non-discrete set (such as a disk or an interval), must necessarily agree everywhere on D. This fact can be viewed as a unique continuation result for the Cauchy-Riemann equations (15). Ill posed problems appear naturally in connection to control theory which deals with unphysical, auxiliary, boundary conditions which are introduced to guide solutions of the system to a desired state.

Besides the traditional questions of classification, local and global well-posedness and solvability, propagation of singularities, and unique continuation of solutions, there are other issues which are addressed by the general theory of linear PDE. A very active area of investigation is *spectral theory.* There is no way I can even begin to give an account of this theory, which is of fundamental importance not only to Quantum Mechanics, and other physical theories, but also to geometry and analytic number theory. A typical problem in spectral theory is to solve the eigenvalue problem in $\mathbb{R}^d$, or a domain $D \subset \mathbb{R}^d$,

$$-\Delta u(x) + V(x)u(x) = \lambda u(x) \tag{267}$$

that is to find the values $\lambda \in \mathbb{R}$, called eigenvalues, for which there exist solutions $u(x)$, localized in space, i.e. bounded in the $L^2(\mathbb{R}^d)$ norm, called eigenfunctions. The existence of an eigenfunction u implies that we can write solutions to the Schrödinger equation,

$$i\partial_t \phi + \Delta\phi - V\phi = 0 \tag{268}$$

of the form $\phi(t,x) = e^{-i\lambda t}u(x)$, called bound states of the physical system described by (268). The eigenvalues λ corresspond to the quanta energy levels of the system. They are very sensitive to the choice of potential V. The distribution of the eigenvalues of the Laplace operator Δ in a domain $D \subset \mathbb{R}^d$ depends on the geometry of the domain D, this is the case, for example, of the very important Weyl asymptotic formula. The *inverse spectral problem* is also important, can one determine the potential V from the knowledge of the corresponding eigenvalues ? The eigenvalue problem can be studied in considerable generality by replacing the operator $-\Delta+V$ with a general elliptic operator. More to the point is the study the eigenvalue problem for the Laplace-Beltrami operator associated to a Riemannian manifold. In the particular case of two dimensional manifolds of constant negative Gauss curvature, i.e. $K = -1$, this problem is important in number theory. A famous problem in differential geometry is to characterize the metric on a 2-dimensional compact manifold, from the spectral properties of the associated Laplace-Beltrami operator.

Related to spectral theory, in a sense opposite to it, is scattering theory. Attempts to formalize the intuition from quantum mechanics that a potential that is suitable small or localized is typically unable to "trap" a quantum particle, which is therefore likely to escape to infinity in a manner resembling that of a free particle. In the case of equation (268), solutions that scatter are those that behave freely as $t \to \infty$. That is, they behave like solutions to the free Schrödinger equation $i\partial_t \psi + \Delta\psi = 0$. A typical problem in scattering theory is to show that, if $V(x)$ tends to zero sufficiently fast as $|x| \to \infty$, all solutions, except the bound states, scatter as $t \to \infty$.

5.2. Conclusions. In the analytic case, the Cauchy Kowalewsky theorem allows us to solve, locally, the IVP for very general classes of PDE. We have a general theory of characteristic hypersurfaces of PDE and understand in considerable generality how they relate to propagation of singularities. We can also distinguish, in considerable generality, the fundamental classes of elliptic and hyperbolic equations and can define general parabolic and dispersive equations. The IVP for a large class of nonlinear hyperbolic systems can be solved locally in time, for sufficiently smooth initial conditions. Similar, local in time, results hold for general classes of nonlinear parabolic and dispersive equations. A lot more can be done for linear equations. We have satisfactory results concerning regularity of solutions for elliptic and parabolic equations and a good understanding of propagation of singularities for a large class of hyperbolic equations. Some aspects of spectral theory and scattering theory and problems of unique continuation can also be studied in considerable generality.

The main defect of the general theory concerns the passage from local to global. Important global features of special equations are too subtle to fit into a too general scheme; on the contrary each important PDE requires special treatment. This is particularly true for nonlinear equations; the large time behavior of solutions is very sensitive to the special features of the equation at hand. Moreover , general points of view may obscure, through unnecessary technical complications, the main properties of the important special cases. A useful general framework is one which provides a simple and elegant treatment of a particular phenomenon, as is the case of symmetric hyperbolic systems in connection to local well posedness and finite speed of propagation. Yet symmetric hyperbolic systems turn out to be simply too general for the study of more refined questions concerning the important examples of hyperbolic equations.

Finally, as another shortcoming of the general theory of linear PDE, we have remarked earlier that hard problems in non-linear PDE are almost always connected with specific linear problems. Yet, often, the linear problems which arise in this way are rather very special and subtle and thus cannot be treated with the degree of generality (and it is not at all necessary that they should be !) one might expect from a general theory.

CHAPTER 6

Equations Derived by the Variational Principle

1. Basic Notions

In this section we will discuss some basic examples of nonlinear wave equations which arise variationally from a relativistic Lagrangian. The fundamental objects of a relativistic field theory are

- **Space-time** $(\mathbf{M}, \mathbf{g})$ which consists of an $n+1$ dimensional manifold $\mathbf{M}$ and a Lorentz metric $\mathbf{g}$; i.e . a nondegenerate quadratic form with signature $(-1, 1, \dots, 1)$ defined on the tangent space at each point of $\mathbf{M}$. We denote the coordinates of a point in $\mathbf{M}$ by x^α, $\alpha = 0, 1, \dots, n$.

  Throughout most of this chapter the space-time will in fact be the simplest possible example - namely, the Minkowski space-time in which the manifold is $\mathbb{R}^{n+1}$ and the metric is given by
$$ds^2 = \mathbf{m}_{\alpha\beta} dx^\alpha dx^\beta = -dt^2 + \left(dx^1\right)^2 + \cdots + \left(dx^n\right)^2 \qquad (269)$$
  with $t = x^0, m_{\alpha\beta} = \mathrm{diag}(-1, 1, \dots, 1)$. Recall that any system of coordinates for which the metric has the form (269) is called inertial. Any two inertial coordinate systems are related by Lorentz transformations.
- **Collection of fields** $\psi = \psi^{(1)}, \psi^{(2)}, \dots, \psi^{(p)}$ which can be scalars, tensors, or some other geometric objects[1] such as spinors, defined on $\mathbf{M}$.
- **Lagrangian density** L which is a scalar function on $\mathbf{M}$ depending only on the tensorfields ψ and the metric[2] $\mathbf{g}$.

We then define the corresponding *action* $\mathcal{S}$ to be the integral,
$$\mathcal{S} = \mathcal{S}[\psi, \mathbf{g} : \mathcal{U}] = \int_{\mathcal{U}} L[\psi] dv_{\mathbf{g}}$$
where $\mathcal{U}$ is any relatively compact set of $\mathbf{M}$. Here $dv_{\mathbf{g}}$ denotes the volume element generated by the metric $\mathbf{g}$. More precisely, relative to a local system of coordinates x^α, we have
$$dv_{\mathbf{g}} = \sqrt{-\mathbf{g}} dx^0 dx^1 \cdots dx^n = \sqrt{-\mathbf{g}} dx$$
with g the determinant of the matrix $(\mathbf{g}_{\alpha\beta})$.

By a compact variation of a field ψ we mean a smooth one-parameter family of fields $\psi_{(s)}$ defined for $s \in (-\epsilon, \epsilon)$ such that,

[1]For simplicity we restrict ourselves to covariant tensors.
[2]as well as its inverse $\mathbf{g}^{-1}$

(1) At $s = 0$, $\psi_{(0)} = \psi$.
(2) At all points $p \in \mathbf{M} \setminus \mathcal{U}$ we have $\psi_{(s)} = \psi$.

Given such a variation we denote $\delta\psi := \dot{\psi} := \frac{d\psi_{(s)}}{ds}\Big|_{s=0}$. Thus, for small s,

$$\psi_{(s)} = \psi + s\dot{\psi} + O(s^2)$$

A field ψ is said to be *stationary* with respect to $\mathcal{S}$ if, for any compact variation $(\psi_{(s)}, \mathcal{U})$ of ψ, we have

$$\frac{d}{ds}\mathbf{S}(s)\Big|_{s=0} = 0$$

where,

$$\mathbf{S}(s) = \mathbf{S}[\psi_{(s)}, \mathbf{g}; \mathcal{U}]$$

We write this in short hand notation as

$$\frac{\delta \mathbf{S}}{\delta \psi} = 0$$

Action Principle, also called the Variational Principle, states that an acceptable solution of a physical system must be stationary with respect to a given Lagrangian density called the Lagrangian of the system. The action principle allows us to derive partial differential equations for the fields ψ called the Euler-Lagrange equations. Here are some simple examples:

1. Scalar Field Equations :

One starts with the Lagrangian density

$$L[\phi] = -\frac{1}{2}\mathbf{g}^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)$$

where ϕ is a complex scalar function defined on $(\mathbf{M}, \mathbf{g})$ and $V(\phi)$ a given real function of ϕ.

Given a compact variation $(\phi_{(s)}, \mathcal{U})$ of ϕ, we set $\mathcal{S}(s) = \mathcal{S}[\phi_{(s)}, \mathbf{g}; \mathcal{U}]$. Integration by parts gives,

$$\begin{aligned} \frac{d}{ds}\mathcal{S}(s)\Big|_{s=0} &= \int_{\mathcal{U}} [-\mathbf{g}^{\mu\nu}\partial_\mu\dot{\phi}\partial_\nu\phi - V'(\phi)\dot{\phi}]\sqrt{-\mathbf{g}}dx \\ &= \int_{\mathcal{U}} \dot{\phi}[\Box_{\mathbf{g}}\phi - V'(\phi)]dv_{\mathbf{g}}] \end{aligned}$$

where $\Box_{\mathbf{g}}$ is the D'Alembertian,

$$\Box_{\mathbf{g}}\phi = \frac{1}{\sqrt{-\mathbf{g}}}\partial_\mu\Big(\mathbf{g}^{\mu\nu}\sqrt{-\mathbf{g}}\,\partial_\nu\phi\Big)$$

.

In view of the action principle and the arbitrariness of $\dot{\phi}$ we infer that ϕ must satisfy the following Euler-Lagrange equation

$$\Box_{\mathbf{g}}\phi - V'(\phi) = 0, \tag{270}$$

Equation (270) is called the *scalar wave equation* with potential $V(\phi)$.

CONFORMAL PROPERTIES **2. Wave Maps** :

The wave map equations will be defined in the context of a space-time $(\mathbf{M}, \mathbf{g})$, a Riemannian manifold N with metric h, and a mapping

$$\phi : \mathbf{M} \longrightarrow N.$$

We recall that if X is a vectorfield on $\mathbf{M}$ then $\phi_* X$ is the vectorfield on N defined by $\phi_* X(f) = X(f \circ \phi)$. If ω is a 1-form on N its pull-back $\phi^* \omega$ is the 1-form on $\mathbf{M}$ defined by $\phi^* \omega(X) = \omega(\phi_* X)$, where X is an arbitrary vectorfield on $\mathbf{M}$. Similarly the pull-back of the metric h is the symmetric 2-covariant tensor on $\mathbf{M}$ defined by the formula $(\phi^* h)(X, Y) = h(\phi_* X, \phi_* Y)$. In local coordinates x^α on $\mathbf{M}$ and y^a on N, if ϕ^a denotes the components of ϕ relative to y^a, we have,

$$(\phi^* h)_{\alpha\beta}(p) = \frac{\partial \phi^a}{\partial x^\alpha} \frac{\partial \phi^b}{\partial x^\beta} h_{ab}(\phi(p)) = \langle \frac{\partial \phi}{\partial x^\alpha}, \frac{\partial \phi}{\partial x^\beta} \rangle$$

where $< \cdot, \cdot >$ denotes the Riemannian scalar product on N.

Consider the following Lagrangian density involving the map ϕ,

$$L = -\frac{1}{2} \mathrm{Tr}_{\mathbf{g}}(\phi^* h)$$

where $\mathrm{Tr}_{\mathbf{g}}(\phi^* h)$ denotes the trace relative to $\mathbf{g}$ of $\phi^* h$. In local coordinates,

$$L[\phi] = -\frac{1}{2} \mathbf{g}^{\mu\nu} h_{ab}(\phi) \frac{\partial \phi^a}{\partial x^\mu} \frac{\partial \phi^b}{\partial x^\nu}.$$

By definition *wave maps* are the stationary points of the corresponding action. Thus by a a straightforward calculation,

$$\begin{aligned} 0 &= \frac{d}{ds} \mathcal{S}(s) \Big|_{s=0} = I_1 + I_2 \qquad (271) \\ I_1 &= -\frac{1}{2} \int_{\mathcal{U}} \mathbf{g}^{\mu\nu} \frac{\partial h_{ab}(\phi)}{\partial \phi^c} \dot{\phi}^c \, \partial_\mu \phi^a \partial_\nu \phi^b \sqrt{-\mathbf{g}} dx \\ I_2 &= -\int_{\mathcal{U}} \mathbf{g}^{\mu\nu} h_{ab}(\phi) \partial_\mu \dot{\phi}^a \partial_\nu \phi^b \sqrt{-\mathbf{g}} dx \end{aligned}$$

After integrating by parts, relabelling and using the symmetry in b, c, we can rewrite I_2 in the form,

$$\begin{aligned} I_2 &= \int_{\mathcal{U}} \dot{\phi}^a \left(h_{ab}(\phi) \Box_{\mathbf{g}} \phi^b + \mathbf{g}^{\mu\nu} \frac{\partial h_{ab}}{\partial \phi^c} \partial_\mu \phi^c \partial_\nu \phi^b \right) dv_{\mathbf{g}} \qquad (272) \\ &= \int_{\mathcal{U}} \dot{\phi}^a \left(h_{ab}(\phi) \Box_{\mathbf{g}} \phi^b + \frac{1}{2} \mathbf{g}^{\mu\nu} \left(\frac{\partial h_{ab}}{\partial \phi^c} + \frac{\partial h_{ac}}{\partial \phi^b} \right) \partial_\mu \phi^b \partial_\nu \phi^c \right) dv_{\mathbf{g}} \end{aligned}$$

Also, relabelling indices

$$I_1 = -\frac{1}{2} \int_{\mathcal{U}} \mathbf{g}^{\mu\nu} \frac{\partial h_{bc}}{\partial \phi^a} \dot{\phi}^a \, \partial_\mu \phi^b \partial_\nu \phi^c \, dv_{\mathbf{g}}.$$

Therefore,

$$\begin{aligned}
0 &= I_1 + I_2 \\
&= \int_{\mathcal{U}} \dot{\phi}^a \left(h_{ab} \Box_{\mathbf{g}} \phi^b + \partial_\mu \phi^b \partial_\nu \phi^c \mathbf{g}^{\mu\nu} \frac{1}{2} \left(\frac{\partial h_{ab}}{\partial \phi^c} + \frac{\partial h_{ac}}{\partial \phi^b} - \frac{\partial h_{bc}}{\partial \phi^a} \right) \right) dv_{\mathbf{g}} \\
&= \int_{\mathcal{U}} \dot{\phi}^a \left(h_{ad} \Box_{\mathbf{g}} \phi^d + \partial_\mu \phi^b \partial_\nu \phi^c \mathbf{g}^{\mu\nu} \frac{1}{2} h^{ds} h_{ad} \cdot \left(\frac{\partial h_{sb}}{\partial \phi^c} + \frac{\partial h_{sc}}{\partial \phi^b} - \frac{\partial h_{bc}}{\partial \phi^s} \right) \right) dv_{\mathbf{g}} \\
&= \int_{\mathcal{U}} \dot{\phi}^a h_{ad} \left(\Box_{\mathbf{g}} \phi^d + \partial_\mu \phi^b \partial_\nu \phi^c \mathbf{g}^{\mu\nu} \Gamma^d_{bc} \right) dv_{\mathbf{g}}
\end{aligned}$$

where $\Gamma^d_{bc} = \frac{1}{2} h^{ds} \left(\frac{\partial h_{sb}}{\partial \phi^c} + \frac{\partial h_{sc}}{\partial \phi^b} - \frac{\partial h_{bc}}{\partial \phi^s} \right)$ are the Christoffel symbols corresponding to the Riemannian metric h. The arbitrariness of $\dot{\phi}$ yields the following equation for wave maps,

$$\Box_{\mathbf{g}} \phi^a + \Gamma^a_{bc}\, \mathbf{g}^{\mu\nu} \partial_\mu \phi^b \partial_\nu \phi^c = 0 \tag{273}$$

Example: Let N be a two dimensional Riemannian manifold endowed with a metric h of the form,

$$ds^2 = dr^2 + f(r)^2 d\theta^2$$

Let ϕ be a wave map from $\mathbf{M}$ to N with components ϕ^1, ϕ^2, relative to the r, θ coordinates. Then, $\Gamma^1_{11} = \Gamma^2_{11} = \Gamma^1_{12} = \Gamma^2_{22} = 0$ and $\Gamma^1_{22} = -f'(r) f(r)$, $\Gamma^2_{12} = \frac{f'(r)}{f(r)}$. Therefore,

$$\begin{aligned}
\Box_{\mathbf{g}} \phi^1 &= f'(r) f(r) \mathbf{g}^{\mu\nu} \partial_\mu \phi^2 \partial_\nu \phi^2 \\
\Box_{\mathbf{g}} \phi^2 &= -\frac{f'(r)}{f(r)} \mathbf{g}^{\mu\nu} \partial_\mu \phi^1 \partial_\nu \phi^2
\end{aligned}$$

The equations of wave maps can be given a simpler formulation when N is a submanifold of the Euclidean space $\mathbb{R}^m$. In this case, the metric h is the Euclidean metric[3] so the first term in (271) vanishes.

$$\begin{aligned}
\frac{d}{ds} \mathcal{S}(s) \Big|_{s=0} &= -\int_{\mathcal{U}} \mathbf{g}^{\alpha\beta} \langle \frac{\partial \phi}{\partial x^\alpha}, \frac{\partial \dot{\phi}}{\partial x^\beta} \rangle dv_{\mathbf{g}} \\
&= \int_{\mathcal{U}} < \Box \phi, \dot{\phi} > dv_{\mathbf{g}}
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product and $\Box$ the D'Alembertian operator on $\mathbf{M}$. Thus the Euler-Lagrange equations take the form,

$$(\Box \phi(p))^T = 0 \tag{274}$$

where T here means the projection onto the tangent space of N at $\phi(p)$.

In the special case when $N \subset \mathbb{R}^m$ is a hypersurface, we can rewrite (274) in a more concrete form. Let ν be the unit normal on N and k the second fundamental form $k(X, Y) = \langle \mathbf{D}_X \nu, Y \rangle$, with $\mathbf{D}_X$ the standard covariant derivative of Euclidean space. The hypersurface N is defined (locally) as the level set of some real valued

[3] Use the standard coordinates of the ambient Euclidean space.

f. Differentiating the equation $f(\phi(x)) = 0$ with respect to local coordinates x^μ on $\mathbf{M}$ yields $0 = < \nu(\phi), \partial_\mu \phi >$ along $\mathbf{M}$. Hence,

$$\begin{aligned} 0 &= \partial^\mu < \nu(\phi), \partial_\mu \phi > = < \Box\phi, \nu > + \mathbf{g}^{\mu\nu} < \partial_\nu \nu(\phi), \partial_\mu \phi > \\ &= < \Box\phi, \nu > + \mathbf{g}^{\mu\nu} < \nabla_{\phi_*(E_\nu)} \nu, \phi_* (E_\mu) > \end{aligned}$$

Where $\phi_* (E_\mu) = \frac{\partial \phi^i}{\partial x^\mu} \frac{\partial}{\partial y^i}$ is the pushforward of $E_\mu = \frac{\partial}{\partial x^\mu}$. In particular, $\phi_*(E_\mu)$ is tangent to N. Therefore,

$$< \Box\phi, \nu > = -k(\phi_*(E^\alpha), \phi_*(E_\alpha)) \tag{275}$$

Thus the equation for wave maps becomes,

$$\Box\phi = -k(\phi_*(E^\alpha), \phi_*(E_\alpha))\, \nu$$

In the case when N is the standard sphere $\mathbb{S}^{m-1} \subset \mathbb{R}^m$, $k(X, Y) = - < X, Y >$ and the equation for wave maps becomes, in standard coordinates x^α in $\mathbb{R}^m$, $\nu^a(\phi) = \phi^a$,

$$\Box\phi^a = -\phi^a \mathbf{g}^{\alpha\beta} < \frac{\partial\phi}{\partial x^\alpha}, \frac{\partial\phi}{\partial x^\beta} >$$

3. Maxwell equations:

An *electromagnetic field* F is an exact two form on a four dimensional manifold $\mathbf{M}$. That is, F is an antisymmetric tensor of rank two such that

$$F = dA \tag{276}$$

where A is a one-form on $\mathbf{M}$ called a *gauge potential* or connection 1-form. Note that A is not uniquely defined - indeed if χ is an arbitrary scalar function then the transformation

$$A \longrightarrow \tilde{A} = A + d\chi \tag{277}$$

yields another gauge potential $\tilde{A}$ for F. This degree of arbitrariness is called gauge freedom, and the transformations (277) are called gauge transformations.

The Lagrangian density for electromagnetic fields is

$$L[F] \quad = \quad -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}.$$

Any compact variation $(F_{(s)}, \mathcal{U})$ of F can be written in terms of a compact variation $(A_{(s)}, \mathcal{U})$ of a gauge potential A, so that $F_{(s)} = dA_{(s)}$. Write

$$\dot{F} = \frac{d}{ds} F_{(s)} \Big|_{s=0}, \qquad \dot{A} = \frac{d}{ds} A_{(s)} \Big|_{s=0}$$

so that relative to a coordinate system x^α we have $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and therefore $\dot{F}_{\mu\nu} = \partial_\mu \dot{A}_\nu - \partial_\nu \dot{A}_\mu$. The action principle gives

$$\begin{aligned} 0 &= \frac{d}{ds}\mathcal{S}(s)\Big|_{s=0} = -\frac{1}{2}\int_{\mathbf{M}} \dot{F}_{\mu\nu} F^{\mu\nu} dv_{\mathbf{g}} \\ &= -\frac{1}{2}\int_{\mathcal{U}} (\partial_\mu \dot{A}_\nu - \partial_\nu \dot{A}_\mu) F^{\mu\nu} dv_{\mathbf{g}} \\ &= -\int_{\mathcal{U}} \partial_\mu \dot{A}_\nu F^{\mu\nu} dv_{\mathbf{g}} = \int_{\mathcal{U}} \dot{A}_\nu \left(\frac{1}{\sqrt{-\mathbf{g}}} \partial_\nu \left(\sqrt{-\mathbf{g}} F^{\mu\nu} \right) \right) dv_{\mathbf{g}} \end{aligned}$$

Note that the second factor in the integrand is just $\mathbf{D}_\mu F^{\mu\nu}$ where $\mathbf{D}$ is the covariant derivative on $\mathbf{M}$ corresponding to $\mathbf{g}$. Hence the Euler-Lagrange equations take the form

$$\mathbf{D}_\nu F^{\mu\nu} = 0. \tag{278}$$

Together, (276) and (278) constitute the Maxwell equations.

Exercise. Given a vector field X^α on $\mathbf{M}$, show

$$\mathbf{D}_\alpha X^\alpha = \frac{1}{\sqrt{-\mathbf{g}}} \partial_\alpha \left(\sqrt{-\mathbf{g}} X^\alpha \right)$$

We can write the Maxwell equations in a more symmetric form by using the Hodge dual of F,

$$^\star F_{\mu\nu} = \frac{1}{2} \in_{\mu\nu\alpha\beta} F^{\alpha\beta}$$

and by noticing that (278) is equivalent to $d\,^\star F = 0$. The Maxwell equations then take the form

$$dF = 0, \qquad d\,^\star F = 0 \tag{279}$$

or, equivalently,

$$\mathbf{D}_\nu F^{\mu\nu} = 0, \qquad \mathbf{D}_\nu\, {}^\star F^{\mu\nu} = 0 \tag{280}$$

Note that since Lorentz transformations commute with both the Hodge dual and exterior differentiation, the Lorentz invariance of the Maxwell equations is explicit in (279).

Definition. Given X an arbitrary vector field, we can define the contractions

$$\begin{aligned} E_\alpha &= (i_X F)_\alpha &= X^\mu F_{\alpha\mu} \\ H_\alpha &= (i_X\, {}^\star F)_\alpha &= X^\mu\, {}^\star F_{\alpha\mu} \end{aligned}$$

called, respectively, the electric and magnetic components of F. Note that both these one-forms are perpendicular to X.

We specialize to the case when $\mathbf{M}$ is the Minkowski space and $X = \frac{d}{dx^0} = \frac{d}{dt}$. As remarked, E, H are perpendicular to $\frac{d}{dt}$, so $E_0 = H_0 = 0$. The spatial components are by definition

$$\begin{aligned} E_i &= F_{0i} \\ H_i &= {}^\star F_{0i} = \frac{1}{2} \in_{0ijk} F^{jk} = \frac{1}{2} \in_{ijk} F^{jk} \end{aligned}$$

We now use (279) to derive equations for E and H from above, which imply

$$\mathbf{D}_\nu \, {}^\star F^{\mu\nu} = 0 \tag{281}$$

and (278), respectively. Setting $\mu = 0$ in both equations of (280) we derive,

$$\partial^i E_i = 0, \qquad \partial^i H_i = 0 \tag{282}$$

Setting $\mu = i$ and observing that $F_{ij} = \in_{ijk} H^k$, $\quad {}^\star F_{ij} = - \in_{ijk} E^k$ we write

$$\begin{aligned} 0 &= -\partial^0 E_i + \partial^j F_{ij} \quad = \partial_0 E_i + \in_{ijk} \partial^j H^k = \partial_t E_i + (\nabla \times H)_i \\ 0 &= \partial_t H_i - \in_{ijk} \partial_j E_k = \partial_t H_i - (\nabla \times E)_i \end{aligned}$$

Therefore,

$$\partial_t E + \nabla \times H = 0 \tag{283}$$

$$\partial_t H - \nabla \times E = 0 \tag{284}$$

Alongside (283) and (284) we can assign data at time $t = 0$,

$$E_i(0, x) = E_i^{(0)}, \qquad H_i(0, x) = H_i^{(0)}$$

Exercise. Show that the equations (282) are preserved by the time evolution of the system (283)-(284). In other words if $E^{(0)}, H^{(0)}$ satisfy (282) then they are satisfied by E, H for all times $t \in \mathbb{R}$.

4. Yang-Mills equations :

The Lagrangians of all classical field theories exhibit the symmetries of the space-time. In addition to these space-time symmetries a Lagrangian can have symmetries called internal symmetries of the field. A simple example is the complex scalar Lagrangian,

$$L = -\frac{1}{2}\mathbf{m}^{\alpha\beta}\partial_\alpha \phi \overline{\partial_\beta \phi} - V(|\phi|)$$

where ϕ is a complex valued scalar defined on the Minkowski space-time $\mathbb{R}^{n+1}$, $\bar{\phi}$ its complex conjugate. We note that L is invariant under the transformations $\phi \to e^{i\theta}\phi$ with θ a fixed real number. It is natural to ask whether the Lagrangian can be modified to allow more general, *local* phase transformations of the form $\phi(x) \to e^{i\theta(x)}\phi(x)$. It is easy to see that under such transformations, the Lagrangian fails to be invariant, due to the term $\mathbf{m}^{\alpha\beta}\partial_\alpha \phi \overline{\partial_\beta \phi}$. To obtain an invariant Lagrangian one replaces the derivatives $\partial_\alpha \phi$ by the covariant derivatives $D_\alpha^{(A)}\phi \equiv \phi_{,\alpha} + iA_\alpha \phi$ depending on a gauge potential A_α. We can now easily check that the new Lagrangian

$$L = -\frac{1}{2}\mathbf{m}^{\alpha\beta} D_\alpha^{(A)} \phi \bar{D}_\beta^{(A)} \phi - V(|\phi|)$$

is invariant relative to the local transformations,

$$\phi(x^\alpha) \to e^{i\theta(x)}\phi(x^\alpha) \quad , \quad A_\alpha \to A_\alpha - \theta_{,\alpha}$$

called gauge transformations.

Remark that the gauge transformations introduced above fit well with the definition of the electromagnetic field F. Indeed setting $F = dA$ we notice that F is invariant. This allows us to consider a more general Lagrangian which includes F,

$$L = -\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} - \frac{1}{2}m^{\alpha\beta}\phi_{,\alpha}\bar{\phi}_{,\beta} - V(|\phi|)$$

called the *Maxwell-Klein-Gordon* Lagrangian.

The Yang-Mills Lagrangian is a natural generalization of the Maxwell-Klein-Gordon Lagrangian to the case when the group $SU(1)$, corresponding to the phase transformations of the complex scalar ϕ, is replaced by a more general Lie group G. In this case the role of the *gauge potential* or *connection 1-form* is taken by a $\mathcal{G}$ valued one form $A = A_\mu dx^\mu$ defined on $\mathbf{M}$. Here $\mathcal{G}$ is the Lie algebra of the Lie group G. Let $[\cdot,\cdot]$ its Lie bracket and $<\cdot,\cdot>$ its Killing scalar product. Typically the Lie group G is one of the classical groups of matrices, i.e. a subroup of either $\mathrm{Mat}(n,\mathbb{R})$ or $\mathrm{Mat}(n,\mathbb{C})$. We pause briefly to recall some facts about the relavent Lie groups and their Lie algebras.

(1) The orthogonal groups $\mathbf{O}(p,q)$. These are the groups of linear transformations of Re^n which preserve a given nondegenerate symmetric bilinear form of signature p, q, $p+q=n$. We denote by $\mathbf{R}^n_{p,q}$ the corresponding space. The case $p=0$ is that of the Euclidean case, the group is then simply denoted by $\mathbf{O}(n)$. The case $p=1, q=n$ is that of the Minkowski space-time $\mathbb{R}^{n+1}$, the group $\mathbf{O}(1,n)$ is the Lorentz group. In general let Q be the diagonal matrix whose first p diagonal elements are -1 and the remaining ones are $+1$. Then,

$$\begin{aligned}\mathbf{O}(p,q) &= \{L \in \mathrm{M}at(n,\mathbb{R}) | L^T Q L = Q\} \\ &= \{L \in \mathrm{M}at(n,\mathbb{R}) | L M L^T = M\}\end{aligned}$$

Note that for $L \in \mathbf{O}(p,q)$, $\det(L) = \pm 1$.
Recall that the special orthogonal groups $\mathbf{SO}(p,q)$ are defined by

$$\mathbf{SO}(p,q) = \{L \in \mathbf{O}(p,q) | \det L = 1\}.$$

They correspond to all orientation preserving isometries of $\mathbf{R}^n_{p,q}$. Both $\mathbf{O}(p,q)$ and $\mathbf{SO}(p,q)$ have as Lie algebra[4]

$$\mathbf{so}(p,q) = \{A \in \mathrm{M}at(n,\mathbb{R}) | AQ + QA^T = 0\}.$$

and that $\dim_{\mathbf{R}}\mathbf{O}(p,q) = \dim_{\mathbf{R}}\mathbf{SO}(p,q) = n(n-1)/2$. The Lie bracket for $\mathbf{so}(p,q)$ is the usual Lie bracket of matrices, i.e. $[A,B] = AB - BA$ and we have the Jacobi identity

$$[A,[B,C]] + [C,[A,B]] + [B,[C,A]] = 0 \tag{285}$$

and its Killing scalar product $<A,B> = -\mathrm{Tr}(AB^T)$ (where Tr is the usual trace for matrices) enjoys the compatibility condition

$$<A,[B,C]> = -<[A,B],C> \tag{286}$$

[4]Recall that the Lie algebra of a Lie group G is simply the tangent space to G at the origin.

(2) The unitary groups $\mathbf{U}(p,q)$. These are the complex analogues of the orthogonal groups. They are the groups of all linear transformations of $\mathbb{C}^n$ which preserve a given nondegenerate hermitian bilinear form. Denote by $\mathbb{C}^n_{p,q}$ the corresponding space. Then, with the matrix Q as above,

$$\mathbf{U}(p,q) = \{U \in \mathrm{M}at(n,\mathbb{C}) \mid U^*QU = Q\}$$

and,

$$\mathbf{SU}(p,q) = \{U \in \mathbf{U}(p,q) \mid \det U = 1\},$$

The corresponding Lie algebras are,

$$\begin{aligned}\mathbf{u}(p,q) &= \{A \in \mathrm{M}at(n,\mathbb{C}) \mid AQ + QA^* = 0\},\\ \mathbf{su}(p,q) &= \{A \in \mathcal{U}(p,q) \mid \mathrm{tr}_M A = 0\},\end{aligned}$$

where the trace $\mathrm{tr}_Q A = Q^{ij}A_{ij}$. The Lie bracket is again the usual one for matrices. The Killing scalar product is given by $< A, B >= -\mathrm{Tr}(AB^*)$. Remark also that $\dim_{\mathbf{R}}\mathbf{U}(p,q) = n^2$, $\dim_{\mathbf{R}}\mathbf{SU}(p,q) = n^2 - 1$.

In the Yang-Mills theory one is interested in compact Lie groups with a positive definite Killing form. This is the case for the groups $O(n), SO(n), U(n), SU(n)$.

In a given system of coordinates the connection 1-form A has the form, $A_\mu dx^\mu$ and we define the (gauge) covariant derivative of a $\mathcal{G}$-valued tensor ψ by

$$\mathbf{D}^{(A)}_\mu \psi = \mathbf{D}_\mu \psi + [A_\mu, \psi] \tag{287}$$

where $\mathbf{D}$ is the covariant derivative on $\mathbf{M}$. Observe that (287) is invariant under the following *gauge transformations*, for a given $\mathcal{G}$-valued gauge potential A and a $\mathcal{G}$- valued tensor ψ,

$$\tilde{\psi} \;=\; U^{-1}\psi U, \qquad \tilde{A}_\alpha = U^{-1}A_\alpha U + \left(\mathbf{D}_\alpha U^{-1}\right)U \tag{288}$$

with $U \in G$.

Proposition 1.1.

$$\begin{aligned}\mathbf{D}^{(\tilde{A})}_\mu \tilde{\psi} &= U^{-1}\left(\mathbf{D}^{(A)}_\mu \psi\right)U\\ &= \widetilde{\mathbf{D}^A\psi}\end{aligned}$$

Proof : This just requires some patience. First we will show

$$\mathbf{D}_\alpha\left(U^{-1}\psi U\right) = U^{-1}\left(\mathbf{D}_\alpha\psi + [\psi, U(\mathbf{D}_\alpha U^{-1})]\right)U$$

Indeed

$$\begin{aligned}\mathbf{D}_\alpha\left(U^{-1}\psi U\right) &= \left(\mathbf{D}_\alpha U^{-1}\right)\psi U + U^{-1}\left(\mathbf{D}_\alpha\psi\right)U + U^{-1}\psi\left(\mathbf{D}_\alpha U\right)\\ &= U^{-1}\left(-\left(\mathbf{D}_\alpha U\right)U^{-1}\psi + \mathbf{D}_\alpha\psi + \psi\left(\mathbf{D}_\alpha U\right)U^{-1}\right)U\\ &= U^{-1}\left(\mathbf{D}_\alpha\psi + [\psi, \left(\mathbf{D}_\alpha U\right)U^{-1}]\right)U\end{aligned}$$

as desired. Hence

$$
\begin{aligned}
\mathbf{D}_\alpha^{(\tilde{A})}\tilde{\psi} &= \mathbf{D}_\alpha\tilde{\psi} + [\tilde{A}_\alpha, \tilde{\psi}] \\
&= U^{-1}\left(\mathbf{D}_\alpha\psi + [\psi, U\left(\mathbf{D}_\alpha U^{-1}\right)]\right) + \left[U^{-1}A_\alpha U + \left(\mathbf{D}_\alpha U^{-1}\right)U, U^{-1}\psi U\right] \\
&= U^{-1}\left(\mathbf{D}_\alpha\psi + [\psi, \left(\mathbf{D}_\alpha U\right)U^{-1}] + [A_\alpha, \psi] + \left[U\left(\mathbf{D}_\alpha U^{-1}\right), \psi\right]\right)U \\
&= U^{-1}\left(\mathbf{D}_\alpha\psi + [A_\alpha, \psi]\right)U = \widetilde{\mathbf{D}_\alpha^{(A)}\psi}
\end{aligned}
$$

■

As in Riemmanian geometry, commuting two (gauge) covariant derivatives produces a fundamental object called the curvature, here denoted by F

$$\mathbf{D}_\alpha\mathbf{D}_\beta\psi - \mathbf{D}_\beta\mathbf{D}_\alpha\psi = [F_{\alpha\beta}, \psi] \tag{289}$$

where the components $F_{\alpha\beta}$ of the curvature can be deduced by the following straightforward computation:

$$
\begin{aligned}
\mathbf{D}_\alpha\mathbf{D}_\beta\psi &= \mathbf{D}_\alpha\left(\mathbf{D}_\beta\psi\right) + [A_\alpha, \mathbf{D}_\beta\psi] \\
&= \mathbf{D}_\alpha\left(\mathbf{D}_\beta\psi + [A_\beta, \psi]\right) + [A_\alpha, \mathbf{D}_\beta\psi + [A_\beta, \psi]] \\
&= \mathbf{D}_\alpha\mathbf{D}_\beta\psi + [\mathbf{D}_\alpha A_\beta, \psi] + [A_\beta, \mathbf{D}_\alpha\psi] + [A_\alpha, \mathbf{D}_\beta\psi] + [A_\alpha, [A_\beta, \psi]]
\end{aligned}
$$

So that

$$
\begin{aligned}
\left(\mathbf{D}_\alpha\mathbf{D}_\beta - \mathbf{D}_\beta\mathbf{D}_\alpha\right)\psi &= [\mathbf{D}_\alpha A_\beta - \mathbf{D}_\beta A_\alpha, \psi] \\
&\quad + \underbrace{[A_\alpha, [A_\beta, \psi]] - [A_\beta, [A_\alpha, \psi]]}_{[[A_\alpha, A_\beta], \psi]}
\end{aligned}
$$

Therefore,

$$F_{\alpha\beta} = \mathbf{D}_\alpha A_\beta - \mathbf{D}_\beta A_\alpha + [A_\alpha, A_\beta] \tag{290}$$

We leave it to the reader to show that the curvature tensor F is invariant under gauge transformations. That is,

$$\widetilde{F^{(\tilde{A})}}\ \left(\equiv U^{-1}F^{(\tilde{A})}U\right) = F^{(A)}$$

and that F satisfies the Bianchi identity

$$\mathbf{D}_\alpha F_{\beta\gamma} + \mathbf{D}_\gamma F_{\alpha\beta} + \mathbf{D}_\beta F_{\gamma\alpha} = 0 \tag{291}$$

We are finally ready to present the generalization of the Maxwell theory provided by the Yang-Mills Lagrangian:

$$L[A] = -\frac{1}{4} < F^{(A)}_{\alpha\beta}, F^{(A)\,\alpha\beta} >_{\mathcal{G}} \tag{292}$$

We derive the Euler-Lagrange equations just as in the Maxwell theory,

$$\begin{aligned}
0 &= \frac{d}{ds}\mathbf{S}(s)\Big|_{s=0} = -\frac{1}{2}\int_{\mathcal{U}} < \dot{F}_{\alpha\beta}, F^{\alpha\beta} >_{\mathcal{G}} dv_{\mathbf{g}} \\
&= -\frac{1}{2}\int_{\mathcal{U}} < \mathbf{D}_\alpha \dot{A}_\beta - \mathbf{D}_\beta \dot{A}_\alpha + [\dot{A}_\alpha, A_\beta] + [A_\alpha, \dot{A}_\beta], F^{\alpha\beta} >_{\mathcal{G}} dv_{\mathbf{g}} \\
&= -\int_{\mathcal{U}} < \mathbf{D}_\alpha \dot{A}_\beta, F^{\alpha\beta} > + < [A_\alpha, \dot{A}_\beta], F^{\alpha\beta} >_{\mathcal{G}} dv_{\mathbf{g}} \\
&= \int_{\mathcal{U}} < \dot{A}_\beta, \mathbf{D}_\alpha F^{\alpha\beta} >_{\mathcal{G}} + < \dot{A}_\beta, \left[A_\alpha, F^{\alpha\beta}\right] >_{\mathcal{G}} dv_{\mathbf{g}}
\end{aligned}$$

which implies

$$\mathbf{D}_\nu F^{\mu\nu} = 0 \tag{293}$$

Together, (291) and (293) form the Yang-Mills equations.

Note that the equations are invariant under the group of gauge transformations. A solution of the Yang-Mills equations, then, is an equivalence class of gauge-equivalent potentials A_α whose curvature F satisfies (293).

In our later treatment of Yang-Mills, we will almost always specify a representative of a solution's equivalence class by imposing additional constraints - called *gauge conditions* - on A. There are three standard ways of doing this, each yielding its own rendition of the Yang-Mills equations with its own faults and advantages:

- **Coulomb Gauge** is defined by,

  $$\nabla^i A_i(t,x) = 0 \quad (t,x) \in \mathbb{R}^{n+1} \tag{294}$$

  To simplify notation, first write (293) in terms of the current J.

  $$\mathbf{D}^\beta F_{\alpha\beta} = J_\alpha = -\left[A^\beta, F_{\alpha,\beta}\right] \tag{295}$$

  When $\alpha = 0$ (294) allows us to write (295) as

  $$J_0 = \partial^i F_{0i} = \partial^i\left(\partial_0 A_i - \partial_i A_0 + [A_0, A_i]\right) - \Delta A_0 + \partial^i [A_0, A_i]$$

  giving us for the time component of A:

  $$\Delta A_0 = 2\left[\partial_i A_0, A_i\right] + \left[A_0, \partial_t A_i\right] + \left[A_i, [A_0, A_j]\right] \tag{296}$$

  When $\alpha = i$, (295) reads

  $$J_i = -\partial_t + \partial^j F_{ij} = -\partial_t\left(\partial_i A_0 + [A_i, A_0]\right) + \partial^j\left(\partial_i A_j - \partial_j A_i + [A_i, A_j]\right)$$

  and after simplifying,

  $$\begin{aligned}
  \Box A_i &= -\partial_t \partial_i A_0 - 2\left[A_j, \partial_j A_i\right] + \left[A_j, \partial_i A_j\right] + \left[\partial_t A_i, A_j\right] \\
  &\quad + 2\left[A_0, \partial_t A_i\right] - \left[A_0, \partial_i A_0\right] - \left[A_j, [A_j, A_i]\right] + \left[A_0, [A_0, A_i]\right]
  \end{aligned} \tag{297}$$

- **Lorentz Gauge** is specified by,

  $$\partial^\mu A_\mu(t,x) = 0 \quad (t,x) \in \mathbb{R}^{3+1} \tag{298}$$

Appealing in its symmetric treatment of the time and space components of A, , the Lorentz gauge also allows (293) to be written as a system of wave equations:

$$\begin{aligned}\mathbf{D}^\beta F_{\alpha\beta} &= \mathbf{D}^\beta\left(\partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta]\right)\\ &= -\Box A_\alpha + \partial^\beta[A_\alpha, A_\beta] + [A_\beta, \partial_\alpha A_\beta] - [A^\beta, \partial_\beta A_\alpha] + \big[A_\beta, [A_\alpha, A^\beta]\big]\end{aligned}$$

The system can be written schematically in the form

$$\Box\Phi \quad = \quad \Phi\cdot\partial\Phi + \Phi^3$$

Again, it is not at all clear that one can transform an arbitrary solution into the Lorentz gauge. In addition, we will have a hard time finding good estimates for this purely hyperbolic system of nonlinear wave equations.

- **Temporal Gauge** is specified by the condition $A_0 = 0$.

5. The Einstein Field Equations:

According to the general relativistic variational principle the space-time metric $\mathbf{g}$ is itself stationary relative to an action,

$$\mathcal{S} = \int_{\mathcal{U}} L dv_{\mathbf{g}}.$$

Here U is a relatively compact domain of $(\mathbf{M}, \mathbf{g})$ and L, the Lagrangian, is assumed to be a scalar function on $\mathbf{M}$ whose dependence on the metric should involve no more than two derivatives[5]. It is also assumed to depend on the matterfields $\psi = \psi^{(1)}, \psi^{(2)}, \dots \psi^{(p)}$ present in our space-time.

In fact we write,

$$\mathcal{S} = \mathcal{S}_G + \mathcal{S}_M$$

with,

$$\begin{aligned}\mathcal{S}_G &= \int_{\mathcal{U}} L_G dv_{\mathbf{g}}\\ \mathcal{S}_M &= \int_{\mathcal{U}} L_M dv_{\mathbf{g}}\end{aligned}$$

denoting, respectively, the actions for the gravitational field and matter. The matter Lagrangian L_M depends only on the matterfields ψ, assumed to be covariant tensorfields, and the inverse of the space-time metric $\mathbf{g}^{\alpha\beta}$ which appears in the contraction of the tensorfields ψ in order to produce the scalar L_M. It may also depend on additional positive definite metrics which are not to be varied [6].

---

[5]In fact we only require that the corrsponding Euler-Lagrange equations should involve no more than two derivatives of the metric.

[6]This is the case of the metric h in the case of wave maps or the Killing scalar product in the case of the Yang-Mills equations.

Now the only possible candidate for the gravitational Lagrangian L_G, which should be a scalar invariant of the metric with the property that the corresponding Euler-Lagrange equations involve at most two derivatives of the metric, is given[7] by the scalar curvature $\mathbf{R}$. Therefore we set,

$$L_G = \mathbf{R}.$$

Consider now a compact variation $(\mathbf{g}_{(s)}, \mathcal{U})$ of the metric $\mathbf{g}$. Let $\dot{\mathbf{g}}_{\mu\nu} = \frac{d}{ds}\mathbf{g}_{\mu\nu}|_{s=0}$. Thus for small s, $\mathbf{g}_{\mu\nu}(s) = \mathbf{g}_{\mu\nu} + s\dot{\mathbf{g}}_{\mu\nu} + O(s^2)$. Also, $\mathbf{g}^{\mu\nu}(s) = \mathbf{g}^{\mu\nu} - s\dot{\mathbf{g}}^{\mu\nu} + O(s^2)$ where $\dot{\mathbf{g}}^{\mu\nu} = \mathbf{g}^{\alpha\mu}\mathbf{g}^{\beta\nu}\dot{\mathbf{g}}_{\alpha\beta}$. Then,

$$\frac{d}{ds}\mathbf{S}_G(s)\Big|_{s=0} = \int_{\mathcal{U}} \dot{\mathbf{R}} dv_{\mathbf{g}} + \int_{\mathcal{U}} \mathbf{R} \dot{dv_{\mathbf{g}}}$$

Now,

$$\dot{dv_{\mathbf{g}}} = \frac{1}{2}\mathbf{g}^{\mu\nu}\dot{\mathbf{g}}_{\mu\nu} dv_{\mathbf{g}}$$

Indeed, relative to a coordinate system, $dv_{\mathbf{g}} = \sqrt{-\mathbf{g}} dx^0 dx^1 \ldots dx^n$ Thus, the above equality follows from,

$$\dot{\mathbf{g}} = \mathbf{g}\mathbf{g}^{\alpha\beta}\dot{\mathbf{g}}_{\alpha\beta},$$

with $\mathbf{g}$ the determinant of $\mathbf{g}_{\alpha\beta}$. On the other hand, writing $\mathbf{R} = \mathbf{g}^{\mu\nu}\mathbf{R}_{\mu\nu}$ and using the formula $\frac{d}{ds}\mathbf{g}^{\mu\nu}_{(s)}\Big|_{s=0} = -\dot{\mathbf{g}}^{\mu\nu}$, we calculate, $\dot{\mathbf{R}} = -\dot{\mathbf{g}}^{\mu\nu}\mathbf{R}_{\mu\nu} + \mathbf{g}^{\mu\nu}\dot{\mathbf{R}}_{\mu\nu}$. Therefore,

$$\frac{d}{ds}\mathbf{S}_G(s)\Big|_{s=0} = -\int_{\mathcal{U}} (\mathbf{R}^{\mu\nu} - \frac{1}{2}\mathbf{g}^{\mu\nu}\mathbf{R})\dot{\mathbf{g}}_{\mu\nu} dv_{\mathbf{g}} + \int_{\mathcal{U}} \mathbf{g}^{\mu\nu}\dot{\mathbf{R}}_{\mu\nu} dv_{\mathbf{g}} \tag{299}$$

To calculate $\dot{\mathbf{R}}_{\mu\nu}$ we make use of the following Lemma,

Lemma 1.2. *Let $\mathbf{g}_{\mu\nu}(s)$ be a family of space-time metrics with $\mathbf{g}(0) = \mathbf{g}$ and $\frac{d}{ds}\mathbf{g}(0) = \dot{\mathbf{g}}$. Set also, $\frac{d}{ds}\mathbf{R}_{\alpha\beta}(s)|_{s=0} = \dot{\mathbf{R}}_{\alpha\beta}$. Then,*

$$\dot{\mathbf{R}}_{\mu\nu} = \mathbf{D}_\alpha \dot{\Gamma}^\alpha_{\mu\nu} - \mathbf{D}_\mu \dot{\Gamma}^\alpha_{\alpha\nu}$$

where $\dot{\Gamma}$ is the tensor,

$$\dot{\Gamma}^\alpha_{\beta\gamma} = \frac{1}{2}\mathbf{g}^{\alpha\lambda}(\mathbf{D}_\beta\dot{\mathbf{g}}_{\gamma\lambda} + \mathbf{D}_\gamma\dot{\mathbf{g}}_{\beta\lambda} - \mathbf{D}_\lambda\dot{\mathbf{g}}_{\beta\gamma})$$

Proof : Since both sides of the identity are tensors it suffices to prove the formula at a point p relative to a particular system of coordinates for which the Christoffel symbols Γ vanish at p. Relative to such a coordinate system the Ricci tensor has the form $\mathbf{R}_{\mu\nu} = \mathbf{D}_\alpha\Gamma^\alpha_{\mu\nu} - \mathbf{D}_\mu\Gamma^\alpha_{\alpha\nu}$. ■

Returning to (299) we find that since $\mathbf{g}^{\mu\nu}\dot{\mathbf{R}}_{\mu\nu}$ can be written as a space-time divergence of a tensor compactly supported in U the corresponding integral vanishes identically. We therefore infer that,

$$\frac{d}{ds}\mathbf{S}_G(s)\Big|_{s=0} = -\int_U \mathbf{E}^{\mu\nu}\dot{\mathbf{g}}_{\mu\nu} dv_{\mathbf{g}} \tag{300}$$

[7] up to an additive constant

where $\mathbf{E}^{\mu\nu} = \mathbf{R}^{\mu\nu} - \frac{1}{2}\mathbf{g}^{\mu\nu}\mathbf{R}$. We now consider the variation of the action integral $\mathbf{S}_M$ with respect to the metric. As remarked before L_M depends on the metric $\mathbf{g}$ through its inverse $\mathbf{g}^{\mu\nu}$. Therefore if we denote $\mathbf{S}_M(s) = \mathbf{S}_M[\psi, \mathbf{g}_{(s)}; \mathcal{U}]$ we have, writing $d\dot{v}_{\mathbf{g}} = \frac{1}{2}\mathbf{g}_{\mu\nu}\dot{\mathbf{g}}^{\mu\nu}dv_{\mathbf{g}}$,

$$\begin{aligned}\frac{d}{ds}\mathbf{S}_M(s)\Big|_{s=0} &= -\int_{\mathcal{U}} \frac{\partial L_M}{\partial \mathbf{g}^{\mu\nu}}\dot{\mathbf{g}}^{\mu\nu}dv_{\mathbf{g}} + \int_{\mathcal{U}} L_M d\dot{v}_{\mathbf{g}} \\ &= -\int_{\mathcal{U}} (\frac{\partial L_M}{\partial \mathbf{g}^{\mu\nu}} - \frac{1}{2}\mathbf{g}_{\mu\nu}L_M)\dot{\mathbf{g}}^{\mu\nu}dv_{\mathbf{g}}\end{aligned}$$

Definition. The symmetric tensor,

$$\mathbf{T}_{\mu\nu} = -\left(\frac{\partial L_M}{\partial \mathbf{g}^{\mu\nu}} - \frac{1}{2}\mathbf{g}_{\mu\nu}L_M\right)$$

is called the *energy-momentum tensor* of the action $\mathbf{S}_M$.

With this definition we write,

$$\frac{d}{ds}\mathbf{S}_M(s)\Big|_{s=0} = \int_{\mathcal{U}} \mathbf{T}^{\mu\nu}\dot{\mathbf{g}}_{\mu\nu}dv_{\mathbf{g}} \tag{301}$$

Finally, combining 300 with 301, we derive for the total action $\mathbf{S}$,

$$\frac{d}{ds}\mathbf{S}(s)\Big|_{s=0} = -\int_{\mathcal{U}} (\mathbf{E}^{\mu\nu} - \mathbf{T}^{\mu\nu})\dot{\mathbf{g}}_{\mu\nu}dv_{\mathbf{g}}$$

Since $\dot{\mathbf{g}}_{\mu\nu}$ is an arbitrary symmetric 2-tensor compactly supported in U we derive the *Einstein field equation*,

$$\mathbf{E}^{\mu\nu} = \mathbf{T}^{\mu\nu}$$

Recall that the Einstein tensor $\mathbf{E}$ satisfies the twice contracted Bianchi identity,

$$\mathbf{D}^{\nu}\mathbf{E}_{\mu\nu} = 0$$

This implies that the energy-momentum tensor $\mathbf{T}$ is also divergenceless,

$$\mathbf{D}_{\nu}\mathbf{T}^{\mu\nu} = 0 \tag{302}$$

which is the concise, space-time expression for the law of conservation of energy-momentum of the matter-fields.

2. The energy-momentum tensor

The conservation law (302) is a fundamental property of a matterfield. We now turn to a more direct derivation.

We consider an arbitrary Lagrangian field theory with stationary solution ψ. Let Φ_s be the one-parameter group of local diffeomorphisms generated by a given vectorfield X. We shall use the flow Φ to vary the fields ψ according to

$$\begin{aligned}\mathbf{g}_s &= (\Phi_s)_*\mathbf{g}\\ \psi_s &= (\Phi_s)_*\psi.\end{aligned}$$

From the invariance of the action integral under diffeomorphisms,

$$\mathbf{S}(s) = \mathbf{S}[\psi_s, \mathbf{g}_s; \mathbf{M}] = \mathbf{S}_M[\psi, \mathbf{g}; \mathbf{M}].$$

So that

$$0 = \frac{d}{ds}\mathbf{S}(s)\Big|_{s=0} = \int_{\mathbf{M}} \frac{\delta \mathbf{S}}{\delta\psi} dv_{\mathbf{g}} + \int_{\mathbf{M}} \mathbf{T}^{\mu\nu}\dot{\mathbf{g}}_{\mu\nu} dv_{\mathbf{g}} \tag{303}$$

The first term is clearly zero, ψ being a stationary solution. In the second term, which represents variations with respect to the metric, we have

$$\dot{\mathbf{g}}_{\mu\nu} = \frac{d}{ds}(\mathbf{g}_s)_{\mu\nu}\Big|_{s=0} = \mathcal{L}_X \mathbf{g}_{\mu\nu} = \mathbf{D}_\mu X_\nu + \mathbf{D}_\nu X_\mu$$

Therefore

$$0 = \int_{\mathbf{M}} \mathbf{T}^{\mu\nu}\mathcal{L}_X\mathbf{g}_{\mu\nu} dv_{\mathbf{g}} = 2\int_{\mathbf{M}} \mathbf{T}^{\mu\nu}\mathbf{D}_\nu X_\mu dv_{\mathbf{g}} = -2\int_{\mathbf{M}} \mathbf{D}_\nu \mathbf{T}^{\mu\nu} X_\mu dv_{\mathbf{g}}$$

As X was arbitrary, we conclude

$$\mathbf{D}_\nu \mathbf{T}^{\mu\nu} = 0. \tag{304}$$

This is again the law of conservation of energy-momentum.

We list below the energy-momentum tensors of the field theories discussed before. We leave it to the reader to carry out the calculations using the definition.

(1) The energy-momentum for the scalar field equation is,

$$\mathbf{T}_{\alpha\beta} = \frac{1}{2}\left(\phi_{,\alpha}\phi_{,\beta} - \frac{1}{2}\mathbf{g}_{\alpha\beta}(\mathbf{g}^{\mu\nu}\phi_{,\mu}\phi_{,\nu} + 2V(\phi))\right)$$

(2) The energy-momentum for wave maps is given by,

$$\mathbf{T}_{\alpha\beta} = \frac{1}{2}\left(< \phi_{,\alpha}, \phi_{,\beta} > - \frac{1}{2}\mathbf{g}_{\alpha\beta}(\mathbf{g}^{\mu\nu} < \phi_{,\mu}, \phi_{,\nu} >)\right)$$

where $< \, , \, >$ denotes the Riemannian inner product on the target manifold.

(3) The energy-momentum tensor for the Maxwell equations is,

$$\mathbf{T}_{\alpha\beta} = F_\alpha^{\cdot\mu} F_{\beta\mu} - \frac{1}{4}\mathbf{g}_{\alpha\beta}(F_{\mu\nu}F^{\mu\nu})$$

(4) The energy-momentum tensor for the Yang-Mills equations is,

$$\mathbf{T}_{\alpha\beta} = < F_\alpha^{\cdot\mu}, F_{\beta\mu} > - \frac{1}{4}\mathbf{g}_{\alpha\beta}(< F_{\mu\nu}, F^{\mu\nu} >)$$

An acceptable notion of the energy-momentum tensor $\mathbf{T}$ must satisfy the following properties in addition of the conservation law (304),

(1) $\mathbf{T}$ is symmetric
(2) $\mathbf{T}$ satisfies the *positive energy condition* that is, $\mathbf{T}(X,Y) \geq 0$, for any future directed time-like vectors X, Y.

The symmetry property is automatic in our construction. The following proposition asserts that the energy-momentum tensors of the field theories described above satisfy the positive energy condition.

Proposition 2.1. *The energy-momentum tensor of the scalar wave equation satisfies the positive energy condition if V is positive. The energy- momentum tensors for the wave maps, Maxwell equations and Yang-Mills satisfy the positive energy condition.*

Proof : To prove the positivity conditions consider two vectors X, Y, at some point $p \in \mathbf{M}$, which are both causal future oriented. The plane spanned by X, Y intersects the null cone at p along two null directions[8]. Let $L, \underline{L}$ be the two future directed null vectors corresponding to the two complementary null directions and normalized by the condition

$$< L, \underline{L} >= -2$$

i.e. they form a null pair. Since the vectorfields X, Y are linear combinations with positive coefficients of $L, \underline{L}$, the proposition will follow from showing that $\mathbf{T}(L,L) \geq 0$, $\mathbf{T}(\underline{L},\underline{L}) \geq 0$ and $\mathbf{T}(L,\underline{L}) \geq 0$. To show this we consider a frame at p formed by the vectorfields $E_{(n+1)} = L$, $E_{(n)} = \underline{L}$ and $E_{(1)}, \ldots, E_{(n-1)}$ with the properties,

$$< E_{(i)}, E_{(n)} >=< E_{(i)}, E_{(n+1)} >= 0$$

and

$$< E_{(i)}, E_{(j)} >= \delta_{ij}$$

for all $i, j = 1, \ldots, n-1$. A frame with these properties is called a *null frame.*

(1) We now calculate, in the case of the wave equation,

$$\begin{aligned} \mathbf{T}(L,L) &= \frac{1}{2} E(\phi)^2 \\ \mathbf{T}(\underline{L},\underline{L}) &= \frac{1}{2} \underline{L}(\phi)^2. \end{aligned}$$

which are clearly non-negative. Now,

$$\mathbf{T}(L,\underline{L}) = \frac{1}{2}\left[L(\phi)\underline{L}(\phi) + (g^{\mu\nu}\phi_{,\mu}\phi_{,\nu} + 2V(\phi))\right]$$

and we aim to express $g^{\mu\nu}\phi_{,\mu}\phi_{,\nu}$ relative to our null frame. To do this, observe that relative to the null frame the only nonvanishing components of the metric $g_{\alpha\beta}$ are,

$$g_{n(n+1)} = -2 \quad , \quad g_{ii} = 1 \;\; i = 1, \ldots, n-1$$

and those of the inverse metric $g^{\alpha\beta}$ are

$$g^{n(n+1)} = -\frac{1}{2} \quad , \quad g^{ii} = 1 \;\; i = 1, \ldots, n-1$$

[8] If X, Y are linearly dependent any plane passing through their common direction will do.

Therefore,

$$g^{\mu\nu}\phi_{,\mu}\phi_{,\nu} = -L(\phi)\underline{L}(\phi) + |\nabla\!\!\!/\phi|^2$$

where

$$|\nabla\!\!\!/\phi|^2 = (E_{(1)}(\phi))^2 + (E_{(2)}(\phi))^2 + \ldots E_{(n-1)}(\phi)^2.$$

Therefore,

$$\mathbf{T}(L, \underline{L}) = \frac{1}{2}|\nabla\!\!\!/\phi|^2 + V(\phi).$$

(2) For wave maps we have, according to the same calculation.

$$\begin{aligned} T(E,E) &= \frac{1}{2} < E(\phi), E(\phi) > \\ T(\underline{E},\underline{E}) &= \frac{1}{2} < \underline{E}(\phi), \underline{E}(\phi) > \\ T(E,\underline{E}) &= \frac{1}{2}\sum_{i=1}^{n-1} < E_{(i)}(\phi), E_{(i)}(\phi) > . \end{aligned}$$

The positivity of T is then a consequence of the Riemannian metric h on the target manifold N.

(3) To show positivity for the energy momentum tensor of the Maxwell equations in $3+1$ dimensions we first write the tensor in the more symmetric form

$$T_{\alpha\beta} = \frac{1}{2}\left(F_{\alpha}{}^{\mu}F_{\beta\mu} + {}^{\star}F_{\alpha}{}^{\mu}\,{}^{\star}F_{\beta\mu}\right) \tag{305}$$

where ${}^{\star}F$ is the Hodge dual of F, i.e. ${}^{\star}F_{\alpha\beta} = \frac{1}{2}\in_{\alpha\beta\mu\nu} F^{\mu\nu}$.

Exercise. Check formula (305).

We introduce the following null decomposition of F at every point $p \in \mathbf{M}$,

$$\begin{aligned} \alpha_A = F_{A4} \qquad &, \qquad \underline{\alpha}_A = F_{A3} \\ \rho = \frac{1}{2}F_{34} \qquad &, \qquad \sigma = \frac{1}{2}\,{}^{\star}F_{34}. \end{aligned}$$

which completely determines the tensor F. Here the indices $A = 1, 2$ correspond to the directions E_1, E_2 tangent to the sphere while the indices $3, 4$ correspond to $E_3 = \underline{L}$ and $E_4 = L$. We then calculate that for ${}^{\star}F$,

$$\begin{aligned} {}^{\star}F_{A4} = -\,{}^{\star}\alpha_A = \qquad &, \qquad {}^{\star}F_{A3} = {}^{\star}\underline{\alpha}_A \\ {}^{\star}F_{34} = 2\sigma \qquad &, \qquad {}^{\star\star}F_{34} = -2\rho \end{aligned}$$

where ${}^{\star}\alpha_A = \in_{AB} \alpha_B$. Here $\in_{AB}$ is the volume form on the unit sphere, hence $\in_{AB} = \frac{1}{2}\in_{AB34}$, i.e. $\in_{11} = \in_{22} = 0$, $\in_{12} = -\epsilon_{21} = 1$. With this notation we calculate,

$$\begin{aligned} T(E_{(4)}, E_{(4)}) &= \frac{1}{2}\sum_{A=1}^{2}\left(F_{4A}\cdot F_{4A} + \frac{1}{4}\,{}^{\star}F_{4A}\cdot {}^{\star}F_{4A}\right) \\ &= \frac{1}{2}\sum_{A=1}^{2}\left(\alpha_A\cdot\alpha_A + {}^{\star}\alpha_A\cdot {}^{\star}\alpha_A\right) \\ &= \sum_{A=1}^{2}\alpha_A\cdot\alpha_A = |\alpha|^2 \geq 0. \end{aligned}$$

Similarly,

$$T(E_{(3)}, E_{(3)}) = \sum_{A=1}^{2} \underline{\alpha}_A \cdot \underline{\alpha}_A = |\underline{\alpha}|^2 \geq 0$$

and in the same vein we find

$$T(E, \underline{E}) = \rho^2 + \sigma^2 \geq 0$$

which proves our assertion.

(4) The positivity of the energy-momentum tensor of the Yang- Mills equations is proved in precisely the same manner as for the Maxwell equations, using the positivity of the Killing scalar product $< \cdot, \cdot >_{\mathcal{G}}$.

■

Another important property which the energy momentum tensor of a field theory may satisfy is the trace free condition, that is

$$\mathbf{g}_{\alpha\beta}\mathbf{T}^{\alpha\beta} = 0.$$

It turns out that this condition is satisfied by all field theories which are conformally invariant.

Definition. A field theory is said to be *conformally invariant* if the corresponding action integral is invariant under conformal transformations of the metric

$$\mathbf{g}_{\alpha\beta} \longrightarrow \tilde{\mathbf{g}}_{\alpha\beta} = \Omega \mathbf{g}_{\alpha\beta}$$

Ω a positive smooth function on the space-time.

Proposition 2.2. *The energy momentum tensor* $\mathbf{T}$ *of a conformally invariant field theory is traceless.*

Proof: Consider an arbitrary smooth function f compactly supported in $\mathcal{U} \subset \mathcal{M}$. Consider the following variation of a given metric $\mathbf{g}$,

$$\mathbf{g}_{\mu\nu}(s) = e^{sf}\mathbf{g}_{\mu\nu}.$$

Let $\mathcal{S}(s) = \mathcal{S}_{\mathcal{U}}[\psi, \mathbf{g}(s)]$. In view of the covariance of $\mathcal{S}$ we have $\mathcal{S}(s) = \mathcal{S}(0)$. Hence,

$$0 = \frac{d}{ds}\mathcal{S}(s)|_{s=0} = \int_{\mathcal{U}} T^{\mu\nu}\dot{\mathbf{g}}_{\mu\nu} dv_{\mathbf{g}}$$

where

$$\dot{\mathbf{g}}_{\mu\nu} = \frac{d}{ds}\mathbf{g}_{\mu\nu}(s)\bigg|_{s=0} = f\mathbf{g}_{\mu\nu}.$$

Hence, $\int_{\mathcal{U}} (T^{\mu\nu}\mathbf{g}_{\mu\nu}) f dv_{\mathbf{g}} = 0$ and since f is arbitrary we infer that,

$$trT = g^{\mu\nu}T_{\mu\nu} \equiv 0.$$

■

We can easily check that the Maxwell and the Yang-Mills equations are conformally invariant in 3×1-dimensions. The wave maps field theory is conformally invariant in dimension $1+1$, i.e. if the space-time $\mathcal{M}$ is two-dimensional[9].

Remark: The action integral of the Maxwell equations, $\mathbf{S} = \int_{\mathcal{U}} F_{\alpha\beta} F^{\alpha\beta} dv_{\mathbf{g}}$ is conformally invariant in any dimension provided that we also scale the electromagnetic field F. Indeed if $\tilde{\mathbf{g}}_{\alpha\beta} = \Omega^2 \mathbf{g}_{\alpha\beta}$ then $dv_{\tilde{\mathbf{g}}} = \Omega^{n+1} dv_{\mathbf{g}}$ and if we also set $\tilde{F}_{\alpha\beta} = \Omega^{-\frac{n-3}{2}} F_{\alpha\beta}$ we get

$$
\begin{aligned}
\tilde{\mathbf{S}}[\tilde{F}, \tilde{\mathbf{g}}] &= \int \tilde{F}_{\alpha\beta} \tilde{F}_{\gamma\delta} \tilde{\mathbf{g}}^{\alpha\gamma} \tilde{\mathbf{g}}^{\beta\delta} dv_{\tilde{\mathbf{g}}} \\
&= \int F_{\alpha\beta} F_{\gamma\delta} \mathbf{g}^{\alpha\gamma} \mathbf{g}^{\beta\delta} dv_{\mathbf{g}} \\
&= \mathbf{S}[F, \mathbf{g}].
\end{aligned}
$$

We finish this section with a simple observation concerning conformal field theories in $1+1$ dimensions. We specialize in fact to the Minkowski space $\mathbb{R}^{1+1}$ and consider the local conservation law, $\partial^\mu \mathbf{T}_{\nu\mu} = 0$. Setting $\nu = 0, 1$ we derive

$$
\partial^0 \mathbf{T}_{00} + \partial^1 \mathbf{T}_{01} = 0, \quad \partial^0 \mathbf{T}_{01} + \partial^1 \mathbf{T}_{11} = 0 \tag{306}
$$

Since the energy-momentum tensor is trace-free, we get $\mathbf{T}_{00} = \mathbf{T}_{11} = A$, say. Set $\mathbf{T}_{01} = \mathbf{T}_{10} = B$. Therefore (**??**) implies that both A and B satisfy the linear homogeneous wave equation;

$$
\Box A \;=\; 0 \;=\; \Box B. \tag{307}
$$

Using this observation it is is easy to prove that smooth initial data remain smooth for all time.

For example, wave maps are conformally invariant in dimension $1+1$. In this case

$$
A \;=\; \mathbf{T}_{00} \;=\; \frac{1}{2} \left(< \partial_t \phi, \partial_t \phi > + < \partial_x \phi, \partial_x \phi > \right),
$$

Given data in $C_0^\infty(\mathbb{R})$, (307) implies that the derivatives of ϕ remain smooth for all positive times. This proves global existence.

3. Conservation Laws

The energy-momentum tensor of a field theory is intimately connected with conservations laws. This connection is seen through Noether's principle,

Noether's Principle: *To any one-parameter group of transformations preserving the action there corresponds a conservation law.*

[9] Similarly for the linear scalar wave equation

We illustrate this fundamental principle as follows: Let $\mathbf{S} = \mathbf{S}[\psi, \mathbf{g}]$ be the action integral of the fields ψ. Let χ_t be a 1-parameter group of isometries of $\mathbf{M}$, i.e., $(\chi_t)_*\mathbf{g} = \mathbf{g}$. Then

$$\begin{aligned} \mathbf{S}[(\chi_t)_*\psi, \mathbf{g}] &= \mathbf{S}[(\chi_t)_*\psi, (\chi_t)_*\mathbf{g}] \\ &= \mathbf{S}[\psi, \mathbf{g}]. \end{aligned}$$

Thus the action is preserved under $\psi \to (\chi_t)_*\psi$. In view of Noether's Principle we ought to find a conservation law for the corresponding Euler-Lagrange equations[10]. We derive these laws using the Killing vectorfield X which generates χ_t.

We begin with a general calculation involving the energy-momentum tensor $\mathbf{T}$ of ψ and an arbitrary vectorfield X. P the one-form obtained by contracting $\mathbf{T}$ with X.

$$P_\alpha = \mathbf{T}_{\alpha\beta} X^\beta$$

Since $\mathbf{T}$ is symmetric and divergence-free

$$\mathbf{D}^\alpha P_\alpha = \left(\mathbf{D}^\alpha \mathbf{T}_{\alpha\beta}\right) X^\beta + \mathbf{T}_{\alpha\beta}\left(\mathbf{D}^\alpha X^\beta\right) = \frac{1}{2}\mathbf{T}^{\alpha\beta}\ {}^{(X)}\pi_{\alpha\beta}$$

where ${}^{(X)}\pi_{\alpha\beta}$ is the deformation tensor of X.

$${}^{(X)}\pi_{\alpha\beta} = \left(\mathcal{L}_X \mathbf{g}\right)_{\alpha\beta} = \mathbf{D}_\alpha X_\beta + \mathbf{D}_\beta X_\alpha$$

Notation. We denote the backward light cone with vertex $p = (\bar{t}, \bar{x}) \in \mathbb{R}^{n+1}$ by

$$\mathcal{N}^-(\bar{t}, \bar{x}) = \{(t, x) \,\big|\, 0 \le t \le \bar{t}; |x - \bar{x}| = \bar{t} - t\}.$$

The restriction of this set to some time interval $[t_1, t_2]$, $t_1 \le t_2 \le \bar{t}$, will be written $\mathcal{N}^-_{[t_1,t_2]}(\bar{t}, \bar{x})$. These null hypersurfaces are null boundaries of,

$$\begin{aligned} \mathcal{J}^{-1}(\bar{t}, \bar{x}) &= \{(t, x) \,\big|\, 0 \le t \le \bar{t};\ |x - \bar{x}| \le \bar{t} - t\} \\ \mathcal{J}^-_{[t_2,t_1]}(\bar{t}, \bar{x}) &= \{(t, x) \,\big|\, t_2 \le t \le t_1;\ |x - \bar{x}| \le \bar{t} - t\} \end{aligned}$$

We shall denote by $S_t = S_t(\bar{t}, \bar{x})$ and $B_t = B_t(\bar{t}, \bar{x})$ the intersection of the time slice Σ_t with $\mathcal{N}^-$, respectively $\mathcal{J}^-$.

At each point $q = (t, x)$ along $\mathcal{N}^-(p)$, we define the null pair (E_+, E_-) of future oriented null vectors

$$\underline{L} = E_+ = \partial_t + \frac{x^i - \bar{x}^i}{|x - \bar{x}|}\partial_i, \qquad L = E_- = \partial_t - \frac{x^i - \bar{x}^i}{|x - \bar{x}|}\partial_i$$

Observe that both $L, \underline{L}$ are null and $< L, \underline{L} >= -2$.

The following is a simple consequence of Stoke's theorem, in the following form.

---

[10] The same argument holds for conformal isometries acting on a conformally invariant field theory. We therefore also expect conservation laws in such a setting.

Proposition 3.1. *Let P_μ be a one-form satisfying $\partial^\mu P_\mu = F$. Then*[11]*, for all $t_1 \leq t_2 \leq \bar{t}$,*

$$\int_{B_{t_2}} \langle P, \partial_t \rangle + \int_{\mathcal{N}^-_{[t_1,t_2]}(p)} \langle P, E_- \rangle = \int_{B_{t_1}} \langle P, \partial_t \rangle - \int_{\mathcal{J}^-_{[t_1,t_2]}(p)} F dt dx \quad (308)$$

where,

$$\int_{\mathcal{N}^-_{[t_1,t_2]}(p)} \langle P, E_- \rangle = \int_{t_1}^{t_2} dt \int_{S_t} \langle P, E_- \rangle \, da_t.$$

Applying this proposition to Stoke's theorem to (308) we get

Theorem 3.2. *Let T be the energy-momentum tensor associated to a field theory and X an arbitrary vector field. Then*

$$\begin{aligned}\int_{B_{t_2}} \mathbf{T}(\partial_t, X) + \int_{\mathcal{N}^-_{[t_1,t_2]}(p)} \mathbf{T}(E_-, X) &= \int_{B_{t_1}} \mathbf{T}(\partial_t, X) \\ &- \int_{\mathcal{J}^-_{[t_1,t_2]}(p)} \mathbf{T}^{\alpha\beta\ (X)}\pi_{\alpha\beta} dt dx\end{aligned} \quad (309)$$

In the particular case when X is Killing, its deformation tensor π vanishes identically. Thus,

Corollary 3.3. *If X is a killing vectorfield,*

$$\int_{B_{t_2}} \mathbf{T}(\partial_t, X) + \int_{\mathcal{N}^-_{[t_1,t_2]}(p)} \mathbf{T}(L, X) = \int_{B_{t_1}} \mathbf{T}(\partial_t, X) \quad (310)$$

Moreover (310) *remains valid if* $\mathbf{T}$ *is traceless and X is conformal Killing.*

The identity (310) is usually applied to time-like future-oriented Killing vectorfields X in which case the positive energy condition for $\mathbf{T}$ insures that all integrands in (310) will be positive. We know that (see appendix 4.2, up to a Lorentz transformation the only Killing, future oriented timelike vectorfield is a constant multiple of ∂_t. Choosing $X = \partial_t$ (310) becomes,

$$\int_{B_{t_2}} \mathbf{T}(\partial_t, \partial_t) + \int_{\mathcal{N}^-_{[t_1,t_2]}(p)} \mathbf{T}(E_-, \partial_t) = \int_{B_{t_1}} \mathbf{T}(\partial_t, \partial_t) \quad (311)$$

In the case of a conformal field theory we can pick X to be the future timelike, conformal Killing vectorfield $X = K_0 = (t^2 + |x|^2)\partial_t + 2tx^i\partial_i$. Thus,

$$\int_{B_{t_2}} \mathbf{T}(\partial_t, K_0) + \int_{\mathcal{N}^-_{[t_1,t_2]}(p)} \mathbf{T}(L, K_0) = \int_{B_{t_1}} \mathbf{T}(\partial_t, K_0) \quad (312)$$

In (311) the term $\mathbf{T}(\partial_t, \partial_t)$ is called *energy density* while $\mathbf{T}(E_-, \partial_t)$ is called *energy flux density.* The corresponding integrals are called *energy contained in B_{t_1}, and*

[11] The brackets $\langle \cdot, \cdot \rangle$ in (308) denote inner product with respect to the Minkowski metric.

B_{t_2} and, respectively, *flux of energy through* $\mathcal{N}^-$. The coresponding terms in (312) are called conformal energy densities, fluxes etc.

Equation (311) can be used to derive the following fundamental properties of relativistic field theories.

(1) Finite propagation speed
(2) Uniqueness of the Cauchy problem

Proof : The first property follows from the fact that, if $\int_{B_{t_1}} \mathbf{T}(\partial_t, \partial_t)$ is zero at time $t = t_1$ then both integrals $\int_{B_{t_2}} \mathbf{T}(\partial_t, \partial_t)$ and $\int_{\mathcal{N}^-_{[t_1,t_2]}} \mathbf{T}(E_-, \partial_t)$ must vanish also. In view of the positivity properties of the $\mathbf{T}$ it follows that the corresponding integrands must also vanish. Taking into account the specific form of $\mathbf{T}$, in a particular theory, one can then show that the fields do also vanish in the domain of influence of the ball B_{t_1}. Conversely, if the initial data for the fields vanish in the complement of B_{t_1}, the the fields are identically zero in the complement of the domain of influence of of B_{t_1}.

The proof of the second property follows immediately from the first for a linear field theory. For a nonlinear theory one has to work a little more. ■

Exercise 1. Formulate an initial value problem for each of the field theories we have encountered so far, scalar wave equation (SWE), Wave Maps (WM), Maxwell equations (ME) and Yang-Mills (YM). Proof uniqueness of solutions to the initial value problem, for smooth solutions.

The following is another important consequence of (311) and (312). To state the results we introduce the following quantities,

$$\mathcal{E}(t) = \int_{\mathbb{R}^n} \mathbf{T}\left(\partial_t, \partial_t\right)(t,x)dx \tag{313}$$

$$\mathcal{E}_c(t) = \int_{\mathbb{R}^n} \mathbf{T}\left(K_0, \partial_t\right)(t,x)dx \tag{314}$$

Theorem 3.4 (Global Energy)**.** *For an arbitrary field theory, if $\mathcal{E}(0) < \infty$, then*

$$\mathcal{E}(t) = \mathcal{E}(0) \tag{315}$$

Moreover, for a conformal field theory, if $\mathcal{E}_c(0) < \infty$,

$$\mathcal{E}_c(t) = \mathcal{E}_c(0) \tag{316}$$

Proof : Follows easily by applying (311) and (312) to past causal domains $\mathcal{J}^-(p)$ with $p = (\bar{t}, 0)$ between $t_1 = 0$ and $t_2 = t$ and letting $\bar{t} \to +\infty$. ■

Exercise 2. Consider the Lagrangian,

$$L = -\frac{1}{2}\mathbf{m}^{\alpha\beta}\partial_\alpha\phi\overline{\partial_\beta\phi} - V(|\phi|)$$

where ϕ is a complex valued scalar defined on the Minkowski space-time $\mathbb{R}^{n+1}$, $\bar{\phi}$ its complex conjugate. As noted before L is invariant under the continuous group of transformations $\phi \to e^{i\theta}\phi$ with $\theta \in \mathbb{R}$. According to Noether's theorem the corresponding Euler-Lagrange equation should have a conservation law. Can you derive it ?

3.5. Energy dispersion in the conformal invariant case. In this section we shall make use of the global conformal energy identity (316) to show how energy dissipates for a filed theories in Minkowski space. Consider a conformal field theory defined on all of $\mathbb{R}^{n+1}$. At each point of $\mathbb{R}^{n+1}$, with $t \geq 0$, define the standard null frame where

$$\begin{aligned} L = E_+ &= \partial_t + \partial_r \\ \underline{L} = E_- &= \partial_t - \partial_r. \end{aligned}$$

Observe that the conformal Killing vectorfield $K_0 = (t^2 + r^2)\partial_t + 2rt\partial_r$ can be expressed in the form,

$$K_0 = \frac{1}{2}\left[(t+r)^2 E_+ + (t-r)^2 E_-.\right]$$

Thus,

$$\begin{aligned} \mathcal{E}_c(t) &= \int_{\mathbb{R}^n} \frac{1}{4}(t+r)^2 \mathbf{T}_{++} + \frac{1}{4}(t-r)^2 \mathbf{T}_{--} + \underbrace{\left((t+r)^2 + (t-r)^2\right)}_{2(t^2+r^2)} \mathbf{T}_{+-} dx. \\ &= \int_{\mathbb{R}^n} \frac{1}{4}(t+r)^2 \mathbf{T}_{++} + \frac{1}{2}(t^2+r^2)\mathbf{T}_{+-} + \frac{1}{4}(t-r)^2 \mathbf{T}_{--} dx \qquad (317) \\ \mathcal{E}_c(0) &= \int_{\mathbb{R}^n} \mathbf{T}(\partial_t, K_0)(0,x) dx = \int_{\mathbb{R}^n} |x|^2 \mathbf{T}(\partial_t, \partial_t) dx \end{aligned}$$

According to (316) we have $\mathcal{E}_c(t) = \mathcal{E}_c(0)$. Assuming that $\mathcal{E}_c(0) = \int_{\mathbb{R}^n} |x|^2 \mathbf{T}(\partial_t, \partial_t) dx$ is finite we conclude that,

$$\begin{aligned} \int_{\mathbb{R}^n} \mathbf{T}_{++}(t,\cdot) dx &\lesssim \frac{\mathcal{E}_c(0)}{t^2} \\ \int_{\mathbb{R}^n} \mathbf{T}_{+-}(t,\cdot) dx &\lesssim \frac{\mathcal{E}_c(0)}{t^2}. \end{aligned}$$

The remaining term in (317) contains the factor $(t-r)^2$ which is constant along outgoing null directions $r = t + c$. Hence for any $0 < \epsilon < 1$

$$\begin{aligned} \int_{|x|>(1+\epsilon)t} \mathbf{T}_{--} &= O(t^{-2}) \\ \int_{|x|<(1-\epsilon)t} \mathbf{T}_{--} &= O(t^{-2}). \end{aligned}$$

We conclude that most of the energy of a conformal field is carried by the $\mathbf{T}_{--}$ component and propagates near the light cone.

3.6. The case of $\Box\phi = 0$. The wave equation $\Box\phi = 0$ is conformal invariant in dimension $n = 1$. However we can still derive useful conservation laws corresponding to conformal Killing vectorfields in any dimension.

Lemma 3.7. *Let $\mathbf{T}_{\alpha\beta} = \mathbf{T}_{\alpha\beta}[\phi]$ the corresponding energy momentum tensor to a solution of $\Box\phi = 0$. Let X be a conformal Killing vectorfield, i.e. $\pi = {}^{(X)}\pi = \mathcal{L}_X m = \Omega m$, and $tr\pi = m^{\alpha\beta}\pi_{\alpha\beta}$. It is easy to check that $\Box\Omega = 0$; in fact, in the particular case of $X = K_0$, $\Omega = 4(n+1)t$. Let*

$$\bar{P}_\alpha = \mathbf{T}_{\alpha\beta}X^\beta + \frac{n-1}{4(n+1)} tr\pi\phi\partial_\alpha\phi - \frac{n-1}{8(n+1)}\partial_\alpha(tr\pi)\phi^2.$$

We have,

$$\partial^\alpha \bar{P}_\alpha = 0.$$

Now consider the null pair $L = \partial_t + \partial_r, \underline{L} = \partial_t - \partial_r$ as in the previous section. We easily check,

$$Q(L, L) = L(\phi)^2 \tag{318}$$

$$Q(L, \underline{L}) = |\nabla\!\!\!/\phi|^2 \tag{319}$$

$$Q(\underline{L}, \underline{L}) = L(\phi)^2 \tag{320}$$

where $|\nabla\!\!\!/\phi| = \sum_A |e_A(\phi)|^2$ with $(e_A)_{A=1,\dots,n-1}$ an orthonormal frame spanning the orthogonal complement of $L, \underline{L}$.

Part 3

Analysis of the Wave Equation in Minkowski Space

CHAPTER 7

Decay estimates

Consider the standard wave equation in Minkowski space $\mathbb{R}^{n+1}$

$$\Box\phi = 0. \tag{321}$$

The canonical, inertial, coordinates in $\mathbb{R}^{n+1}$ are denoted by x^μ, $\mu = 0, 1, \dots, n$ relative to which the Minkowski metric takes the diagonal form $\mathbf{m}_{\mu\nu} = \mathrm{diag}(-1, 1, \dots, 1)$. We have $x^0 = t$ and $x = (x^1, \dots, x^n)$ denote the spatial coordinates. We make use of the standard summation convention over repeted indices and those concerning raising and lowering the indices of vectors and tensors. In particular, if $x_\mu = m_{\mu\nu}x^\nu$, we have $x_0 = -t$ and $x_i = x^i$, $i = 1, \dots, n$. We denote by Σ_{t_0} the spacelike hyperplanes $t = t_0$. The wave operator is defined by $\Box = \mathbf{m}^{\alpha\beta}\partial_{\alpha\beta} = -\partial_t^2 + \sum_i \partial_i^2$. We study the initial value problem,

$$\phi(0, x) = f(x), \quad \partial_t\phi(0, x) = g(x) \tag{322}$$

For convenience we denote $\phi[0] = (f, D^{-1}g)$ with D^{-1} the pseudodifferential operator with symbol $|\xi|^{-1}$. Let,

$$E[\phi](t) = \int_{\Sigma_t} (|\partial_t\phi|^2 + \sum_i |\partial_i\phi|^2)dx \tag{323}$$

be the total energy of ϕ at time t. The conservation law for the energy is,

$$E[\phi](t) = E[\phi](0) \tag{324}$$

As a consequence we have the energy inequalities, for all $s \geq 0$,

$$\|\partial\phi(t)\|_{H^s(\mathbb{R}^n)} \leq \| \partial\phi(0) \|_{H^s(\mathbb{R}^n)}.$$

The energy identity can be proved both by the geometric techniques discussed in the previous sections, involving only integration by parts, or by the Fourier method, using Plancherel formula together with the Fourier representation formula,

$$\phi(t, x) = (2\pi)^{-n} \int e^{ix\cdot\xi} \big(\cos t|\xi| f\hat{\ }(\xi) + \frac{\sin t|\xi|}{|\xi|} g\hat{\ }(\xi) \big) d\xi \tag{325}$$

Remark 0.8. The standard Sobolev embedding $H^s(\mathbb{R}^n) \subset L^\infty(\mathbb{R}^n)$, for $s > \frac{n}{2}$ allows us to get L^∞ bounds of solutions to (321) without using the explicit representation. This procedure generalizes to nonlinear equations and plays an important role in the proof of the local existence theorem.

Proposition 0.9 (Dispersive inequality)**.** *The solutions to* (321)*,* (322) *verify,*

$$\|\phi(t)\|_{L^\infty} \leq c|t|^{-\frac{n-1}{2}} \|\phi[0]\|_{B_{1,1}^{\frac{n+1}{2}}} \tag{326}$$

with $B^{\frac{n+1}{2}}_{1,1}$ *the Besov space slightly larger than* $W^{\frac{n+1}{2}}, 1$*. More precisely,*

$$\|f\|_{B^{\frac{n+1}{2}}_{1,1}} \approx \sum_{k\in\mathbb{Z}} 2^{k\frac{n+1}{2}} \|P_k f\|_{L^1(\mathbb{R}^n)}$$

Exercise. Show that the inequality (326) follows from its frequency localized version. In other words show that it suffices to prove the following inequality,

$$\|\phi(t)\|_{L^\infty} \le c|t|^{-\frac{n-1}{2}} \|\phi[0]\,\|_{L^1} \tag{327}$$

for initial data f, g whose Fourier transforms are localized to $\frac{1}{2} \le |\xi| \le 2$.

Proof The standard proof of (326) is based on the method of stationary phase applied to the representation (325). In odd dimensions one can prove a related form of the dispersive estimate using the spherical means representation of solutions. This is particularly easy to do for $n = 3$. We shall later discuss a derivation of (326) which avoids any representation formulas. ■

Remark 0.10. The dispersive inequality provides two types of information. The first concerns the precise decay rate of $\|\phi(t)\|_{L^\infty}$ as $t \to \infty$ while the second provides information about the regularity properties of $\|\phi(t)\|_{L^\infty}$ for $t > 0$. As far as improved regularity is concerned the estimate (326) gains, for $t > 0$, $\frac{n-1}{2}$ derivatives when compared to the Sobolev embedding $L^\infty(\mathbb{R}^n) \subset W^{1,n}(\mathbb{R}^n)$.

In many applications, especially to nonlinear equations, (326) is not very useful. A more effective procedure to derive the asymptotic properties of solutions of the wave equation is based on generalized energy estimates, obtained by the commuting vectorfields method, together with global Sobolev inequalities. In what follows we review the commuting vectorfields method for deriving the above decay rate estimate. The idea is to use the energy identity (324) together with the vectorfields which commute with the wave operator $\Box$ and and a global version of the classical Sobolev inequalities We refer the reader to [**?**] and [**?**] for details.

The Minkowski space-time $\mathbb{R}^{n+1}$ is equipped, see appendix 4.2, with a family of Killing and conformal Killing vector fields, the translations $\mathbf{T}_\mu = \partial_\mu$, Lorentz rotations $\mathbf{L}_{\mu\nu} = x_\mu\partial_\nu - x_\nu\partial_\mu$, scaling $\mathbf{S} = t\partial_t + x^i\partial_i$ and the inverted translations $\mathbf{K}_\mu = -2x_\mu\mathbf{S} + < x, x > \partial_\mu$. Recall that x^μ, denote the standard variables $x^0 = t$, $x^1, \dots, x^n$, and $x_\mu = m_{\mu\nu}x^\nu$. The Killing vector fields $\mathbf{T}_\mu$ and $\mathbf{L}_{\mu\nu}$ commute with $\Box$ while $\mathbf{S}$ preserves the space of solutions in the sense that $\Box\phi = 0$ implies $\Box\mathbf{S}\phi = 0$ as $[\Box, S] = 2\Box$. One can split the operators $\mathbf{L}_{\mu\nu}$ into the angular rotation operators ${}^{(ij)}\mathbf{O} = x_i\partial_j - x_j\partial_i$ and the boosts ${}^{(i)}\mathbf{L} = x_i\partial_t + t\partial_i$, for $i, j, k = 1, \dots, n$. Recall the energy expression in (323). Based on the commutation properties described above we define the following "generalized energies "

$$E_k[\phi] = \sum_{X_{i_1},..,X_{i_j}} E[X_{i_1}X_{i_2}...X_{i_j}\phi] \tag{328}$$

with the sum taken over $0 \le j \le k$ and over all Killing vector fields $\mathbf{T}, \mathbf{L}_{\mu\nu}$ as well as the scaling vector field $\mathbf{S}$. The crucial point of the commuting vectorfield method is that the quantities E_k, $k \ge 1$ are conserved by solutions to (321). Therefore, if,

$$\sum_{0 \le k \le s} \int (1+|x|)^{2k} \Big(|\nabla^{k+1} f(x)|^2 + |\nabla^k g(x)|^2 \Big) dx \le C_s < \infty \tag{329}$$

then for all t, $E_s[\phi](t) \le C_s$. The desired decay estimates of solutions to (321) can now be derived from the following global version of the Sobolev inequalities (see [**?**], [**?**]):

Theorem 0.11 (Global Sobolev)**.** *Let ϕ be an arbitrary function in R^{n+1} such that $E_s[\phi]$ is finite for some integer $s > \frac{n}{2}$. Then,*

$$|\partial\phi(t,x)| \lesssim (1+t+|x|)^{-\frac{n-1}{2}} (1+|t-|x||)^{-\frac{1}{2}} \sup_{0 \le t' \le t} E_s[\phi](t') \tag{330}$$

for all $t > 0$. Therefore if the data f, g in (321) *satisfy 329, with $s > \frac{n}{2}$, then for all $t \ge 0$,*

$$|\partial\phi(t,x)| \lesssim \frac{1}{(1+t+|x|)^{\frac{n-1}{2}} (1+|t-|x||)^{\frac{1}{2}}} \tag{331}$$

Remark 0.12. Clearly this estimate, whose proof is purely geometric[1], implies the decay properties given by the dispersive inequality (326). In fact it provides more information outside the wave zone $|x| \sim t$ which fit very well with the expected propagation properties of the linear equation $\square\phi = 0$. On the other hand, as (330) is really a global version of the Sobolev inequality, it seems that the estimates of the Proposition 0.11 have no bearing on the improved regularity features of (326). This is however not quite true as we shall see, later.

Proof We only sketch the main ideas of the proof below. Consider the canonical null pair $L_\pm = \partial_t \pm \partial_r$,, an associated null frame $e_1, \ldots e_{n-1}, e_n = L_-, e_{n+1} = L_+$ as well as the angular vectorfields, $A_i = \partial_i - \frac{x_i}{r}\partial_r$. Clearly,

$$\sum_i |A_i\phi| \lesssim |\nabla\!\!\!\!/\,\phi| \lesssim \sum_i |A_i\phi|.$$

where $|\nabla\!\!\!\!/\,\phi|^2 = \sum_{i=1}^{n-1} |e_i(\phi)|^2$. Also,

$$|\partial_r\phi| + \sum_i |A_i\phi| \lesssim |\nabla\phi| \lesssim |\partial_r\phi| + \sum_i |A_i\phi|$$

We can also easily check the following simple algebraic identities,

$$\begin{aligned}
\frac{1}{2}\big((t+r)L_+ + (t-r)L_-\big) &= \mathbf{S} \\
\frac{1}{2}\big((t+r)L_+ - (t-r)L_-\big) &= \sum_i \frac{x_i}{|x|}\mathbf{L}_i \\
tA_i &= L_i - \frac{x_i}{|x|}\sum_j \frac{x_j}{|x|} L_j \\
t\mathbf{O}_{ij} &= x_i\mathbf{L}_j - x_j\mathbf{L}_i
\end{aligned}$$

[1] In particular it does not require any explicit representation of solutions

From the first two identities we easily derive,

$$
\begin{aligned}
|L_+\phi(t,x)| &\lesssim \frac{1}{t}|\Gamma\phi(t,x)| \\
|L_-\phi(t,x)| &\lesssim \frac{1}{\big|t-|x|\big|}|\Gamma\phi(t,x)|
\end{aligned}
\tag{332}
$$

with $|\Gamma\phi| = |\mathbf{S}\phi| + |\mathbf{L}\phi|$.

$$
|\nabla\!\!\!\!/\,\phi(t,x)| \lesssim \frac{1}{t}|\Gamma\phi(t,x)|. \tag{333}
$$

Clearly, we also have,

$$
|\partial\phi(t,x)| \lesssim \frac{1}{\big|t-|x|\big|}|\Gamma\phi(t,x)|
$$

or, more generally,

$$
|\partial^N\phi(t,x)| \lesssim \frac{1}{\big|t-|x|\big|^N}|\Gamma^N\phi(t,x)| \tag{334}
$$

where $|\Gamma^N\phi| = \sum|\Gamma_1\ldots\Gamma_N\phi|$ with $\Gamma_1,\ldots,\Gamma_N$ any of the vectorfields $S, L_1, \ldots L_n$.

Combining the above inequalities with the definition of our norms we derive

$$
\begin{aligned}
t\|E_+\phi(t)\|_{L^2} &\lesssim \|\Gamma\phi(t)\|_{L^2(\mathbb{R}^n)} \\
t\|\nabla\!\!\!\!/\,\phi(t)\|_{L^2} &\lesssim \|\Gamma\phi(t)\|_{L^2(\mathbb{R}^n)} \\
\|uE_-\phi(t)\|_{L^2} &\lesssim \|\Gamma\phi(t)\|_{L^2(\mathbb{R}^n)}
\end{aligned}
$$

where $u = |t-|x||$ It remains to derive sup-norm estimates from the L^2 estimates above.

Proposition 0.13. *Let $\Box\phi = 0$ with initial data verifying the assumptions above. Then, for all $t \geq 0$, $s > \frac{n}{2}$,*

$$
\|\phi(t)\|_{L^\infty} \lesssim \Big(\frac{1}{t}\Big)^{\frac{n-1}{2}}\|\Gamma^s\phi(t)\|_{L^2(\mathbb{R}^n)} \tag{335}
$$

$$
\|(1+|u|)^k\partial^k\phi(t)\|_{L^\infty} \lesssim \Big(\frac{1}{t}\Big)^{\frac{n-1}{2}}\|\Gamma^s\phi(t)\|_{L^2(\mathbb{R}^n)} \tag{336}
$$

Also,

$$
\begin{aligned}
\|E_+\phi(t)\|_{L^\infty} &\lesssim \Big(\frac{1}{t}\Big)^{\frac{n+1}{2}}\|\Gamma^{s+1}\phi(t)\|_{L^2(\mathbb{R}^n)} \\
\|\nabla\!\!\!\!/\,\phi(t)\|_{L^\infty} &\lesssim \Big(\frac{1}{t}\Big)^{\frac{n+1}{2}}\|\Gamma^{s+1}\phi(t)\|_{L^2(\mathbb{R}^n)} \\
\|(1+|u|)E_-\phi(t)\|_{L^\infty} &\lesssim \Big(\frac{1}{t}\Big)^{\frac{n-1}{2}}\|\Gamma^{s+1}\phi(t)\|_{L^2(\mathbb{R}^n)}
\end{aligned}
$$

The proof is based on the following Lemma

Lemma 0.14. *Let $u(x)$ be a smooth, compactly supported function on $\mathbb{R}^n$, $n \geq 2$. We have,*

$$|u(x)| \leq C \frac{1}{|x|^{n-1}} \Big(\|\partial_r u\|_{L^1} + \|(r \nabla\!\!\!/)^{n-1} \partial_r u\|_{L^1} \Big) \tag{337}$$

■

Proposition 0.15 (see[**?**]). *The commuting vectorfields method implies the dispersive inequality* (326)*.*

Proof Without loss of generality we may assume that $\partial_t \phi = g = 0$ and that the Fourier transform of $f = \phi(0)$ is supported in the shell $\frac{\lambda}{2} \leq |\xi| \leq 2\lambda$ for some $\lambda \in 2^{\mathbf{N}}$. By a simple scaling argument we may in fact assume $\lambda = 1$. For such initial conditions, with Fourier supports restricted to $1/2 \leq |\xi| \leq 2$, it suffices to prove,

$$\|\phi(t)\|_{L^\infty} \lesssim (1 + |t|)^{-\frac{n-1}{2}} \|f\|_{L^1(\mathbb{R}^n)}$$

Since $\hat{\phi}$, the Fourier transform of ϕ relative to the space variables x, is also supported in the same shell it suffices to prove the estimates for $\nabla \phi$ instead of ϕ.

Next we cover $\mathbb{R}^n$ by an union of discs D_I centered at points $I \in \mathbf{Z}^n$ with integer coordinates such that each D_I intersects at most a finite number c_n of discs D_J with c_n depending only on the dimension n. Consider a smooth partition of unity $(\chi_I)_{I \in \mathbf{Z}^n}$ with $\operatorname{supp} \chi_I \subset D_I$ and each χ_I positive. Clearly we can arrange to have, for all k,

$$\sum_{I \in \mathbf{Z}^n} |\nabla^k \chi_I(x)| \leq C_{k,n} \tag{338}$$

uniformly in $x \in \mathbb{R}^n$, with a constant $C_{n,k}$ depending only on n and k. Now set, $f_I = \chi_I \cdot f$, and ϕ_I the corresponding solution to (321) with data $\phi_I(0) = f_I, \partial_t \phi_I(0) = 0$. Clearly $f = \sum_I f_I,\ \phi = \sum_I \phi_I$. It suffices to prove that for all I,

$$\|\nabla^k \phi_I(t)\|_{L^\infty} \lesssim (1 + t)^{-\frac{n-1}{2}} \sum_{j=0}^{n+k+1} \|D^j f_I\|_{L^1} \tag{339}$$

Indeed if (339) holds true we easily infer that,

$$\|\nabla^k \phi(t)\|_{L^\infty} \quad \leq \quad \sum_I \|\nabla^k \phi_I(t)\|_{L^\infty} \lesssim (1 + t)^{-\frac{n-1}{2}} \sum_{j=0}^{n+k+1} \sum_I \|D^j f_I\|_{L^1}$$

In view of (338) we have,

$$\begin{aligned}\sum_I \|D^j f_I\|_{L^1} &= \sum_I \|D^j(\chi_I f)\|_{L^1} \lesssim \sum_I \sum_{0\le i\le j} \int_{\mathbb{R}^n} |D^i\chi_I(x)||D^{j-i}f(x)|dx \\ &= \sum_{0\le i\le j} \int_{\mathbb{R}^n} \Big(\sum_I |D^i\chi_I(x)| \Big) \big|D^{j-i}f(x)|dx \\ &\lesssim \sum_{0\le i\le j} c_{i,n} \|D^{j-i}f\|_{L^1(\mathbb{R}^n)} \lesssim \|f\|_{L^1(\mathbb{R}^n)}\end{aligned}$$

Hence,

$$\|\nabla^k\phi(t)\|_{L^\infty} \quad \lesssim \quad \le (1+t)^{-\frac{n-1}{2}} \|f\|_{L^1(\mathbb{R}^n)}$$

as desired.

It therefore remains to check (339). Without loss of generality, by performing a space translation, we may assume that $I = 0$. Applying the proposition 0.11 to $\psi = \nabla\phi_0$ we derive, for s_* the first integer strictly larger than $\frac{n}{2}$,

$$\begin{aligned}\|\psi(t)\|_{L^\infty} &\le c(1+t)^{-\frac{n-1}{2}} E_{s_*}[\phi_0](t) \\ &\le c(1+t)^{-\frac{n-1}{2}} E_{s_*}[\phi_0](0).\end{aligned}$$

Since the support of ϕ_0 is included in in the ball of radius 1 centered at the origin we have,

$$E_{s_*}[\phi_0](0) \le C_n \sum_{j=0}^{s_*+1} \|D^j f_0\|_{L^2}.$$

Finally, according to the standard Sobolev inequality in $\mathbb{R}^n$, $\|f\|_{L^2} \le c\|\nabla^{\frac{n}{2}} f\|_{L^1}$, we conclude with,

$$\|\psi(t)\|_{L^\infty} \le c(1+t)^{-\frac{n-1}{2}} \sum_{j=0}^{n+2} \|D^j f_0\|_{L^1}$$

as desired.

∎

CHAPTER 8

Strichartz Inequalities

Strichartz inequalities are an important tool in the study of linear and nonlinear wave equations. They are intimately tied to restriction theorems. In this chapter we shall only consider the case of the standard linear wave equation. Similar inequalities hold true however for linear dispersive equations such as the Schrödinger, linear KdV etc.

0.15.1. *Homogeneous wave equation.* Consider solutions $u = u(t,x), t \in \mathbb{R}, x \in \mathbb{R}^n$ to the equation

$$\Box u = F, \tag{340}$$

$$u(0,x) = f(x), \quad \partial_t u(0,x) = g(x), \tag{341}$$

with $\Box$ the wave operator $\Box = -\partial_t^2 u + \Delta$. Clearly, a solution to eqrefeq:genwave can be written as a superposition between a solution to the homogeneous wave equation,

$$\Box u = 0, \tag{342}$$

verifying the initial condition (341) at time $t = 0$, and a solution to the purely inhomogeneous wave equation

$$\Box u = F, \tag{343}$$

with zero initial data

$$u(0,x) = 0, \quad \partial_t u(x,0) = 0.$$

We denote by $W(t)h$ the fundamental solution of the homogeneous problem (342), i.e. $u(t,x) = \big(W(t)h\big)(x)$ is the unique solution of (342) which verifies the initial conditions

$$u(0,x) = 0, \quad \partial_t u(0,x) = h(x)$$

By Duhamel's principle any solution of the inhomogeneous equation can itself be written as a superposition of solutions to the homogeneous equation according to the formula,

$$u(t) = \int_0^t W(t-t')F(t')\mathrm{d}t', \tag{344}$$

Before stating the main result of this section we make the following definition.

Definition 0.16. We say that the pair of real numbers (q, r) is an admissible wave pair if they satisfy the conditions

$$\begin{aligned} q &\geq 2, \\ \frac{2}{q} &\leq (n-1)\left(\frac{1}{2}-\frac{1}{r}\right), \\ (q, r, n) &\neq (2, \infty, 3). \end{aligned}$$

We are now ready to state the following.

Theorem 0.17. *Suppose that $n \geq 2$ and (q, r) is a wave admissible pair [1] with $r < \infty$.*

(1) *Assume the dimensional condition, $\frac{1}{q}+\frac{n}{r}=\frac{n}{2}-\gamma$. Then, if u verifies the homogeneous equation (342) with initial conditions (341),*

$$\|u\|_{L_t^q L_x^r} + \|u\|_{L_t^\infty \dot{H}_x^\gamma} + \|\partial_t u\|_{L_t^\infty \dot{H}_x^{\gamma-1}} \lesssim \|f\|_{\dot{H}^\gamma} + \|g\|_{\dot{H}^{\gamma-1}} \tag{345}$$

(2) *Assume the dimensional condition,[2] $\frac{1}{q}+\frac{n}{r}=\frac{n}{2}-\gamma=\frac{1}{q'}+\frac{n}{r'}-2$, with q' dual to q and r' dual to r. Then, if u verifies the purely inhomogeneous problem (0.15.1) with zero initial conditions, then on a finite time interval $[0, T]$:*

$$\|u\|_{L^q([0,T];L^r)} + \|u\|_{C([0,T];\dot{H}^\gamma)} + \|\partial_t u\|_{C([0,T];\dot{H}^{\gamma-1})} \lesssim \|F\|_{L^{q'}([0,T];L^{r'})} \tag{346}$$

(3) *We also have the following more general version of (346) for admissible pairs (q_1, r_1), (q_2, r_2) with $r_1, r_2 < \infty$ verifying the dimensional condition,*

$$\frac{1}{q_1}+\frac{n}{r_1}=\frac{n}{2}-\gamma=\frac{1}{q_2'}+\frac{n}{r_2'}-2$$

Then,

$$\|u\|_{L^{q_1}([0,T];L^{r_1})} + \|u\|_{C([0,T];\dot{H}^\gamma)} + \|\partial_t u\|_{C([0,T];\dot{H}^{\gamma-1})} \lesssim \|F\|_{L^{q_2'}([0,T];L^{r_2'})} \tag{347}$$

Remark 0.18. For $n \geq 4$, the region of admissable exponents corresponds to a quadrilateral $OEPQ$ in the plane $(1/q, 1/r)$ with vertices $O = (1/\infty, 1/\infty)$, $E = (1/\infty, 1/2)$, $P = (1/2, \frac{n-3}{2(n-1)})$ and $Q = (1/2, 1/\infty)$. When $n = 3$ the point P coincides with Q and the region reduces to the triangle OEQ. When $n = 2$ we have a smaller triangle OEQ_2 where $Q_2 = (1/4, 1/\infty)$.

For $n = 3$, the boundary of the triangular region is allowed except for the endpoint P. For $n \geq 4$, the boundary of the quadrilateral region is entirely allowed, as we will note below.

The interesting cases are the ones on the segment EP and the ones on PQ close to P, since all the others can be deduced from these using Sobolev embeddings. The point E corresponds to the energy estimates. There are counterexamples that

[1] The case when $r = \infty$ can also be included provided that we modify the spaces on the left of the estimates below to appropriate Besov spaces.

[2] Thus, in fact, $\gamma = 1/2$.

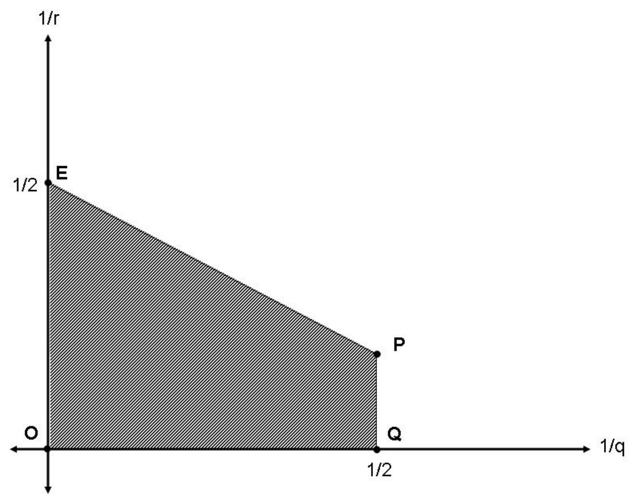

FIGURE 1. Admissable exponents for $n \geq 4$

exclude the point P when $n = 3$, while the inclusion of P in higher dimensions were recently obtained by Keel and Tao [**K-T**].

The standard Strichartz estimate[3] corresponds to the point $S = (\frac{n-1}{2(n+1)}, \frac{n-1}{2(n+1)})$.

Remark 0.19. We remark that in even though the end-point case $n = 3, q = \infty, r = 2$ is forbidden, the estimates holds in the spherically symmetric case. Indeed let ϕ be a solution of the homogeneous wave equation $\Box\phi = 0$ in $\mathbb{R}^{3+1}$ subject to the initial conditions

$$\phi(0, x) = 0, \quad \partial_t\phi(0, x) = f(x)$$

and assume that f is spherically symmetric i.e. $f(x) = f(|x|)$. Then,

$$\int_0^\infty \|\phi(t, \cdot)\|^2_{L^\infty(\mathbb{R}^3)} dt \leq c\|f\|^2_{L^2}. \tag{348}$$

The proof is an immediate consequence of the Hardy–Littlewood maximal theorem[4] in view of the fact that, for spherically symmetric f,

$$\phi(x, t) = \frac{c}{|x|} \int_{\left||x|-t\right|}^{|x|+t} \lambda f(\lambda)\, d\lambda.$$

Remark 0.20. We give an elementary example below to illustrate how the end point result $n = 3, q = \infty, r = 2$ fails in the general case due to possible concentrations along null rays. We show below that there exists a sequence of functions f_n in $C_0^\infty(\mathbb{R}^3)$, with $\|f_n\|_{L^2} = 1$ such that for the corresponding solutions ϕ_n,

$$\int_0^\infty |\phi_n(t, t, 0, 0)|^2\, dt \geq n. \tag{349}$$

[3] i.e. the one actually proved by Strichartz.

[4] This is obviously so in the region $r \leq t$ while for $r \geq t$ the argument is elementary.

assume by contradiction that in fact, $J := \int_0^\infty \phi(t,t,0,0)\varphi(t)\,dt < C$ for all $f \in C_0^\infty(\mathbb{R}^3)$ with $\|f\|_{L^2} = 1$ and some $\varphi \in \mathcal{S}(\mathbb{R}), \varphi \not\equiv 0$. In view of the formula (see section on the fundamental solution of $\Box$ in $\mathbb{R}^{3+1}$),

$$\phi(t,x) = (4\pi)^{-1} t \int_{|\xi|=1} f(x+t\xi)\,d\xi$$

we find that,

$$J = (4\pi)^{-1} \int_{\mathbb{R}^3} |y|^{-1} f_1(y_1 + |y|, y_2, y_3)\varphi(|y|)\,dy$$

or, changing the variables $z = y + (|y|, 0, 0)$

$$J = (4\pi)^{-1} \int\limits_{z_1>0} \frac{1}{z_1} f(z)\varphi\Big(\frac{|z|^2}{2z_1}\Big) dz < c.$$

Since f is an arbitrary $C_0^\infty(\mathbb{R}^3)$ function, $\|f\|_{L^2} = 1$, we must have that,

$$z \to \frac{1}{z_1}\varphi\Big(\frac{|z|^2}{2z_1}\Big)$$

is in $L^2(\mathbb{R}^3_+)$ which is false whenever $\varphi \not\equiv 0$. In fact,

$$\int_{\mathbb{R}^3_+} \frac{1}{z_1^2}\varphi^2\Big(\frac{|z|^2}{2z_1}\Big) dz \quad = \quad \int_{\mathbb{R}^3} \frac{1}{(y_1+|y|)|y|}\varphi^2(|y|)dy = 2\pi \int_0^\infty \varphi^2(\lambda) \int_0^\pi \frac{\sin\theta}{1+\cos\theta}\,d\theta$$

diverges logarithmically if $\varphi \not\equiv 0$.

0.21. Fourier representation of solutions. We can solve the homogeneous problem (342) by the Fourier method. To recall, If we apply the Fourier transform with respect to the space variables, the initial value problem (342), (341) becomes a Cauchy problem for an ordinary differential equation:

$$\partial_t^2 \widehat{u} + |\xi|^2 \widehat{u} = 0, \quad \widehat{u}(0,\xi) = \hat{f}(\xi), \quad \partial_t \widehat{u}(0,\xi) = \hat{g}(\xi),$$

which can be solved explicitly:

$$\widehat{u}(t,\xi) = \cos(t|\xi|)\,\hat{f}(\xi) + \sin(t|\xi|)\,\frac{\hat{g}(\xi)}{|\xi|} \tag{350}$$

Thus the fundamental solution $W(t)h$, defined above, takes the form,

$$W(t)h(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi} \frac{\sin(t|\xi|)}{|\xi|} \hat{h}(\xi) d\xi. \tag{351}$$

By Duhamel principle, see (344), the general solution of the inhomogeneou equation $\Box u = F$ can be expressed in the form,

$$u(t) = \partial_t W(t) f + W(t) g + \int_0^t W(t-s)F(s)ds. \tag{352}$$

let $D = (-\Delta)^{1/2}$ be the operator whose symbol in Fourier space is given by $|\xi|$. Observe that,

$$(DW(t))f(x) \quad = \quad (W(t)Df)(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi} \sin t|\xi|)\hat{f}(\xi)d\xi.$$

Since $\sin t|\xi|$ and $\cos t|\xi|$ are bounded the operators $\partial_t W(t)$ and $DW(t)$ map $H^s(\mathbb{R}^n)$ in itself. In particular, solutions u of (342), (341) preserves the (Sobolev) regularity of the initial data f and g. More precisely, If $f, D^{-1}g \in H^s$ for some $s \in \mathbb{R}$, then $u(t), D^{-1}\partial_t u(t) \in H^s$ uniformly for $t \in \mathbb{R}$. We can also write,

$$\|u\|_{L_t^\infty \dot{H}_x^\gamma} + \|\partial_t u\|_{L_t^\infty \dot{H}_x^{\gamma-1}} \lesssim \|f\|_{\dot{H}^\gamma} + \|g\|_{\dot{H}^{\gamma-1}}$$

which provides the easy part of estimate[5] (345). Therefore to prove (345) it suffices to prove,

$$\|u\|_{L_t^q L_x^r} \lesssim \|f\|_{\dot{H}^\gamma} + \|g\|_{\dot{H}^{\gamma-1}} \tag{353}$$

for and wave admissible pair (q, r).

We also remark that,

$$\partial_t W(t)h(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi}\cos(t|\xi|)\hat{h}(\xi)d\xi$$

and,

$$D^{-1}W(t)h(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi}\frac{\cos(t|\xi|)}{|\xi|}\hat{h}(\xi)d\xi$$

We can rewrite (350) as

$$\widehat{u}(t,\xi) = e^{it|\xi|}\hat{f}^+(\xi) + e^{-it|\xi|}\hat{f}^-(\xi),$$

where $f^\pm = \frac{1}{2}\left(f \pm D^{-1}g\right)$. It follows that $u = u^+ + u^-$ where

$$u^\pm = \int e^{i(x\cdot\xi \pm t|\xi|)}\hat{f}_\pm(\xi)d\xi$$

Observe that to prove (353) it suffices to prove,

$$\|u^+\|_{L_t^q L_x^r} \lesssim \|f^+\|_{\dot{H}^\gamma} \tag{354}$$

and a similar estimate for f^-.

0.22. Energy estimates. We will derive a simple L^2 estimate for general solutions of $\Box u = F$ by integration by parts. It all follows from the simple algebraic identity:

$$-\frac{1}{2}\partial_t(|\partial_t u|^2 + |\nabla u|^2) + \partial_i(\partial_t u \partial_i u) = \partial_t u \cdot F \tag{355}$$

where $|\nabla u|^2 = \sum_{i=1}^n (\partial_i u)^2$ and $\partial_i = \partial_{x^i}$. Integrating with respect to x, and assuming that u and its derivatives vanish[6] at infinity we derive,

$$\partial_t \int_{\mathbb{R}^n} \left((\partial_t u)^2 + |\nabla u|^2\right)dx = 2\int_{\mathbb{R}^n} \partial_t u \cdot F dx$$

Thus integrating in t,

$$\|\partial_t u(t)\|_{L^2}^2 + \|\nabla u(t)\|_{L^2}^2 \le \|\partial_t u(0)\|_{L^2}^2 + \|\nabla u(0)\|_{L^2}^2 + 2\int_0^t \int_{\mathbb{R}^n} \partial_t u \cdot F dx ds$$

[5]Another derivation, based on energy identities, is given in the next subsection.

[6]This can easily be justified by the finite propagation speed property of solutions to the wave equation

which we rewrite, with $|\partial u|^2 = |\partial_t u|^2 + |\nabla u|^2$,

$$\|\partial u(t)\|_{L^2}^2 = \|\partial u(0)\|_{L^2}^2 + 2\int_0^t \int_{\mathbb{R}^n} \partial_t u \cdot F dx ds. \tag{356}$$

In particular, applying Hölder,

$$\|\partial u(t)\|_{L^2}^2 \le \|\partial u(0)\|_{L^2}^2 + 2\int_0^t \|\partial_t u(s)\|_{L^2}\|F(s)\|_{L^2} ds$$

from which we derive the inhomogeneous energy estimate,

$$\sup_{t\in[0,T]} \|\partial u(t)\|_{L^2} \lesssim \|\partial u(0)\|_{L^2} + \int_0^T \|F(s)\|_{L^2} ds \tag{357}$$

Now let D^s be the operator $D^s = (-\Delta)^{s/2}$ whose symbol in Fourier space is given by $|\xi|^s$. Since D^s commutes with $\Box$ we easily derive,

$$\|\partial D^s u(t)\|_{L^2}^2 = \|\partial D^s u(0)\|_{L^2}^2 + 2\int_{\mathbb{R}^n} \partial_t D^s u \cdot D^s F dx$$

We can write, using Plancherel with respect to the x variables,

$$\int_{\mathbb{R}^n} \partial_t D^s u \cdot D^s F dx = \int_{\mathbb{R}^n} \partial_t D^{2s} u \cdot F dx,$$

Therefore, by Hölder, in the slab $\mathcal{D}_T = [0,T]\times\mathbb{R}^n$,

$$\sup_{t\in[0,T]} \|\partial D^s u(t)\|_{L^2}^2 \le \|\partial D^s u(0)\|_{L^2}^2 + 2\|D^{2s}\partial_t u\|_{L_t^q L_x^r(\mathcal{D}_T)}\|F\|_{L_t^{q'} L_x^{r'}(\mathcal{D}_T)}$$

Choosing $s = -1$ we infer that,

$$\sup_{t\in[0,T]} \|\partial D^{-1/2} u(t)\|_{L^2}^2 \le \|\partial D^{-1/2} u(0)\|_{L^2}^2 + 2\|D^{-1}\partial_t u\|_{L_t^q L_x^r(\mathcal{D}_T)}\|F\|_{L_t^{q'} L_x^{r'}(\mathcal{D}_T)}$$

We apply this energy estimate to solution of the inhomogeneous problem (0.15.1) with zero initial conditions. We also assume that the dimensional condition $\frac{1}{q} + \frac{n}{r} = \frac{n}{2} - \gamma = \frac{1}{q'} + \frac{n}{r'} - 2$ is verified. That implies $\gamma = \frac{1}{2}$. We thus have,

$$\sup_{t\in[0,T]} \|\partial D^{-1/2} u(t)\|_{L^2}^2 \le 2\|D^{-1}\partial_t u\|_{L_t^q L_x^r(\mathcal{D}_T)}\|F\|_{L_t^{q'} L_x^{r'}(\mathcal{D}_T)}$$

Assume for a moment that we can prove the estimate,

$$\|D^{-1}\partial_t u\|_{L_t^q L_x^r(\mathcal{D}_T)} \lesssim \|F\|_{L_t^{q'} L_x^{r'}(\mathcal{D}_T)} \tag{358}$$

Then,

$$\sup_{t\in[0,T]} \|\partial D^{-1/2} u(t)\|_{L^2} \lesssim \|F\|_{L_t^{q'} L_x^{r'}(\mathcal{D}_T)}$$

which is equivalent to,

$$\sup_{t\in[0,T]} \left(\|u(t)\|_{\dot H^{\gamma})} + \|\partial_t u\|_{\dot H^{\gamma-1})}\right) \quad\lesssim\quad \|F\|_{L^{q'}([0,T];L^{r'})}$$

thus proving half of estimate (346). Therefore the inhomogeneous estimate (346) reduces to proving,

$$\|u\|_{L^q([0,T];L^r)} + \|D^{-1}\partial_t u\|_{L^q([0,T];L^r)} \quad\lesssim\quad \|F\|_{L^{q'}([0,T];L^{r'})} \tag{359}$$

0.23. Homogenous Case. In this section we prove estimate (354) and thus complete the proof for the homogeneous Strichartz estimate of theorem 0.17. Using the space-time Fourier transform, i.e. Fourier transform with respect to both t and x,

$$\widetilde{u}_+(\tau,\xi) = \delta(\tau - |\xi|)\hat{f}_+(\xi), \quad \widetilde{u}_-(\tau,\xi) = -\delta(\tau + |\xi|)\hat{f}_-(\xi), \tag{360}$$

These are the components of $\widetilde{u}$ living on the forward null cone $C_+ = \{\tau = |\xi|\}$ and on the backward null cone $C_- = \{\tau = -|\xi|\}$, respectively. Thus we can interpret (354) from the point of view of a restriction theorem for the half light cones C_+ or C_-. We next show that it suffices to prove (354) for the case when $\hat{f}_+$ is included in fixed dyadic piece. More precisely, dropping the label + it suffices to show that,

$$\|u_k^+\|_{L_t^q L_x^r} \quad \lesssim \quad 2^{k\gamma}\|f_k^+\|_{L^2} \tag{361}$$

where $u^+ = \sum_{k\in 2^{\mathbb{Z}}} u_k^+$, $u_k^+ = P_k u^+$, $f_k^+ = P_k f^+$ and P_k the standard LP projections with respect to the spatial variables x.

To show that (362) implies (354) is highly nontrivial[7] as we need to rely on corollary 6.15 adapted to the mixed norms $L_t^q L_x^r$ with both q and r larger than 2. Thus,

$$\|u^+\|^2_{L_t^q L_x^r} \quad \lesssim \quad \sum_{k\in\mathbb{Z}} \|u_k^+\|^2_{L_t^q L_x^r} \lesssim \sum_{k\in\mathbb{Z}} 2^{2k\gamma}\|f_k^+\|^2_{L^2} \lesssim \|f^+\|_{\dot{H}^\gamma}$$

Finally we observe, using a simple scaling argument, that (362) follows from,

$$\|u_0^+\|_{L_t^q L_x^r} \quad \lesssim \quad \|f_0^+\|_{L^2} \tag{362}$$

We now define the *truncated cone operator* C to be the operator

$$Cf(t,x) = \int e^{it|\xi|} e^{ix\cdot\xi}\chi(\xi)\hat{f}(\xi)\mathrm{d}\xi, \tag{363}$$

where χ is a cut-off function supported in $1.2 \le |\xi| \le 2$, such as the one used in the definition of the LP projections, see (116). The operator C can be viewed as the adjoint of the restriction of the Fourier transform to a truncated cone,

$$\widehat{C^*F}(\xi) = \overline{\chi(\xi)}\ \widetilde{F}(|\xi|,\xi)$$

Estimate (362) is an immediate consequence of the following theorem.

Theorem 0.24. *Let $(q,r),(q_1,r_1),(q_2,r_2)$ be admissable pairs of exponents. Then we have the estimates*

$$\|Cf\|_{L_t^q L_x^r} \lesssim \|f\|_{L^2}, \tag{364}$$

and also

$$\|CC^*F\|_{L_t^{q_1} L_x^{r_1}} \lesssim \|F\|_{L_t^{q_2'} L_x^{r_2'}}. \tag{365}$$

Composing C with C^* we derive,

$$CC^*F(t,x) \simeq \int e^{i[(t-s)|\xi| + (x-y)\cdot\xi]}|\beta(\xi)|^2 F(s,y)\mathrm{d}s\mathrm{d}y\mathrm{d}\xi,$$

[7]Without using corollary 6.15 we would only derive a weaker estimate with the Besov norm $\dot{B}^\gamma_{2,1}$ replacing $\dot{H}^\gamma$ norm on the right.

which can be rewritten as the convolution

$$CC^*F(t,\cdot) = \int U(t-s)F(s,\cdot)\mathrm{d}s, \tag{366}$$

with the evolution operator

$$U(t)f(x) = \int e^{i(t|\xi|+x\cdot\xi)}|\chi(\xi)|^2\hat{f}(\xi)\mathrm{d}\xi. \tag{367}$$

(Observe that U is essentially the same operator as C !) By the TT^* principle, we know that the estimate (364) is equivalent to the following estimate for CC^*,

$$\|CC^*F\|_{L^q_t L^r_x} \lesssim \|F\|_{L^{q'}_t L^{r'}_x}. \tag{368}$$

which is also equivalent to the polarized form (365). Thus, to prove the theorem it suffices to prove (368). As in the second proof of the restriction theorem presented in the previous section to prove (368) we need to prove the following properties for the evolution operators $U(t)$.

Proposition 0.25. *Let $\chi(\xi)$ be a fixed $C_0^\infty(\mathbb{R}^n)$ function supported in $1/2 \le |\xi| \le 2$ and,*

$$U(t)f(x) = \int e^{i(t|\xi|+x\cdot\xi)}\chi(\xi)\hat{f}(\xi)d\xi. \tag{369}$$

Then,

$$\|U(t)f\|_{L^2} \lesssim C\|f\|_{L^2} \tag{370}$$

$$\|U(t)f\|_{L^\infty} \lesssim (1+|t|)^{-\frac{n-1}{2}}\|f\|_{L^1} \tag{371}$$

from which, interpolating, for all $2 \le r \le \infty$,

$$\|U(t)f\|_{L^r} \lesssim (1+|t|)^{-\frac{n-1}{2}(1-\frac{2}{r})}\|f\|_{L^{r'}} \tag{372}$$

Moreover, if in addition, $\chi = \chi_\mu$ is supported in a cube of size μ, then (371) *can be strengthened to*

$$\|U(t)f\|_{L^\infty} \lesssim \mu(1+|t|)^{-\frac{n-1}{2}}\|f\|_{L^1} \tag{373}$$

Proof We prove directly the stronger version (373). We only need to check (**??**). We write,

$$U(t)f = K_t * f, \qquad K_t(x) = \int e^{i(x\cdot\xi+t|\xi|)}\chi_\mu(\xi)d\xi$$

It suffices to show that,

$$|K_t(x)| \lesssim \mu\frac{1}{(1+|t|+|x|)}$$

In the regions $|x| < |t|/2$ and $|x| \ge 2|t|$ we integrate by parts k times with respect to the operator $L = -i\sum_j \frac{x_j+t\frac{\xi_j}{|\xi}}{|x+t\frac{\xi}{|\xi|}|^2}\partial_{\xi_j}$, such that $L(e^{i(x\cdot\xi+t|\xi|)}) = e^{i(x\cdot\xi+t|\xi|)}$. We also make use of the straightforward estimate, $|\partial_\xi^\alpha\chi_\mu(\xi)| \lesssim \mu^{-|\alpha|}$ to derive, $|K_t(x)| \lesssim (1+|t|)^{-k}\mu^{n-k}$ or, choosing $k = \frac{n-1}{2}$,

$$|K_t(x)| \lesssim (1+|t|)^{-\frac{n-1}{2}}\mu^{\frac{n+1}{2}}.$$

On the other hand, in the region $|t| \approx |x|$, we write, with $\beta(|\xi|)$ vanishing on the support of h_μ,

$$K_t(x) \quad = \quad \int_{1-2\mu}^{1+2\mu} e^{it\lambda}\chi(\lambda)\int_{|\xi|=\lambda} e^{ix\cdot\xi}h_\mu(\xi)d\sigma(\xi)$$

We now need to rely on the following estimate,

$$\sup_{1/2\le\lambda\le 2} |\int_{|\xi|=\lambda} e^{ix\cdot\xi}h(\xi)d\sigma(\xi)| \lesssim (1+|x|)^{-\frac{n-1}{2}} \tag{374}$$

which follows easily from the decay of the Fourier transform of measures supported on $\mathbb{S}^{n-1}$ discussed in the previous section, see lemma 8.8. Therefore, for $|t| \sim |x|$,

$$|K_t(x)| \lesssim \mu(1+|x|)^{-\frac{n-1}{2}} \lesssim \mu(1+|t|)^{-\frac{n-1}{2}}$$

as desired. ■

We are now ready to prove (368) by following the same argument as in the second proof of the restriction theorem. Indeed, in view of (366) and (372) we derive,

$$\|CC^*F\|_{L^r_x}(t) \quad \lesssim \quad \int_{-\infty}^{+\infty}(1+|t-s|)^{-\gamma(r)}\|F(s)\|_{L^{r'}_x}ds \tag{375}$$

where $\gamma(r) = -\frac{n-1}{2}(1-\frac{2}{r})$. We are now precisely in the same situation as in the second proof of the restriction theorem, see the argument following formula (202). If $0 < \gamma(r) < 1$ we can apply the Hardy-Littlewood-Sobolev inequality to obtain

$$\|CC^*F\|_{L^q_t L^r_x} \lesssim \|F\|_{L^{q'}_t L^{r'}_x},$$

when $-\gamma(r) + 1 + 1/q = 1/q'$, hence $\gamma(r) = 2/q$. This proves (362), and thus theorem 0.24, in the case $0 < \gamma(r) = 2/q < 1$. If $q = 2$ and $\gamma(r) > 1$ we have from (375),

$$\|CC^*F\|_{L^2_t L^r_x} \lesssim \|F\|_{L^2_t L^{r'}_x},$$

by an application of the standard Hausdorff-Young inequality.

Finally, if $2/q < 1$ and $\gamma(r) > 2/q$ the result follows from the case $\gamma(r) = 2/q$ using Sobolev inequalities. Due to the fact that one of the principal curvatures of the light cone vanishes, the Strichartz estimates for the wave equation is not as strong as it could be. Using the improved dispersive estimate (373) we can however derive a stronger statement ,which is very useful in applications.

Proposition 0.26. *Let $0 < \mu < 1$. Let f be an L^2 function with Fourier transform supported in a cube of size μ at a distance 1 from the origin. Let (q, r) be an admissable pair of exponents for the Strichartz estimates. Then*

$$\|Cf\|_{L^q_t L^r_x} \lesssim \mu^{\left(\frac{1}{2}-\frac{1}{r}\right)}\|f\|_{L^2}. \tag{376}$$

The proof is based on the improved dispersive estimate (373). Interpolating it with (370) we derive,

$$\|U(t)f\|_{L^r} \quad \lesssim \quad \mu^{1-\frac{2}{r}}(1+|t|)^{-\frac{n-1}{2}(1-\frac{2}{r})}\|f\|_{L^{r'}}$$

The proof the continues exactly as above to derive,

$$\|CC^*F\|_{L^q_t L^r_x} \lesssim \mu^{1-\frac{2}{r}}\|F\|_{L^{q'}_t L^{r'}_x},$$

and therefore, by the TT^* argument, $\|Cf\|_{L^q_t L^r_x} \lesssim \mu^{\left(\frac{1}{2}-\frac{1}{r}\right)}\|f\|_{L^2}$, as desired. As a straightforward corollary to the proposition we derive:

Theorem 0.27. *Consider a general solution of $\Box u = 0$ with data f, g supported, in Fourier space, on a cube of size μ situated in a dyadic shell of size λ, with λ much larger than μ, say $\lambda \geq 8\mu$. Then,*

$$\|u\|_{L^q_t L^r_x} \lesssim \mu^{1-\frac{2}{r}}\left(\|f\|_{\dot{H}^\gamma} + \|f\|_{\dot{H}^{\gamma-1}}\right) \tag{377}$$

Proof The proof follows easily by a scaling argument from the proposition above. ■

Finally we state below another result, which follows easily from the decay estimate (371).

Theorem 0.28. *Let u be a free wave, i.e. solution of the homogeneous equation $\Box u = 0$, with initial data (f, g). Then,*

$$\begin{aligned} \|u(t)\|_{L^\infty} &\lesssim |t|^{-\frac{n-1}{2}} \sum_{\lambda\in 2^{\mathbb{Z}}} \left(\lambda^{\frac{n+1}{2}}\|f_\lambda\|_{L^1} + \lambda^{\frac{n-1}{2}}\|g_\lambda\|_{L^1}\right) \\ &= |t|^{-\frac{n-1}{2}}\left(\|f\|_{\dot{B}^{n+1/2}_{1,1}} + \|g\|_{\dot{B}^{n-1/2}_{1,1}}\right). \end{aligned}$$

The uniform decay rate $|t|^{-\frac{n-1}{2}}$, for large t, plays a very important role in the study of nonlinear perturbations of the standard wave equation.

0.29. Inhomogeneous Strichartz estimates. We have already reduced the inhomogeneous Strichartz estimate (346) of theorem 0.17 to estimate (359). Proceeding as in the case of the homogeneous estimates we can now reduce (359) to the case when the spatial Fourier transform of F is supported in the unit dyadic ring $1/2 \leq |\xi| \leq 2$. Moreover, decomposing u as before in the $\pm$ parts it suffices to prove the estimates separately for u_+ and u_-. Therefore we need to prove,

$$\|u^+|_{L^q([0,T];L^r)} + \|D^{-1}\partial_t u^+\|_{L^q([0,T];L^r)} \lesssim \|F\|_{L^{q'}([0,T];L^{r'})} \tag{378}$$

We have,

$$\begin{aligned} u_+(t,\cdot) &= \int_0^t U(t-s)F(s,\cdot)ds \\ D^{-1}\partial_t u_+(t,\cdot) &= \int_0^t \partial_t D^{-1}U(t-s)F(s,\cdot)ds \end{aligned}$$

Since, in view of the dyadic restriction, $\partial_t D^{-1}U(t) \sim U(t)$ it suffices to prove the estimate for $\|u^+|_{L^q([0,T];L^r)}$. Clearly, u^+ differs from CC^*F in (366) only by the restriction of the interval of integration to $[0,t]$. In view of this fact we write $u_+ = (CC^*)_R F$. We are thus led to the following theorem, from which (378) and thus (346).

Theorem 0.30. *Let $U(t)$ defined as in (369) and let*

$$(CC^*)_R F(t,\cdot) = \int_0^t U(t-s)F(s,\cdot)ds$$

Then, for all admissible pairs (q_1, r_1), (q_2, r_2),

$$\|(CC^*)_R F\|_{L_t^{q_1} L_x^{r_1}([0,T]\times\mathbb{R}^n)} \lesssim \|F\|_{L_t^{q_2'} L_x^{r_2'}([0,T]\times\mathbb{R}^n)} \tag{379}$$

Proof The proof is straightforward in the case $(q_1, r_1) = (q_2, r_2) = (q, r)$. Indeed in this case we can simply repeat the proof of estimate (368) and just take into account the limits of integration. We have also treated the case when $q_1 = \infty$, $r_1 = 2$, see the subsection on energy estimates. The other non-diagonal case cases are a little more difficult and will be treated in the more general abstract setting discuss later in this section. The proof we have given covers however the most interesting case of estimate (346). We have thus given complete proofs for the first two parts of theorem 0.17 ■

0.31. Necessity of the admissibility conditions. To understand what is the optimal range of exponents q and r we consider the analog of the Knapp counterexample in the context of the truncated cone operator C defined in (363).

For some small $\delta > 0$, let

$$D = \{\xi \in \mathbb{R}^n : |\xi_1 - 1| < 1/2, |\xi'| < \delta\},$$

and consider $f = \chi_D$. We have

$$Cf(t,x) = e^{i(t+x_1)} \int_D e^{i[t(|\xi|-\xi_1)+(t+x_1)(\xi_1-1)+x'\cdot\xi']} \mathrm{d}\xi,$$

and observe that

$$|\xi| - \xi_1 = \frac{|\xi'|^2}{|\xi| + \xi_1} \lesssim \delta^2.$$

We can then choose a region of space-time R defined by

$$|t| \lesssim \delta^{-2}, \quad |t + x_1| \lesssim 1, \quad |x'| \lesssim \delta^{-1},$$

such that, when $(t,x) \in R$ and $\xi \in D$, then the oscillatory factor inside the last integral can be treated as a constant. Hence, $|Cf(t,x)| \gtrsim |D|$ for $(t,x) \in R$ and we have

$$\frac{\|Cf\|_{L_t^q L_x^r}}{\|f\|_{L^2}} \gtrsim \frac{|D| \|\chi_R\|_{L_t^q L_x^r}}{|D|^{1/2}} \sim \delta^{\frac{n-1}{2} - \frac{2}{q} - \frac{n-1}{r}}.$$

In the limit $\delta \to 0$, an estimate of the form (368) will necessarily imply that q and r satisfy the condition

$$\frac{2}{q} \leq (n-1)\left(\frac{1}{2} - \frac{1}{r}\right). \tag{380}$$

The other restriction on the range for q, i.e. $q \geq 2$ is a consequence of the invariance of the operator CC^* under time translations. Indeed for translation invariant operators we have the following general result due to Hörmander, [**?**].

Proposition 0.32. *Let* $T : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$ *be a (non trivial) linear operator which commutes with translations, in the sense that* $(Tf) \circ \tau_y = T(f \circ \tau_y)$, *where* $\tau_y(x) = x + y$, *for* $x, y \in \mathbb{R}^n$. *If* T *is bounded from* L^p *to* L^q *then we necessarily have* $q \geq p$.

The proof is based on the following lemma.

Lemma 0.33. *Let $f \in L^p(\mathbb{R}^n)$, then*

$$\lim_{|y|\to\infty} \|f + f \circ \tau_y\|_{L^p} = 2^{1/p}\|f\|_{L^p}.$$

Proof For every $R > 0$ consider the decomposition $f = g_R + h_R$, where $g_R(x) = f(x)$ if $|x| < R$ and 0 if $|x| \geq R$, and $h_R(x) = 0$ if $|x| < R$ and $f(x)$ if $|x| \geq R$. Then

$$\lim_{R\to\infty} \|g_R\|_{L^p} = \|f\|_{L^p}, \quad \lim_{R\to\infty} \|h_R\|_{L^p} = 0.$$

For $R = |y|/2$ we have

$$f + f \circ \tau_y = g_R + g_R \circ \tau_y + h_R + h_R \circ \tau_y.$$

The functions g_R and $g_R \circ \tau_y$ have disjoint supports, so that

$$\|g_R + g_R \circ \tau_y\|_{L^p}^p = \|g_R\|_{L^p}^p + \|g_R \circ \tau_y\|_{L^p}^p = 2\|g_R\|_{L^p}^p,$$

while

$$\lim_{|y|\to\infty} \|h_R + h_R \circ \tau_y\|_{L^p} \leq \lim_{|y|\to\infty} 2\|h_R\|_{L^p} = 0,$$

hence

$$\lim_{|y|\to\infty} \|f + f \circ \tau_y\|_{L^p} = \lim_{|y|\to\infty} 2^{1/p}\|g_R\|_{L^p} = 2^{1/p}\|f\|_{L^p}.$$

■

Proof [Proof of Proposition 0.32] Let $C > 0$ be the optimal constant for the estimate

$$\|Tf\|_{L^q} \leq C\|f\|_{L^p}, \quad \forall f \in L^p.$$

Then by linearity and the translation invariance,

$$\|Tf + (Tf) \circ \tau_y\|_{L^q} \leq C\|f + f \circ \tau_y\|_{L^p}.$$

When $|y| \to \infty$, applying the lemma we obtain

$$2^{1/q}\|Tf\|_{L^q} \leq C2^{1/p}\|f\|_{L^p}, \quad \forall f \in L^p.$$

The optimality of C implies that $2^{\frac{1}{p}-\frac{1}{q}} \geq 1$, hence $q \geq p$. ■

The proposition generalizes easily to vector valued L^p spaces and if we consider CC^* as an operator from $L^{q'}(\mathbb{R}; L_x^{r'})$ to $L^q(\mathbb{R}; L_x^r)$, then we must have $q \geq q'$, which is the condition $q \geq 2$.

0.34. A general, abstract framework. It turns out that the method of proving Strichartz estimates described above applies to many other equations, such as Schrödinger, KdV etc. It thus pays to have a general framework which applies to all these cases.

Let $(X, \mathrm{d}\mu)$ be a measure space and H a Hilbert space. Consider a family $(U(t))_{t\in\mathbf{R}}$ of operators $U(t) : H \to L^2(X)$, which describes the *evolution* of some system with *data* in H. We assume that this evolution satisfies the following two properties:

- for all $t \in \mathbf{R}$ and $f \in H$ we have the energy estimate:
$$\|U(t)f\|_{L^2(X)} \lesssim \|f\|_H; \tag{381}$$
- for all $t \neq s$ and $g \in L^1(X)$ we have the dispersive inequality:
$$\|U(t)U^*(s)g\|_{L^\infty(X)} \lesssim |t-s|^{-\gamma_0}\|g\|_{L^1(X)}, \tag{382}$$
for some $\gamma_0 > 0$.

Interpolating between (381) and (382) we obtain the estimate
$$\|U(t)U^*(s)g\|_{L^r(X)} \lesssim |t-s|^{-\gamma(r)}\|g\|_{L^{r'}(X)}, \tag{383}$$
for $r \geq 2$, where
$$\gamma(r) = \gamma_0\left(1 - \frac{2}{r}\right).$$

Theorem 0.35. *If the evolution operator $U(t)$ satisfies* (381) *and* (382)*, then the estimates*
$$\|U(t)f\|_{L^q_t L^r_X} \lesssim \|f\|_H, \tag{384}$$
hold for all $q, r \geq 2$ verifing:
$$\frac{2}{q} = \gamma(r), \quad (q, r, \gamma_0) \neq (2, \infty, 1). \tag{385}$$

Remark 0.36. This form of the Strichartz inequalities applies to linear dispersive equations such as Schrödinger.

Proof If we consider the operator $T : H \to L^q_t L^r_X$ defined by $Tf(t,x) = (U(t)f)(x)$ then it is easy to verify that the dual of T is the operator $T^* : L^{q'}_t L^{r'}_X \to H$ given by $T^*F = \int U^*(s)F(s,\cdot)\mathrm{d}s$. By the TT^* method, (384) is then equivalent to the estimate
$$\|\int U(t)U^*(s)F(s)\mathrm{d}s\|_{L^q_t L^r_X} \lesssim \|F\|_{L^{q'}_t L^{r'}_X}. \tag{386}$$
By duality and symmetry considerations, this is in turn equivalent to
$$|B(F,G)| \lesssim \|F\|_{L^{q'}_t L^{r'}_X}\|G\|_{L^{q'}_t L^{r'}_X}, \tag{387}$$
where $B(F,G)$ is the bilinear form
$$B(F,G) = \iint_{s<t} \langle U^*(t)F(t), U^*(s)G(s)\rangle \mathrm{d}t\mathrm{d}s. \tag{388}$$

From the bilinear version of (383) we have that
$$|B(F,G)| \lesssim \iint \frac{\|F(t)\|_{L^{r'}}\|G(s)\|_{L^{r'}}}{|t-s|^{\gamma(r)}}\mathrm{d}s\mathrm{d}t. \tag{389}$$
If $\gamma(r) < 1$, we can apply the Hardy-Littlewood-Sobolev inequality and obtain (387). This concludes the proof for the cases $q = 2/\gamma(r) > 2$.

The endpoint case, corresponding to $\gamma(r) = 2/q = 1$, is allowed when $r < \infty$. Its proof will be described in the next section. ∎

Remark 0.37. If we strengthen the dispersive condition (382) to

$$\|U(t)U^*(s)g\|_{L^\infty(X)} \lesssim (1+|t-s|)^{-\gamma_0}\|g\|_{L^1(X)}, \tag{390}$$

then (389) can be improved to

$$|B(F,G)| \lesssim \iint \frac{\|F(t)\|_{L^{r'}}\|G(s)\|_{L^{r'}}}{(1+|t-s|)^{\gamma(r)}}\mathrm{d}s\mathrm{d}t. \tag{391}$$

Now we can obtain (387) from Young's inequality when $2/q = 1/p$ and $(1+|t|)^{-\gamma(r)} \in L^p(\mathbf{R})$, i.e. $\gamma(r)p > 1$. Hence, (390) allows us to extend the Strichartz estimates (384) in Theorem 0.35 to the range

$$\frac{2}{q} \leq \gamma(r), \quad (q,r,\gamma_0) \neq (2,\infty,1). \tag{392}$$

This case applies to the linear wave equations.

Remark 0.38. We observe that there is a natural scaling associated to the objects in this abstract formulation. More precisely, the estimates (384) in Theorem 0.35 are invariant under the change of scale defined by

$$U(t) \leftarrow U(t/\lambda), \quad U^*(s) \leftarrow U^*(s/\lambda), \quad \mathrm{d}\mu \leftarrow \lambda^{\gamma_0}\mathrm{d}\mu, \quad \langle f,g\rangle_H \leftarrow \lambda^{\gamma_0}\langle f,g\rangle_H. \tag{393}$$

We can also consider the endpoint case.

$$q=2, \quad r=\frac{2\gamma_0}{\gamma_0-1}, \quad \gamma_0>1.$$

This, in fact, is more difficult than the previous non-endpoint case, and requires a two-parameter estimate which is better than the one-parameter family given by the interpolation (383). This proof is presented in the previously mentioned paper by Keel and Tao, "Endpoint Strichartz Estimates". We omit it here.

0.39. Inhomogeneous estimates. Saying that an operator T maps the Hilbert space H into $L_t^q L_X^r$, is equivalent to saying that its dual T^* maps $L_t^{q'} L_X^{r'}$ into H, and is also equivalent to saying that the TT^* operator maps $L_t^{q'} L_X^{r'}$ into $L_t^q L_X^r$. If the pair (q,r) is allowed to vary in a set E of admissable exponents, we can view TT^* as a composition of two operators associated with different pairs of exponents. It follows that TT^* actually satisfies a larger set of mapping properties, since it maps $L_t^{\tilde{q}'} L_X^{\tilde{r}'}$ into $L_t^q L_X^r$, for any couple of pairs $(q,r),(\tilde{q},\tilde{r}) \in E$.

The operator $Tf(t) = U(t)f$ defined in the previous subsection can be viewed as the solution of some homogenous, translation invariant, linear evolution equation. The solution of the corresponding inhomogenoues problem, using Duhamel's principle, would be represented by the retarded operator

$$RF(t) = \int_{s<t} U(t)U^*(s)F(s)\mathrm{d}s.$$

Observe that operator R looks very similar to the TT^* operator, which is given by

$$TT^*F(t) = \int U(t)U^*(s)F(s)\mathrm{d}s.$$

The restriction $s < t$ in the definition of R, however, destroys the composition structure of TT^*. Fortunately, all the mapping properties of TT^*, which we have derived above, can be transfered to R.

Theorem 0.40. *The operator R maps $L_t^{\tilde{q}'} L_X^{\tilde{r}'}$ into $L_t^q L_X^r$, for any couple of pairs $(q, r), (\tilde{q}, \tilde{r})$ for which the Strichartz estimate 384 holds.*

Proof First of all observe that in the proof of theorem 0.35 we have actually proved the diagonal case $(q, r) = (\tilde{q}, \tilde{r})$. Indeed, the bilinear form defined in (388) can be written as $B(F, G) = \iint R(F) \cdot G \mathrm{d}x\mathrm{d}t$ and (387) is the dual formulation of the mapping property for R.

The non diagonal cases with $\frac{1}{q} + \frac{1}{\tilde{q}} < 1$ follow from the mapping properties of TT^* by using a general argument about integral operators due to Christ and Kiselev (see [] and []) which we summarize in Proposition 0.42 below.

It remains to consider the cases with $q = \tilde{q} = 2$ and $r \neq \tilde{r}$, under the assumption that the evolution $U(t)$ satisfies the stronger dispersive inequality (390) with $\gamma_0 > 1$. Since, we have already proved the case $r = \tilde{r}$, by interpolation it is enough to consider the extreme case: $r = r_* = \frac{2\gamma_0}{\gamma_0 - 1}$, $\tilde{r} = \infty$, and show that

$$|B(F, G)| \lesssim \|F\|_{L_t^2 L_X^{r_*'}} \|G\|_{L_t^2 L_X^1}.$$

This estimate follows by decomposing $B(F, G)$ into dyadic pieces, $B = \sum_{\lambda \in 2^{\mathbb{Z}}} B_\lambda$, where

$$B_\lambda(F, G) = \iint_{\lambda/2 \leq |t-s| \leq 2\lambda} \langle U^*(t)F(t), U^*(s)G(s)\rangle \mathrm{d}t\mathrm{d}s, \tag{394}$$

The desired conclusion follows immediately from the lemma below. ■

Lemma 0.41. *Let $B_\lambda(F, G)$ be the bilinear form defined in* (394)*. Then, there exists an $\varepsilon > 0$ such that*

$$|B_\lambda(F, G)| \lesssim \min\left\{\lambda, \lambda^{-1}\right\}^\varepsilon \|F\|_{L_t^2 L_X^{r_*'}} \|G\|_{L_t^2 L_X^1}.$$

Proof We may assume that F and G are supported on disjoint time intervals of length $O(\lambda)$ separated by a distance $O(\lambda)$. Then $B_\lambda(F, G) = \langle T^*F, T^*G\rangle_H$. We use the energy estimate to bound $\|T^*F\|_H$ and the Strichartz estimate with $q = 2$ and $r = \infty$ to bound $\|T^*G\|_H$, so that

$$|B_\lambda(F, G)| \lesssim \|F\|_{L_t^1 L_X^2} \|G\|_{L_t^2 L_X^1}.$$

We then apply Holder inequality and use the assumption on the support of F to obtain

$$|B_\lambda(F, G)| \lesssim \lambda^{1/2} \|F\|_{L_t^2 L_X^2} \|G\|_{L_t^2 L_X^1}.$$

We can also write $B_\lambda(F, G) = \iiint F(t) \cdot U(t)U^*(s)G(s) \mathrm{d}x\mathrm{d}s\mathrm{d}t$ and make use of the dispersive inequality,

$$|B_\lambda(F, G)| \lesssim (1 + \lambda)^{-\gamma_0} \|F\|_{L_t^1 L_X^1} \|G\|_{L_t^1 L_X^1}.$$

Again, we apply Holder inequality and use the assumption on the support of F and G to obtain

$$|B_\lambda(F,G)| \lesssim \frac{\lambda}{(1+\lambda)^{\gamma_0}} \|F\|_{L^2_t L^1_X} \|G\|_{L^2_t L^1_X}.$$

Hence, B_λ is bounded on $L^2_t L^2_X \times L^2_t L^1_X$ with constant $\lambda^{1/2}$ and on $L^2_t L^1_X \times L^2_t L^1_X$ with constant $\frac{\lambda}{(1+\lambda)^{\gamma_0}}$. By standard interpolation of L^p spaces we obtain that B_λ is bounded on $L^2_t L^{r'_*}_X \times L^2_t L^1_X$ with constant C_λ, where

$$C_\lambda = \lambda^{\theta/2} \left(\frac{\lambda}{(1+\lambda)^{\gamma_0}} \right)^{1-\theta}, \quad \frac{1}{r'_*} = \frac{\theta}{2} + \frac{1-\theta}{1}, \quad r_* = \frac{2\gamma_0}{\gamma_0 - 1}.$$

Simplyfing the expression we find that

$$C_\lambda = \frac{\lambda^{\frac{\gamma_0+1}{2\gamma_0}}}{1+\lambda} \lesssim \min\left\{\lambda, \lambda^{-1}\right\}^{\varepsilon},$$

with

$$\varepsilon = \min\left\{ \frac{\gamma_0+1}{2\gamma_0}, 1 - \frac{\gamma_0+1}{2\gamma_0} \right\} = \frac{\gamma_0 - 1}{2\gamma_0} = \frac{1}{r_*} > 0.$$

■

0.41.1. *Integral operators with restricted kernel.* In this subsection we give a self contained exposition of the results of Christ-Kisselev mentioned above. Consider an integral operator with a measurable kernel $K(s,t)$,

$$Tf(t) = \int_{\mathbb{R}} K(s,t) f(s) \mathrm{d}s,$$

and its restricted version associated with the kernel $K(s,t)\chi(s<t)$,

$$Rf(t) = \int_{s<t} K(s,t) f(s) \mathrm{d}s.$$

If T maps L^p into L^q and $1 \le p < q \le \infty$ then we have that R also maps L^p into L^q. An equivalent formulation of this fact is given in the following proposition.

Proposition 0.42. *Let $K(s,t)$ be a measurable function on $\mathbf{R} \times \mathbf{R}$. Let $B(f,g)$ be the bilinear form with kernel K,*

$$B(f,g) = \iint K(s,t) f(s) g(t)\, ds\, dt,$$

and $\widetilde{B}(f,g)$ the bilinear form with kernel restricted to the region $s<t$,

$$\widetilde{B}(f,g) = \iint_{s<t} K(s,t) f(s) g(t)\, ds\, dt.$$

Let $p,q \ge 1$, with the condition

$$\frac{1}{p} + \frac{1}{q} > 1. \tag{395}$$

If B is bounded on $L^p \times L^q$,

$$|B(f,g)| \lesssim \|f\|_{L^p} \|g\|_{L^q},$$

then $\widetilde{B}$ is also bounded on $L^p \times L^q$,

$$\left|\widetilde{B}(f,g)\right| \lesssim \|f\|_{L^p}\|g\|_{L^q}.$$

Remark 0.43. There are cases for which equality in condition (395) is not allowed. Consider for the example the case of the Hilbert transform, which corresponds to the kernel $K(s,t) = \frac{1}{s-t}$, with $p = q = 2$.

Proof Let $f \in L^p$ and $g \in L^q$ with $\|f\|_{L^p} = \|g\|_{L^q} = 1$.

Define $F(t) = \int_{s<t} |f(s)|^p \mathrm{d}s$. F is a continuous non-decreasing function which maps $[-\infty, +\infty]$ onto $[0,1]$. In particular, the inverse image of an interval of the type $I = [a,b] \subset [0,1]$ will be an interval of the same type, $F^{-1}(I) = [A,B]$, with $F(A) = a$, $F(B) = b$, and $\int_A^B |f(s)|^p \mathrm{d}s = F(B) - F(A) = b - a$. Hence,

$$\|f\|_{L^p(F^{-1}(I))} = |I|^{1/p}. \tag{396}$$

Consider now a Whitney decomposition of the set $\Omega = \left\{(x,y) \in \mathbb{R}^2 : x < y\right\}$ into disjoint dyadic squares, as in Lemma 6.7, $\Omega = \cup_Q Q$, where each square $Q = I \times J$ has the property

$$\mathrm{dist}\,(I,J) \approx |I| = |J| = \lambda, \tag{397}$$

for some dyadic value of λ. If we look only at those squares needed to cover the triangle $\Omega \cap [0,1]^2$, then $\lambda \leq 1/2$.

Observe that $s < t$ implies that either $F(s) < F(t)$ or $f \equiv 0$ almost everywhere on the interval $[s,t]$. Hence, we can write

$$\widetilde{B}(f,g) = \iint_{F(s)<F(t)} K(s,t) f(s) g(t) \mathrm{d}s\mathrm{d}t = \sum_Q B(\chi_{F^{-1}(I)} f, \chi_{F^{-1}(J)} g).$$

Using the boundedness of B on $L^p \times L^q$ we obtain

$$\left|\widetilde{B}(f,g)\right| \lesssim \sum_Q \|f\|_{L^p(F^{-1}(I))} \|g\|_{L^q(F^{-1}(J))}.$$

Now we use (396), (397) and the fact that, for each given dyadic interval J, the number of intervals I for which $I \times J$ is one of the squares in the decomposition of Ω is bounded by a universal constant. Hence,

$$\left|\widetilde{B}(f,g)\right| \lesssim \sum_{\lambda \leq 1/2} \lambda^{\frac{1}{p}} \sum_{|J|=\lambda} \|g\|_{L^q(F^{-1}(J))}.$$

Next, we apply Hölder's inequality to the summation over the dyadic intervals J of length λ and since there are λ^{-1} of them in $[0,1]$ we have

$$\left|\widetilde{B}(f,g)\right| \lesssim \sum_{\lambda \leq 1/2} \lambda^{\frac{1}{p}} \lambda^{-\frac{1}{q'}} \|g\|_{L^q} = \sum_{\lambda \leq 1/2} \lambda^{\frac{1}{p}+\frac{1}{q}-1} \lesssim 1.$$

$\blacksquare$

CHAPTER 9

Bilinear Estimates

1. Bilinear proofs of some Strichartz estimates

Consider the homogeneous wave equation $\Box u = 0$ in $\mathbb{R}^{1+3}$. The Strichartz estimate (345) with $q = r = 4$ and $\gamma = 1/2$. Takes the form,

$$\|u\|_{L^4(\mathbb{R}^{1+3})} \lesssim \|f\|_{\dot{H}^{1/2}} + \|g\|_{\dot{H}^{-1/2}}$$

Writing $u = u^+ + u^-$ it suffices to prove,

$$\|u^+\|_{L^4(\mathbb{R}^{1+3})} \lesssim \|f^+\|_{\dot{H}^{1/2}} \tag{398}$$

where

$$u_+(t,x) = \int e^{ix\cdot\xi + t|\xi|}\hat{f}(\xi)d\xi$$

Clearly,

$$\|u^+\|^2_{L^4(\mathbb{R}^{1+3})} = \|u^+ \cdot u^+\|_{L^2} = \|\widetilde{u^+} * \widetilde{u^+}\|_{L^2}$$

Now, recalling (360), and dropping the index $+$,

$$\begin{aligned}\tilde{u} * \tilde{u}(\tau,\xi) &= \int\int \delta(\tau - \lambda - |\xi - \eta|)\hat{f}(\xi - \eta)\delta(\lambda - |\eta|)\hat{f}(\eta)d\lambda d\eta \\ &= \int \delta((\tau - |\eta| - |\xi - \eta|)\hat{f}(\eta)\hat{f}(\xi - \eta)d\eta\end{aligned}$$

Clearly, (398) follows from the following:

Theorem 1.1. *The bilinear operator,*

$$B(F,G) = \int \delta(\tau - |\eta| - |\xi - \eta|)\frac{F(\xi - \eta)}{|\xi - \eta|^{1/2}}\frac{G(\eta)}{|\eta|^{1/2}}d\eta.$$

verifies the estimate,

$$\|B(F,G)\|_{L^2(\mathbb{R}^{1+3})} \lesssim \|F\|_{L^2(\mathbb{R}^3)}\|G\|_{L^2(\mathbb{R}^{1+3})} \tag{399}$$

Proof By Cauchy-Schwartz,

$$\begin{aligned}|B(F,G)(\tau,\xi)|^2 &\lesssim J(\tau,\xi)\int \delta(\tau - |\eta| - |\xi - \eta|)|F(\xi - \eta)|^2|G(\eta)|^2 d\eta \\ J(\tau,\xi) &= \int \delta(\tau - |\eta| - |\xi - \eta|)\frac{1}{|\xi - \eta|}\frac{1}{|\eta|}d\eta\end{aligned}$$

It suffices to show that J is uniformly bounded. Indeed, if that is the case,

$$\begin{aligned}\|B(F,G)\|_{L^2(\mathbb{R}^{1+3})} &\lesssim \sup_{\tau,\xi} J(\tau,\xi)\int\int \delta(\tau-|\eta|-|\xi-\eta|)F(\xi-\eta)|^2|G(\eta)|^2 d\eta d\tau d\xi \\ &\lesssim \sup_{\tau,\xi} J(\tau,\xi)\|F\|_{L^2}^2\|G\|_{L^2}^2\end{aligned}$$

Therefore the bilinear estimate is an immediate consequence of the uniform boundedness of J. This follows from the following more general lemma below. ■

Lemma 1.2. *Let F be an arbitrary function of two variables and J_F the integral*

$$J_F^{\mp}(\tau,\xi) = \int_{\mathbb{R}^n}\delta(\tau-|\eta|\mp|\xi-\eta|)F(|\eta|,|\xi-\eta|)$$

Then,

$$J_F^{-}(\tau,\xi) = (\tau^2-|\xi|^2)^{\frac{n-3}{2}}\int_{-1}^{1} F\big(\frac{\tau+s|\xi|}{2},\frac{\tau+s|\xi|}{2}\big)(\tau^2-x^2|\xi|^2)(1-|x|^2)^{\frac{n-3}{2}}dx, \tag{400}$$

$$J_F^{+}(\tau,\xi) = (\tau^2-|\xi|^2)^{\frac{n-3}{2}}\int_{1}^{\infty} F\big(\frac{\tau+s|\xi|}{2},\frac{\tau+s|\xi|}{2}\big)(\tau^2-x^2|\xi|^2)(1-|x|^2)^{\frac{n-3}{2}}dx \tag{401}$$

Proof : Observe that in the case $\mp = -$ the measure $\delta(\tau-|\eta|-|\xi-\eta|)$ is supported on the *ellipsoid* of revolution with foci at 0 and ξ, $\mathcal{E}(\tau,\xi)=\{\eta\in\mathbb{R}^n : |\eta|+|\xi-\eta|=\tau\}$,. In this case $|\xi|\le\tau$. In the $\mp=+$ the measure $\delta(\tau-|\eta|+|\xi-\eta|)$ is supported in the *hyperboloid* of revolution with foci at 0 and ξ, $\mathcal{H}(\tau,\xi)=\{\eta\in\mathbb{R}^n : |\eta|-|\xi-\eta|=\tau\}$, which is an unbounded hypersurface with infinite volume. In this case $|\xi|^2\le\tau^2$. In the sense of distributions, we have the identity

$$\begin{aligned}\delta(\tau-|\eta|\mp|\xi-\eta|) &= \delta\big(\frac{(\tau-|\eta|)^2-|\xi-\eta|^2}{2(\tau-|\eta|)}\big)\\ &= 2(\tau-|\eta|)\delta\big((\tau-|\eta|)^2-|\xi-\eta|^2\big)\\ &= 2(\tau-|\eta|)\delta\big(\tau^2-|\xi|^2-2\tau\lambda+2\lambda\xi\cos\theta\big)\\ &= 2(\tau-|\eta|)\delta\big(\tau^2-|\xi|^2-2\tau\lambda+2a|\xi|\big)\end{aligned}$$

with a the cosine of the angle between η and ξ. Thus, for fixed τ and ξ we must have, on the support of the measure,

$$a = -\frac{\tau^2-|\xi|^2-2\tau\lambda}{2|\xi|\lambda} \tag{402}$$

Observe that in the ellipsoidal case a can take any values in the interval $[-1,1]$ and thus, since $\lambda=\frac{\tau^2-|\xi|^2}{2(\tau-a|\xi|)}$, we have $\frac{\tau-|\xi|}{2}\le\lambda\le\frac{\tau+|\xi|}{2}$. On the other hand, in the hyperboloidal case when $|\xi|^2>\tau^2$, we must also have the restriction,

$$\frac{\tau}{|\xi|}\le a.$$

and thus, $\lambda=\frac{-\tau^2+|\xi|^2}{2(-\tau+a|\xi|)}\ge\frac{\tau+|\xi}{2}$.

Thus, since $d\eta = \lambda^{n-1} d\lambda dS_\omega = (1-a^2)^{\frac{n-3}{2}} \lambda^{n-1} d\lambda dS_{\omega'}$,

$$
\begin{aligned}
J_F^- &= \frac{1}{|\xi|} \int_{\frac{\tau-|\xi|}{2}}^{\frac{\tau+|\xi|}{2}} F(\lambda, \tau-\lambda)(\tau-\lambda)\lambda^{n-2} \big[1 - \big(\frac{\tau^2 - |\xi|^2 - 2\tau\lambda}{2|\xi|\lambda}\big)\big]^{\frac{n-3}{2}} d\lambda \\
&= \frac{(\tau^2 - |\xi|^2)^{\frac{n-3}{2}}}{|\xi|^{n-2}} \int_{\frac{\tau-|\xi|}{2}}^{\frac{\tau+|\xi|}{2}} F(\lambda, \tau-\lambda)(\tau-\lambda)\lambda \big[\big(\frac{\tau+|\xi|}{2} - \lambda\big)\big(\lambda - \frac{\tau - |\xi|}{2}\big)\big]^{\frac{n-3}{2}}
\end{aligned}
$$

At last we perform the change of variables $x = \frac{2\lambda - \tau}{|\xi|}$ to derive the desired formula (400). The proof for (401) follows in the same manner.

■

2. Improved Bilinear Strichartz

Consider two solutions of the homogeneous wave equations, $\square u = \square v = 0$. For simplicity, and without loss of generality, we assume that u, v verify the reduced initial data at $t = 0$,

$$u(0,x) = f(x),\ v(0,x) = g(x),\ \partial_t u(0,x) = \partial_t v(0,x) = 0.$$

We consider estimates of the form,

$$\|D^{-b}(uv)\|_{L_t^{q/2} L_x^{r/2}} \lesssim \|f\|_{\dot{H}^a} \|g\|_{\dot{H}^a}$$

with (q, r) an acceptable pair. By dimensional analysis and recalling the exponent $\gamma = n((\frac{1}{2} - \frac{1}{r})) - \frac{1}{q}$ in (345), we must have,

$$2a = -b + 2\big(n(\frac{1}{2} - \frac{1}{r}) - \frac{1}{q}\big) = -b + 2\gamma \tag{403}$$

We decompose the product $u \cdot v$ by the trichotomy formula,

$$
\begin{aligned}
u \cdot v &= \sum_{\mu<\lambda} u_\mu v_\lambda + \sum_{\mu<\lambda} v_\mu u_\lambda + \sum_{\mu\le\lambda} P_\mu(u_\lambda v_\lambda) \\
&= (u\cdot v)_{LH} + (u\cdot v)_{HL} + (u\cdot v)_{HH}
\end{aligned}
$$

Here $\mu, \lambda \in 2^{\mathbb{Z}}$, $u_\lambda = P_\lambda u$ and P_λ the usual LP projections. Now,

$$\|D^{-b}(uv)_{LH}\|_{L_t^{q/2} L_x^{r/2}} \le \sum_{\mu\le\lambda} \lambda^{-b} \|u_\mu v_\lambda\|_{L_t^{q/2} L_x^{r/2}} \le \sum_{\mu\le\lambda} \lambda^{-b} \|u_\mu\|_{L_t^q L_x^r} \|v_\lambda\|_{L_t^q L_x^r}$$

in view of the Strichartz estimates of the previous section

$$
\begin{aligned}
\|u_\mu\|_{L_t^q L_x^r} &\lesssim \mu^{(\gamma-a)} \|f_\mu\|_{\dot{H}^a} = \mu^{b/2} \|f_\mu\|_{\dot{H}^a} \\
\|v_\lambda\|_{L_t^q L_x^r} &\lesssim \lambda^{(\gamma-a)} \|g_k\|_{\dot{H}^a} = \lambda^{b/2} \|g_\lambda\|_{\dot{H}^\alpha}
\end{aligned}
$$

and therefore, for $b > 0$,

$$
\begin{aligned}
\|D^{-b}(uv)_{LH}\|_{L_t^{q/2} L_x^{r/2}} &\lesssim \sum_{\mu\le\lambda} \big(\frac{\mu}{\lambda}\big)^b \|f_\mu\|_{\dot{H}^a} \|g_\lambda\|_{\dot{H}^a} \\
&\lesssim \|f\|_{\dot{H}^a} \|g\|_{\dot{H}^a}
\end{aligned}
$$

By symmetry,

$$\|D^{-b}(uv)_{LH}\|_{L_t^{q/2}L_x^{r/2}} \quad \lesssim \quad \|f\|_{\dot H^a}\|g\|_{\dot H^a}$$

It thus only remains to estimate the high-high term $\|(u\cdot v)_{HH}\|_{L_t^{q/2}L_x^{r/2}}$. This requires a more subtle argument based on theorem **??**. We write,

$$\|D^{-b}(u\cdot v)_{HH}\|_{L_t^{q/2}L_x^{r/2}} \lesssim \sum_{\mu\le\lambda}\mu^{-b}\|P_\mu(u_\lambda v_\lambda)\|_{L_t^{q/2}L_x^{r/2}}$$

If we use the standard Strichartz estimate, i.e.,

$$\begin{aligned}\|P_\mu(u_\lambda v_\lambda)\|_{L_t^{q/2}L_x^{r/2}} \quad &\lesssim \quad \|u_\lambda\|_{L_t^qL_x^r}\|v_\lambda\|_{L_t^qL_x^r}\lambda^{2(\gamma-a)}\|f\|_{\dot H^a}\|g\|_{\dot H^a}\\ &= \quad \lambda^b\|f\|_{\dot H^a}\|g\|_{\dot H^a} \end{aligned} \tag{404}$$

we would derive,

$$\|D^{-b}(u\cdot v)_{HH}\|_{L_t^{q/2}L_x^{r/2}} \lesssim \sum_{\mu\le\lambda}\lambda^b\mu^{-b}\|f\|_{\dot H^a}\|g\|_{\dot H^a}$$

which diverges. We need to replace (404) by a stronger estimate which takes into account the presence of P_μ in front of $u_\lambda v_\lambda$. To achieve this, we need first to exploit some orthogonality properties. We decompose the the data f_λ, g_λ, in Fourier space, into pieces supported on cubes of size μ, $f_\lambda = \sum_Q f_Q$, $g_\lambda = \sum_Q g_Q$ and denote by u_Q, v_Q the corresponding solutions. Clearly the decomposition commutes with the wave operator $\Box$. Thus, $u_\lambda \sim \sum_Q u_Q$, $v_\lambda \sim \sum_Q v_Q$ and

$$P_\mu(u_\lambda\cdot v_\lambda) \quad \sim \quad \sum_{Q_1,Q_2} P_\mu\big(u_{Q_1}v_{Q_2}\big)$$

Observe that $P_\mu(u_{Q_1}u_{Q_2}) \neq 0$ only if Q_1+Q_2 intersects the region of frequencies of size μ where P_μ is supported. For each cube Q_1, of size μ, there are only a finite number (which depends only on n) of cubes Q_2 for which this happens. Morally, by enlarging the cubes if necessary we may assume that $Q_2 = -Q_1$ and thus,

$$P_\mu(u_\lambda\cdot v_\lambda) \quad \sim \quad \sum_Q u_Q v_{-Q}.$$

Hence,

$$\|P_\mu(u_\lambda v_\lambda)\|_{L_t^{q/2}L_x^{r/2}} \lesssim \sum_Q \|u_Q v_{-Q}\|_{L_t^qL_x^r} \lesssim \sum_Q \|u_Q\|_{L_t^qL_x^r}\|v_{-Q}\|_{L_t^qL_x^r}.$$

We are now in a position to apply theorem 0.27. Thus,

$$\|u_Q\|_{L_t^qL_x^r} \quad \lesssim \quad \big(\frac{\mu}{\lambda}\big)^{\frac12-\frac1r}\|f_Q\|_{\dot H^\gamma}$$

and similarly for v_{-Q}. Hence,

$$\begin{aligned}\|P_\mu(u_\lambda v_\lambda)\|_{L_t^{q/2}L_x^{r/2}} \quad &\lesssim \quad \big(\frac{\mu}{\lambda}\big)^{1-\frac2r}\sum_Q \|f_Q\|_{\dot H^\gamma}\|g_Q\|_{\dot H^\gamma}\\ &\lesssim \quad \big(\frac{\mu}{\lambda}\big)^{1-\frac2r}\|f_\lambda\|_{\dot H^\gamma}\|g_\lambda\|_{\dot H^\gamma}\\ &\lesssim \quad \big(\frac{\mu}{\lambda}\big)^{1-\frac2r}\lambda^{2\gamma-2a}\|f_\lambda\|_{\dot H^a}\|g_\lambda\|_{\dot H^a}\\ &\lesssim \quad \big(\frac{\mu}{\lambda}\big)^{1-\frac2r}\lambda^{b}\|f_\lambda\|_{\dot H^a}\|g_\lambda\|_{\dot H^a}\end{aligned}$$

and, consequently,

$$\begin{aligned}\|D^{-b}(u\cdot v)_{HH}\|_{L_t^{q/2}L_x^{r/2}} &\lesssim \sum_{\mu<\lambda}\big(\frac{\mu}{\lambda}\big)^{1-\frac{2}{r}-b}\|f_\lambda\|_{\dot H^a}\|g_\lambda\|_{\dot H^a}\\ &\lesssim \|f\|_{\dot H^a}\|g\|_{\dot H^a}\end{aligned}$$

provided that $b<1-\frac{2}{r}$. We have just proved the following bilinear estimate, see [**?**].

Theorem 2.1. *The following estimate* [1] *holds for solutions* $\Box u=\Box v=0$*, any admissible pair* (q,r) *and any* $0\le b<1-\frac{2}{r}$*,*

$$\|D^{-b}(u\cdot v)_{L_t^{q/2}L_x^{r/2}} \lesssim \|u[0]\|_{\dot H^a}\|v[0]\|_{\dot H^a} \tag{405}$$

provided that the dimensional condition,

$$a=-\frac{b}{2}+\gamma,\qquad \gamma=n(\frac12-\frac1r)-\frac1q \tag{406}$$

3. Bilinear estimates for null forms.

In this subsection we discuss the simplest bilinear estimates for null quadratic forms, see [**?**], [**?**], [**?**] and [**?**].

Definition 3.1. Let u,v be two smooth solutions of $\Box=\Box v=0$ on $\mathbb{R}^{n+1}$. The standard null quadratic forms are $Q_0(u,v)=-\partial_t u\,\partial_t v+\sum_{i=1}^n\partial_i u\partial_i v$, as well as $Q_{ij}(u,v)=\partial_i u\partial_j v-\partial_i v\partial_j u$, and $Q_{0i}(u,v)=\partial_i u\partial_t v-\partial_i v\partial_t u$ for $i,j=1,\dots,n$.

Theorem 3.2. *For any null form* Q *and any solutions to* $\Box=\Box v=0$ *on* $\mathbb{R}^{n+1}$*,* $n\ge2$*, we have,*

$$\|Q(u,v)\|_{L^2(\mathbb{R}^{n+1})} \lesssim \|u[0]\|_{\dot H^1(\mathbb{R}^n)}\|v[0]\|_{H^{\frac{n+1}{2}}(\mathbb{R}^n)} \tag{407}$$

Remark 3.3. Without loss of generality, it suffices to consider the reduced initial value problems

$$u(0,x)=f(x),\ v(0,x)=g(x),\ \partial_t u(0,x)=\partial_t v(0,x)=0 \tag{408}$$

In what follows we show how to deduce the estimate (3.2) from a more general form of bilinear estimates presented in the next section.

Definition 3.4. Let D^α, D_+^α and D_-^α be the operators in $\mathbb{R}^{n+1}$ defined by the multipliers with symbols, respectively

$$|\xi|^\alpha,\quad (|\tau|+|\xi|)^\alpha,\quad \big||\tau|-|\xi|\big|^\alpha.$$

Observe that we can write, for any smooth functions u,v,

$$2Q_0(u,v)=\Box(uv)-\Box uv-u\Box u$$

[1]Here $\|u[0]\|_{\dot H^a}=\|u(0)\|_{\dot H^a}+\|\partial_t u(0)\|_{\dot H^a}$

Thus, if $\Box u = \Box v = 0$, using Plancherel,

$$\begin{aligned} \|Q_0(u,v)\|_{L^2(\mathbb{R}^{n+1})} &\leq \frac{1}{2}\|\Box(uv)\|_{L^2(\mathbb{R}^{n+1})} = \frac{1}{2}(2\pi)^{-n}\|(\tau^2 - |\xi|^2)\widetilde{uv}\|_{L^2(\mathbb{R}^{n+1})} \\ &\lesssim \|D_+D_-(uv)\|_{L^2(\mathbb{R}^{n+1})} \end{aligned}$$

Therefore,

$$\|Q_0(u,v)\|_{L^2(\mathbb{R}^{n+1})} \leq \|D_+D_-(uv)\|_{L^2(\mathbb{R}^{n+1})} \tag{409}$$

Thus, in the case of the null form Q_0, theorem 3.2 reduces to,

$$\|D_+D_-(uv)\|_{L^2(\mathbb{R}^{n+1})} \lesssim \|u[0]\|_{\dot{H}^1(\mathbb{R}^n)}\|v[0]\|_{H^{\frac{n+1}{2}}(\mathbb{R}^n)} \tag{410}$$

which is a special case of theorem **??**.

Below we show that similar estimates hold true for the other null forms, Q_{ij}, Q_{0i}.

Remark 3.5. Given a solution u of $\Box u = 0$ with initial data $u(0,x) = f(x)$, $\partial_t u(0,x) = 0$ we denote by u' the solution of the same equation with data $u'(0,x) = f'(x)$, $\partial_t u'(0,x) = 0$ where $f' = \mathcal{F}^{-1}(|\hat{f}|)$. Observe, of course, that $\|f'\|_{\dot{H}^a} = \||f\|_{\dot{H}^a}$ and thus, from the point of view of the L^2 type estimates we are considering u and u' are indistinguishable.

Proposition 3.6. *Let u, v be smooth solutions of the homogeneous wave equation with initial . The following estimates hold true:*

$$\|Q_{ij}(u,v)\|_{L^2(\mathbb{R}^{n+1})} \lesssim \|D^{1/2}D_-^{1/2}(D^{1/2}u' \cdot D^{1/2}v')\|_{L^2(\mathbb{R}^{n+1})} \tag{411}$$

$$\|Q_{0i}(u,v)\|_{L^2(\mathbb{R}^{n+1})} \lesssim \|D_+^{1/2}D_-^{1/2}(D^{1/2}u' \cdot D^{1/2}v')\|_{L^2(\mathbb{R}^{n+1})} \tag{412}$$

Proof : We first decompose, as before, $u = u^+ + u^-$, $v = v^+ + v^-$ We write, in Fourier variables,

$$Q_{ij}\widetilde{(u^+, v^\pm)}(\tau,\xi) = \int q_{ij}(\eta, \xi-\eta)\delta(\tau - |\eta| \pm |\xi-\eta|)\hat{f}(\eta)\hat{g}(\xi-\eta)d\eta$$

where $q_{ij}(\eta, \xi-\eta) = \eta_i(\xi-\eta)_j - \eta_j(\xi-\eta)_i = (\xi\wedge\eta)_{ij}$ We now rely on the following simple lemma.

Lemma 3.7. *The following inequalities hold true,*

$$|\xi\wedge\eta| \lesssim |\xi|^{1/2}|\eta|^{1/2}|\xi+\eta|^{1/2}(|\xi| + |\eta| - |\xi+\eta|)^{1/2} \tag{413}$$

$$|\xi\wedge\eta| \lesssim |\xi|^{1/2}|\eta|^{1/2}|\xi+\eta|^{1/2}(|\xi+\eta| - \big||\xi| - |\eta|\big|)^{1/2} \tag{414}$$

We have indeed,

$$\begin{aligned} 4|\xi\wedge\eta|^2 &= 4(|\xi||\eta| - \xi\cdot\eta)(|\xi||\eta| + \xi\cdot\eta) \\ &= ((|\xi| + |\eta| - |\xi+\eta|)((|\xi| + |\eta| + |\xi+\eta|) \\ &\quad (|\xi+\eta| - \big||\xi| - |\eta|\big|)(|\xi+\eta| + \big||\xi| - |\eta|\big|) \end{aligned}$$

from which the lemma immediately follows.

Therefore, in both cases, $|\widetilde{Q_{ij}(u^+, v^\pm)}(\tau, \xi)|$ can be bounded by the expression,

$$\int |q_{ij}(\eta, \xi - \eta)| \; \delta(\tau - |\eta| \pm |\xi - \eta|) |\hat{f}(\eta)| \, |\hat{g}(\xi - \eta)| d\eta$$

$$\lesssim \big| |\tau| - |\xi| \big|^{1/2} |\xi|^{1/2} \int \delta(\tau - |\eta| \pm |\xi - \eta|) \, |\eta|^{1/2} |\xi - \eta|^{1/2} |\hat{f}(\eta)| \, |\hat{g}(\eta)| d\eta$$

$$= D^{1/2} D_-^{1/2} (D^{1/2} u' D^{1/2} v')$$

as desired.

■

According to proposition 3.6, theorem 3.2 reduces, for $Q = Q_{ij}$, resp. $Q = Q_{0i}$, to the statements,

$$\begin{aligned} \|D^{1/2} D_-^{1/2} (u \cdot v)\|_{L^2(\mathbb{R}^{n+1})} &\lesssim \|u[0]\|_{\dot{H}^{1/2}} \cdot \|u[0]\|_{\dot{H}^{n/2}} \\ \|D_+^{1/2} D_-^{1/2} (u \cdot v)\|_{L^2(\mathbb{R}^{n+1})} &\lesssim \|u[0]\|_{\dot{H}^{1/2}} \cdot \|u[0]\|_{\dot{H}^{n/2}} \end{aligned}$$

which are particular cases of theorem **??**.

APPENDIX A

A Review of Integration over Submanifolds

In this appendix, we provide a review of the integration of functions over lower-dimensional submanifolds of $\mathbb{R}^n$ (including the computation of length, surface area, etc.) and the relevant algebra. We will not discuss the integration of differential forms, even though the ideas are similar. In the last section, we prove the Coarea Formula, which allows one to compute the density of a measure pushed forward by a smooth map with surjective derivative – this operation is used in the lecture notes to define the pullback of a distribution. We also give a proof of the change of variables formula.

0.8. Algebraic Preliminaries and the Computation of Volumes.

Before discussing integration, we consider the related problem of determining the k-dimensional volume of a parallelogram $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k)$ in $\mathbb{R}^n$ with edges v_i and a vertex at the origin – in fact we will study an abstraction of this problem. We are forced to discuss this problem from an abstract point of view, speaking, for example, of "an" inner product space $W, g(\cdot,\cdot)$ with a subspace V, or of the dual space V^* of linear functionals from V to $\mathbb{R}$. This abstraction is necessary for our application to integration, where every linear object will locally approximate some nonlinear object, and we need the freedom to change variables or basis at whim. For example $W = W_x$ will approximate some kind of ambient three or four-dimensional space around the point x, and the subspace $V = V_x$ will approximate some surface therein which contains the point x. The dual space $V^* = V_x^*$ of linear functions $v_* : V \to \mathbb{R}$ will approximate the span of coordinate functions on that surface around the point x, and each of these spaces will have its own inner product to play the role of the Euclidean dot product in providing the geometric notions of angle, length, projection, perpendicularity, etc.

Now, consider a parallelogram $v^1 \wedge \ldots \wedge v_k$ with edges $v_1, \ldots, v_k \in \mathbb{R}^n$. To compute the number $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k)$ in general, it suffices to observe three geometrically intuitive facts:

The Axioms of k-dimensional Volume

- *Normalization:* $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k) = 1$ whenever $\{v_1, \ldots, v_k\}$ form an orthonormal set
- *Homogeneity:* Replacing any edge v_i by the vector αv_i multiplies the associated volume by $|\alpha|$

- *Parallel Shift Invariance:* $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k)$ remains unchanged if we shift one vertex by any amount within the span of the others, e.g.
$$|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k) = |\mathrm{Vol}| \left((v_1 + \sum_2^k c^i v_i) \wedge v_2 \wedge \ldots \wedge v_k \right)$$

In particular, we have $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k) = 0$ whenever $\{v_1, \ldots, v_k\}$ are linearly dependent. Combining these observations allows us to compute k-dimensional volumes of arbitrary parallelograms by making appropriate parallel shifts until the edges are in fact orthogonal. Hence, there is a unique such k-dimensional volume function satisfying the above properties for any inner product.

Exercise 0.9. From the above axioms, deduce the additional properties:

- $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k)$ remains invariant under arbitrary permutation of the edges
- If the span of $\{v_1, \ldots, v_l\}$ is orthogonal to the span of $\{v_{l+1}, \ldots, v_k\}$, then in fact $|\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k) = |\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_l) \cdot |\mathrm{Vol}|(v_{l+1} \wedge \ldots \wedge v_k)$

For the computation and analysis of angle and length, the linearity provided by an inner product is very helpful, and in the same way it is helpful to have a linear tool for the computation of volume. After some experimentation with parallelograms in the plane, the reader will be convinced that $|\mathrm{Vol}|$ is sort of almost a linear function in each edge. In order to take precise hold of this helpful linearity, we consider the special case $n = k$, and we find it necessary to introduce a *signed* volume function $\mathrm{Vol}(\cdot) : V^k \to \mathbb{R}$ for k-dimensional parallelograms. We will quickly run into the issue of orientation, which will force us to take seriously the order of the edges.

Definition 0.10. A *volume element* associated to a k-dimensional real vector space V and inner product $g(\cdot, \cdot)$ is a function $\mathrm{Vol}(v_1, \ldots, v_k) : V^k \to \mathbb{R}$ which satisfies the following properties:

- Vol is linear in each of its variables
- $|\mathrm{Vol}(v_1, \ldots, v_k)| = |\mathrm{Vol}|(v_1 \wedge \ldots \wedge v_k)$ is the unsigned volume of the parallelogram $v_1 \wedge \ldots \wedge v_k$

The second property implies that $\mathrm{Vol}(v_1, \ldots, v_k) = 0$ for all linearly dependent sets $\{v_1, \ldots, v_k\}$ – together with multilinearity, it then follows that Vol is also invariant under parallel shifts of the edges. Using these properties, it is clear that the choice of a value $\mathrm{Vol}(e_1, \ldots, e_k) = \pm 1$ for any particular orthonormal basis $\{e_j, j = 1, \ldots, k\}$ fully determines the function $\mathrm{Vol}(v_1, \ldots, v_k)$ [1]. Indeed, the value at any other point $(v_1, \ldots, v_k)$ can be obtained after a finite sequence of parallel shifts and scalar multiplications of entries.

Thus, while the unsigned, k-dimensional volume function $|\mathrm{Vol}|$ is uniquely determined by the inner product, there are two volume elements Vol which give rise to

[1] In this case, $\mathrm{Vol}(v_1, \ldots, v_k) = \pm \det(g(v_i, e_j))$ is essentially the determinant of the matrix of coefficients of the vectors $(v_1, \ldots, v_k)$ expressed relative to $\{e_j\}$ up to a sign.

the volume function in this way. Choosing one such volume element allows one to speak of whether an ordered basis $(e_1, \dots, e_k)$ is "positively" or "negatively" oriented depending on the sign of $\mathrm{Vol}(e_1, \dots, e_n)$. The issue of whether two bases are differently oriented boils down to whether one can be obtained (at least up to a positive multiple) from a sequence of parallel shifts of the other – for example the "left hand rule" and "right hand rule" determined by one's thumb, index and middle fingers may be compared to differently oriented bases of $\mathbb{R}^3$.

Exercise 0.11. Prove the uniqueness of Vol up to the choice of one of two "orientations".

The vector space of all k-linear functions $\omega(\cdot) : V^k \to \mathbb{R}$ with the property that $\omega(v_1, \dots, v_k) = 0$ whenever $\{v_1, \dots, v_k\}$ are linearly dependent is a one-dimensional vector space by similar considerations of parallel shifts. Any nonzero k-linear function ω on V^k with this property is also called a **volume form** on V, and every other volume form is a constant multiple thereof[2]. Such a function is a special case of a more general and similarly useful class of multilinear functions:

Definition 0.12. A multilinear function $\eta : V^t \to \mathbb{R}$ is called *alternating* if it satisfies any of the following equivalent properties:

- The value $\eta(v_1, \dots, v_t)$ remains invariant under parallel shifts of $v_1, \dots, v_t$
- $\eta(v_1, \dots, v_t) = 0$ if $v_i = v_j$ for some $i \neq j$
- A transposition of any two vectors in the input changes the sign of the output. For example,

$$\eta(v_1, v_2, v_3, \dots, v_t) = -\eta(v_3, v_2, v_1, \dots, v_t) = +\eta(v_3, v_t, v_1, \dots, v_{t-1}, v_2)$$

- $\eta(v_1, \dots, v_t) = 0$ whenever the vectors $\{v_i\}$ form a linearly dependent set

Exercise 0.13. Let W be an n-dimensional real vector space with basis $\{e_1, \dots, e_n\}$. We temporarily introduce the notation $\vec{j} = (j_1, j_2, \dots, j_k)$ for an increasing multi-index $1 \leq j_1 < j_2 < \dots < j_k \leq n$. Show that there is a basis for the space of alternating k-linear forms on W given by $\{\eta^{\vec{i}} \ : \ 1 \leq i_1 < i_2 < \dots i_k \leq n\}$ so that

$$\eta^{\vec{i}}(e_{j_1}, e_{j_2}, \dots, e_{j_k}) = 1 \cdot (\vec{i} = \vec{j}) + 0 \cdot (\vec{i} \neq \vec{j}).$$

In particular, the space of alternating k-linear forms has dimension $\binom{n}{k}$.

0.13.1. *Exterior Powers and the Determinant.* Whenever functions and other objects in mathematics possess some symmetry, it is very common to regard these objects as living on a new space which internally captures their symmetry. For example, a periodic function on the line can be regarded as a function on a circle, or a function in the plane which only depends on one of the variables can be regarded as a function on the line. The spaces which internalize the symmetry are constructed by quotienting out by the symmetry relation satisfied by the object of study just as the circle is a quotient of the line. This common theme in mathematics motivates another approach to the study of alternating, multilinear functions.

[2]While a volume element as defined earlier is a volume form, not every volume form on an inner product space is a volume element, and the notion of a volume form requires no inner product to make sense.

In this spirit, we can approach the theory of alternating k-linear functions by constructing a vector space $\Lambda^k(W)$ – called the k'th exterior power of W – which internalizes the symmetry of alternating, k-linear functions on W. This vector space is essentially the vector space generated by nondegenerate parallelograms called "k-vectors" $w_1 \wedge \ldots \wedge w_k$ of **ordered** edges $w_i \in W$. These k-vectors cannot truly be identified with parallelograms because they are subject to some equivalence relations which imply that various different $w_1 \wedge \ldots \wedge w_k$ and $w_1' \wedge \ldots \wedge w_k'$ may still represent the same element of $\Lambda^k(W)$. The following rules provide enough information to perform any computation in $\Lambda^k(W)$.

- The map $\pi : (w_1, \ldots, w_k) \to w_1 \wedge \ldots \wedge w_k \in \Lambda^k(W)$ is linear in each variable. For instance,

$$(u + 7v) \wedge x \wedge y = u \wedge x \wedge y + (7v) \wedge x \wedge y = u \wedge x \wedge y + 7 \cdot (v \wedge x \wedge y)$$

- $w_1 \wedge \ldots \wedge w_k = 0$ in $\Lambda^k(W)$ if and only if the $\{w_i\}$ are linearly dependent
- $w_1 \wedge \ldots \wedge w_i \wedge \ldots \wedge w_k$ represents the same element as $w_1 \wedge \ldots \wedge w_i' \wedge \ldots \wedge w_k$ if $w_i' = w_i + \sum_{j \neq i} c^j w_j$ can be obtained from w_i by a parallel shift in the span of the the remaining vertices, and more generally
- $w_1 \wedge \ldots \wedge w_k = w_1' \wedge \ldots \wedge w_k'$ are equivalent in $\Lambda^k(W)$ if and only if one parallelogram can be obtained from the other by a finite number of such parallel shifts. For example,

$$u \wedge v \wedge w = (u + w) \wedge v \wedge w = (u + w) \wedge v \wedge (-u) = w \wedge v \wedge (-u)$$

  all give different expressions for $-(w \wedge v \wedge u)$.

We see in particular that the order of the edges in $w_1 \wedge \ldots \wedge w_k$ is very important, since an odd permutation of the edges will negate the element in $\Lambda^k(W)$ [3].

Such a vector space exists, its dimension is $\binom{n}{k}$ whenever W has dimension n [4], and one can take as a basis the set of k-vectors $\{e_{\vec{i}} = e_{i_1} \wedge \ldots \wedge e_{i_k} : 1 \leq i_1 < \ldots < i_k \leq n\}$ whenever $\{e_i, i = 1, \ldots n\}$ form a basis for W.

The purpose of the vector space $\Lambda^k(W)$ is the following **universal property**, which completely characterizes $\Lambda^k(W)$.

Proposition 0.14. *For every alternating k-linear map* [5] $A : W^k \to V$ *there is a unique* **linear** *map* $\tilde{A} : \Lambda^k(W) \to V$ *so that* $\tilde{A} \circ \pi = A$.

Example Any linear map $T : W \to V$ induces a linear map $\Lambda^k T : \Lambda^k(W) \to \Lambda^k(V)$ so that $T(w_1 \wedge \ldots \wedge w_k) = (Tw_1) \wedge \ldots \wedge (Tw_k)$ since the map $(w_1, \ldots, w_k) \to (Tw_1) \wedge \ldots \wedge (Tw_k)$ from $W^k \to \Lambda^k(V)$ is alternating. As is common with many

[3] An odd permutation of the k letters $\{w_1, \ldots, w_k\}$ may be defined as a permutation of the edges $\{w_1, \ldots, w_k\}$ which negates $w_1 \wedge \ldots \wedge w_k$. In particular, a transposition, which interchanges exactly two vertices, is an odd permutation by the previous calculation. Even transpositions of k letters are those which preserve $w_1 \wedge \ldots \wedge w_k$.

[4] There are a few special cases to note: $\Lambda^0(W) = \mathbb{R}$, $\Lambda^1(W) = W$, and of course $\Lambda^k(W) = 0$ for all $k > n$

[5] "Alternating" means the same as it does for multilinear functions into $\mathbb{R}$, and the same alternative characterizations from the above fact apply.

objects in algebra, using the universal property of Λ^k is the primary means by which one constructs a well-defined linear map on the space.

The universal property described in Proposition (0.14) is exactly the usefulness of the exterior powers – it states that every alternating multilinear function can be viewed as a linear function on the appropriate vector space.

As an important application of Proposition (0.14), the dual space, $\Lambda^k(W)^*$, may be identified with the space of alternating k-linear functions by defining

$$\eta(w_1, \ldots, w_n) \equiv \eta(w_1 \wedge \ldots \wedge w_n), \quad \eta \in \Lambda^k(W)^*$$

This definition identifies the basis dual to $\{e_{\vec{i}} \in \Lambda^k(W)\}$ with the basis of alternating k-linear functions $\{\eta^{\vec{i}}\}$ in Exercise (0.13). At the same time, it is useful to identify $\Lambda^k(W^*) \simeq \Lambda^k(W)^*$ by the linear extension of the association $[\ \cdot\ ] : (W^*)^k \to \Lambda^k(W)^*$ defined by

$$\left[(v_*^1 \wedge \ldots \wedge v_*^k)\right](w_1, \ldots, w_k) \equiv \det\left(v_*^i(w_j)\right)_{i,j=1\ldots k}, \quad v_*^1 \wedge \ldots \wedge v_*^k \in \Lambda^k(W^*)$$

This extension is well-defined by the universal property of $\Lambda^k(W^*)$ since $[\ \cdot\ ]$ is alternating and k-linear. It is characterized by the following properties, which can be useful for computation:

- The map $(v_*^1, \ldots, v_*^k) \to \left[(v_*^1 \wedge \ldots v_*^k)\right]$ is an alternating, k-linear map $(W^*)^k \to \Lambda^k(W)^*$
- For any basis $\{e_j\ :\ 1 \le j \le n\}$ of W with dual basis [6] $\{e_*^i\ :\ 1 \le i \le n\}$ of W^*, and any k-multi-indices $\vec{i}, \vec{j}$, we have, in the notation of Exercise (0.13),
$$\left[e_*^{i_1} \wedge \ldots \wedge e_*^{i_k}\right](e_{\vec{j}}) = 1 \cdot (\vec{i} = \vec{j}) + 0 \cdot (\vec{i} \neq \vec{j})$$

It is also important (although obvious) to notice the property that for any linear map $T : W \to W$

$$[v_*^1 \wedge \ldots \wedge v_*^k](Tw_1, \ldots, Tw_k) = [v_*^1 \circ T \wedge \ldots \wedge v_*^k \circ T](w_1, \ldots, w_k)$$

Since the association $[\ \cdot\ ]$ really identifies elements of $\Lambda^k(W^*)$ with alternating, k-linear functions, there is no particular need to continue writing the brackets, and it makes perfect sense to only write

$$(v_*^1 \wedge \ldots \wedge v_*^k)(w_1, \ldots, w_k) \equiv \det\left(v_*^i(w_j)\right)_{i,j=1\ldots k}$$

We have yet to define here what the determinant of a $k \times k$ matrix is. In order to be more self-contained, let us give a construction based on exterior powers.

□

Exercise 0.15. The Determinant Recall that $\Lambda^n(W)$ is one-dimensional for W of dimension n (in other words, a finite sequence of scalar multiplications and parallel shifts can construct any basis from an initial basis). From this fact, we

[6] That is, the e_*^i are the linear maps from $W \to \mathbb{R}$ so that $e_*^i(e_j) = \delta_{ij}$ – hence $e_*^i(v)$ gives the "i'th" coordinate of v relative to the basis $\{e_i\}$

associate to every linear map $T : W \to W$ a scalar (called the "determinant" of T), which satisfies the equality

$$T_{\#}\alpha = (\det T)\alpha, \quad \alpha \in \Lambda^n(W)$$

Equivalently, $\det T$ is the trace of $\Lambda^n T$. Prove that $\det T = 0$ if and only if T fails to be a bijection. Also note that $\det(T_1 \circ T_2) = \det T_1 \cdot \det T_2$. Try to use this definition and the rules for computation in $\Lambda^n(W)$ to compute the determinant of an explicit matrix with entries of your choice. If we define the transpose $T^t : W^* \to W^*$ by $w_* \to w_* \circ T$, show that $\det T^t = \det T$.

What are the geometric meanings of column operations to compute a determinant in this context? Row operations?

The only serious difference between these approaches is that a linear map T between two vector spaces pushes forward parallelograms by linear extension of the rule $w_1 \wedge \ldots \wedge w_k \to T_{\#}(w_1 \wedge \ldots \wedge w_k) \equiv (Tw_1) \wedge \ldots \wedge (Tw_k)$, whereas the same linear map induces a pull-back of alternating k-linear functions $T^*\eta(w_1, \ldots, w_k) = \eta(Tw_1, \ldots Tw_k)$.

0.15.1. *Volumes, Angles, and the Induced Inner Product.* For the computation of volumes, we will be interested in the case where W possesses an inner product $g(\cdot, \cdot)$. The space $\Lambda^k(W)$ inherits its own inner product induced by g, which is given by bilinear extension of the explicit formula

$$g(v_1 \wedge \ldots \wedge v_k, w_1 \wedge \ldots \wedge w_k) = \det\left(g(v_i, w_j)\right), \quad i, j = 1, \ldots, k$$

The existence of such an induced inner product confirms the intuition that oriented k-dimensional parallelograms also possess a notion of size (the k-dimensional volume) and angle (e.g. the angle between two planes), both of which are invariant under parallel shifts of the vertices. In fact, the induced norm

$$|v_1 \wedge \ldots \wedge v_k| \equiv \left[\det\left(g(v_i, v_j)\right)\right]^{1/2} \tag{415}$$

of a k-vector turns out to be equal to the k-dimensional volume of the associated parallelogram (either simply by definition or because the norm apparently satisfies the axioms of a k-dimensional volume function described at the beginning of this appendix), and the angles between different parallelograms u, v (or equivalently the oriented subspaces spanned thereby) may be computed through the usual rule

$$\cos\theta = g(u, v)/(\|u\| \cdot \|v\|)$$

for computing angles between vectors in an inner product space. For example, when two planes intersect in a line spanned by a vector e of unit norm $g(e, e) = 1$, we can see the correct angle between the planes is computed as follows: if $u = e \wedge u_1$ is an oriented 2-parallelogram with $e \perp u_1$, and $v = e \wedge v_1$ is an oriented 2-parallelogram with $e \perp v_1$, then $g(u, v) = g(u_1, v_1)$ and the angle between the planes described by u and v is equal to the angle between u_1 and v_1 as desired.

With the induced inner product, an orthonormal basis $\{e_1, \ldots, e_n\}$ of W clearly gives rise to an orthonormal basis $\{e_{\vec{i}}\}$ for $\Lambda^k(W)$ as before. Of course, the space $\Lambda^k(W)^*$ of alternating k-linear maps therefore inherits an inner product by duality, and for this inner product the basis $\{\eta^{\vec{i}}(\cdot)\}$ dual to $\{e_{\vec{i}}\}$ becomes an orthonormal basis. Combining all these observations allows one to take advantage of multilinearity (and even apply changes of variable) to compute volumes in several ways.

Exercise 0.16. Compute the area of the 2-parallelogram $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \wedge \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}$ in three different ways:

- Use the Gram-Schmidt process and the axioms of two-dimensional volume
- Use the formula for the norm of a 2-parallelogram in $\Lambda^2(\mathbb{R}^3)$
- Decompose relative to the orthonormal basis $$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Exercise 0.17. Compute the angle between the oriented plane spanned by $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \end{pmatrix} \right\}$ and the oriented plane spanned by $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\}$.

Note that these planes intersect in a line and therefore are contained within a three-dimensional subspace of $\mathbb{R}^4$.

Exercise 0.18. Let V be a 3-dimensional vector space with inner product $g(\cdot, \cdot)$ and a chosen orientation, and let $\mathrm{Vol}(\cdot, \cdot, \cdot)$ be the associated volume element. Define an antisymmetric, bilinear product $(u, v) \to u \times v \in V$ by the identity

$$g(u \times v, w) = \mathrm{Vol}(u, v, w)$$

Show that $u \times v$ is orthogonal to both u and v, its length is exactly the two-dimensional area of $u \wedge v$ and that the ordered trio $(u, v, u \times v)$ determines a positively oriented basis for V unless u and v are linearly dependent.

Exercise 0.19. For the induced inner product on $\Lambda^2(V)$ to be nondegenerate, we need $||u \wedge v||^2 \geq 0$ with equality if and only if u and v are linearly dependent – geometrically, the area of a nondegenerate parallelogram is positive. Show that the above inequality is actually the Cauchy-Schwartz inequality for the inner product on V. (One can use this idea to motivate a standard proof of the Cauchy-Schwartz inequality. The Cauchy-Schwartz inequality is also equivalent to the triangle inequality, and the geometric relation between these two interpretations may be demonstrated by forming a parallelogram to illustrate the identity $u + v = v + u$.)

Exercise 0.20. Suppose A is an $m \times n$ matrix and B is an $n \times m$ matrix with $m \leq n$. For any m-index $\vec{i}$ $(1 \leq i_1 < \ldots < i_m \leq n)$, we let $A^{\vec{i}}$ denote the $m \times m$ matrix $A^{\vec{i}}_{jr} = A_{ji_r}$ for the matrix composed of the columns of A corresponding to $\vec{i}$. Similarly, we write $B^{\vec{i}}$ for the $m \times m$ matrix $B^{\vec{i}}_{jr} = B_{i_j r}$ whose rows are those rows of B that have indices from $\vec{i}$. Prove the Cauchy-Binet formula:

$$\det(AB) = \sum_{\vec{i}} \det A^{\vec{i}} \cdot \det B^{\vec{i}}$$

(**Hint**: Decompose the identity operator on $\Lambda^m(\mathbb{R}^n)$ as a sum of projection operators. Use the universal property of Λ^m to define the appropriate maps.)

Exercise 0.21. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map, $\Lambda^k A : \Lambda^k(\mathbb{R}^n) \to \Lambda^k(\mathbb{R}^n)$ the induced map. Define the characteristic polynomial of A to be the degree n polynomial

$$\chi(x) = \det(xI - A)$$

Prove the formula

$$\frac{d}{dx}\mathrm{tr}\Lambda^k(xI - A) = (n - (k-1))\mathrm{tr}\Lambda^{k-1}(xI - A)$$

by using the multilinearity of the wedge product. Interpret this calculation geometrically by considering difference quotients (Why does differentiating decrease k by 1? What does the factor $(n-k+1)$ count?). By corollary, deduce the following formula for the coefficients of the characteristic polynomial:

$$\chi(x) = \sum_{k=0}^{n} \mathrm{tr}\Lambda^{n-k}(-A) \cdot x^k$$

□

Now let us see how all this algebra interplays as we attempt to patch local (or infinitesimal) notions of volume to global ones and define integration on submanifolds of $\mathbb{R}^n$. The reader unfamiliar with the notion of a submanifold should have in mind the set of solutions to an equation $f_1(x) = f_2(x) = \ldots = f_m(x) = 0$ in $\mathbb{R}^n$ in the generic situation to which the implicit function theorem applies (for instance, the transverse intersection of the cylinder $x^2+y^2 = 1$ with the sphere $x^2+y^2+z^2 = 2^2$ in $\mathbb{R}^3$).

1. Integration on Submanifolds of $\mathbb{R}^n$

For each point $x \in \mathbb{R}^n$, we define the n-dimensional vector space $T_x(\mathbb{R}^n)$ of all velocities $(w_1, \ldots, w_n)$ which can be realized at the point x by a smooth path in the ambient space $\gamma : \mathbb{R} \to \mathbb{R}^n$ [7]. The space $T_x(\mathbb{R}^n)$ is called the tangent space to $\mathbb{R}^n$ at the point x, and a typical element of $T_x(\mathbb{R}^n)$ may be written $w = w_1\frac{\partial}{\partial x^1} + \ldots + w_n\frac{\partial}{\partial x^n}$. We choose this notation not only to distinguish the tangent space from the elements of $\mathbb{R}^n$ themselves, but also because it suggests another way to think of tangent vectors as differential operators on smooth functions.

For $K \subseteq \mathbb{R}^n$, one similarly defines $T_x(K)$ (the tangent space to K at x) as the set of velocities which can be realized at the point x by a smooth path $\gamma : \mathbb{R} \to K$ **within** K. This definition makes sense for arbitrary subsets $K \subseteq \mathbb{R}^n$; at the moment, we are only interested in K for which $T_x(K)$ is in fact a k-dimensional subspace of $T_x(\mathbb{R}^n)$ for all points $x \in K$, and for which the tangent space $T_x(K)$ varies "smoothly" along K in some sense. At any rate, in these notes, K usually arises as the level set of a smooth map $f : \mathbb{R}^n \to \mathbb{R}^{n-k}$ whose derivative $Df(x)$ at every point is a surjective linear map $Df : T_x(\mathbb{R}^n) \to T_{f(x)}(\mathbb{R}^{n-k})$.

[7] It makes no difference here, whether one replaces $\mathbb{R}$ by any open subinterval thereof.

Exercise 1.1. Suppose $f(x_0) = 0$. Define the derivative as a linear map between $Df(x_0) : T_{x_0}(\mathbb{R}^n) \to T_0(\mathbb{R}^{n-k})$ and use the implicit function theorem to show that the tangent space to $K = f^{-1}(0)$ at the point x_0 is exactly the null space of $Df(x_0)$ under the assumption that $Df(x_0)$ is surjective. In this way, the vectors tangent to the level sets of f are exactly those directions along which the value of f remains basically unchanged. Compare to the case $\{xy = 0\}$ in $\mathbb{R}^2$.

The vector space $T_x(K)$ inherits an inner product $g(\cdot,\cdot)$ from the ambient space $W = T_x(\mathbb{R}^n)$ by restriction, and this "local" notion of geometry patches together to allow one to discuss more global geometric quantities on K. For example, since the inner product allows one to measure the "length" of a vector $|v| = \sqrt{(}g(v,v))$, one may define the "speed" of a trajectory with velocity $v = \dot{\gamma}$ at a point x and thereby approximate the length of a small portion of the curve near the point x by $|\text{Vol}|(\gamma([t, t+\Delta t])) \sim |\dot{\gamma}|(t)|\Delta t|$. But then one can extend the local approximation to a global notion of length for a piecewise-smoothly parameterized curve $\gamma : [a,b] \hookrightarrow K$ by the formula

$$L(\gamma) = \int_a^b |\dot{\gamma}(t)|\ dt$$

One then uses the homogeneity of the integrand to show that this notion of length is in some way independent of a change in parameterization where we replace $\gamma(t) \to \gamma(\Phi(t))$ with $\Phi(\cdot)$ a continuously differentiable bijection of intervals and $\Phi'(\cdot) \neq 0$. In fact, what one determines in this way is an arclength measure on the image of γ with respect to which a continuous function (say) may be integrated by the formula

$$\int_\gamma u\ dl = \int_a^b u(\gamma(t))|\dot{\gamma}(t)|\ dt$$

Example. Consider a path $\gamma(t)$ in the (x,y) plane, but expressed in the polar coordinates $r(t), \theta(t)$. The velocity of γ is given by $\dot{r}(t)\partial_r + \dot{\theta}(t)\partial_\theta$. It is easy to see that the vectors ∂_r and ∂_θ are orthogonal and have respective lengths 1 and r at the point $(r\cos\theta, r\sin\theta)$ in the (x,y) plane. Therefore, the speed of γ satisfies $|\dot{\gamma}|^2(t) = \dot{r}^2 + (r\dot{\theta})^2$ by the Pythagorean theorem, and so

$$L(\gamma) = \int_a^b \sqrt{\dot{r}^2 + (r\dot{\theta})^2}\ dt$$

Exercise 1.2. Verify rigorously that the above calculation is correct and accords with our definitions. (Note that γ is really the composition of a path in the (r,θ) plane and the polar coordinate map.)

In a similar way, if one can parameterize an l-dimensional subset $L \subseteq K$ by a bijective map $\Gamma : \mathcal{R} \to L$ on some l-dimensional, open region $\mathcal{R} \subseteq \mathbb{R}^l$, one can define a parameterization-independent integral so that

$$\int_L u\ dS_L = \int_{\mathcal{R}^l} u(\Gamma(t))\rho_\Gamma(t)\ dt^1 \dots dt^l$$

for all (decent) functions on L whose support lies within the image of Γ. The density factor $\rho_\Gamma(t)$ measures the factor by which Γ shrinks or enlarges the volumes of small subsets near the point t. In the particular case where $u(p) = 1\cdot(p \in \Omega)+0\cdot(p \notin \Omega)$ is the characteristic function of a subregion Ω of the image of Γ, the integral calculates the l-dimensional volume of Ω. The density factor $\rho_\Gamma(t)$ must be present (otherwise every subset of L would appear to have the same volume). More precisely, to measure the distortion of volume by Γ near a point $t = (t_1, \dots t_l) \in \mathcal{R}$, we would like to have an approximation

$$\begin{aligned}|\mathrm{Vol}|_L \Gamma([t_1, t_1 + \Delta t_1] \times \ldots \times [t_l, t_l + \Delta t_l])\\ \sim \rho_\Gamma(t)|\mathrm{Vol}|_{\mathcal{R}}[t_1, t_1 + \Delta t_1] \times \ldots \times [t_l, t_l + \Delta t_l]\end{aligned}$$

for small rectangles $[t_1, t_1 + \Delta t_1] \times \ldots \times [t_l, t_l + \Delta t_l]$, and also for arbitrary small parallelograms at t whose vertices may not be perpendicular. Therefore it makes sense to require that Γ be continuously differentiable (or at least possess some amount of regularity), and to define

$$\int_L u \; dS_L = \int_{\mathcal{R}} u(\Gamma(t)) \; |D\Gamma(\partial_{t^1} \wedge \ldots \wedge \partial_{t^l})|(t) \; dt^1 \ldots dt^l \tag{416}$$

where the density factor is the l-dimensional volume of the parallelogram

$$D\Gamma(\partial_{t^1} \wedge \ldots \wedge \partial_{t^l}) \equiv (D\Gamma(\partial_{t^1}) \wedge \ldots \wedge D\Gamma(\partial_{t^l})) \in \Lambda^l(T_{\Gamma(t)}(K))$$

We have already discussed several ways to compute this quantity in the abstract setting (for example, by the Gram-Schmidt process), but let us unravel our definitions and write down a more explicit, general description.

By definition, $D\Gamma(\partial_{t^i}) = \sum_j \frac{\partial \Gamma^j}{\partial t^i} \frac{\partial}{\partial x^j}$, and applying the formula (415), we use the Euclidean inner product $g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) = \delta_{ij}$ on $T_{\Gamma(t)}(K)$ to form an $l \times l$ matrix, and compute its determinant:

$$\begin{aligned}|D\Gamma(\partial_{t^1} \wedge \ldots \wedge \partial_{t^l})|^2(t) &= \det\left(g(\frac{\partial \Gamma^j}{\partial t^i}(t)\frac{\partial}{\partial x^j}, \frac{\partial \Gamma^r}{\partial t^m}(t)\frac{\partial}{\partial x^r})\right)_{i,m=1\ldots l}\\ &= \det\left(\delta_{jr}\frac{\partial \Gamma^j}{\partial t^i}\frac{\partial \Gamma^r}{\partial t^m}\right)_{i,m=1\ldots l}\end{aligned}$$

In the above formula we have employed the tremendously useful *Einstein summation convention* – namely, we understand a summation over the indices j and r because they are repeated – in order to condense expressions such as

$$\delta_{jr}\frac{\partial \Gamma^j}{\partial t^i}\frac{\partial \Gamma^r}{\partial t^m} = \sum_{j,r=1}^{n} \delta_{jr}\frac{\partial \Gamma^j}{\partial t^i}\frac{\partial \Gamma^r}{\partial t^m} = \sum_{j=1}^{n} \frac{\partial \Gamma^j}{\partial t^i}\frac{\partial \Gamma^j}{\partial t^m}$$

and $D\Gamma(\partial_{t^i}) = \sum_j \frac{\partial \Gamma^j}{\partial t^i}\frac{\partial}{\partial x^j} = \frac{\partial \Gamma^j}{\partial t^i}\frac{\partial}{\partial x^j}$. We will continue to use this notation from this point onward.

In some ways, our discussion thus far is not sufficiently general for applications. For example, if one makes a coordinate change $(x^1, \dots, x^n) = \Phi(y^1, \dots, y^n) =$

$(\Phi^1(y), \ldots, \Phi^n(y))$ for some diffeomorphism Φ of $\mathbb{R}^n$, one runs into the complication that the inner product on $T_y(\mathbb{R}^n)$ is in general no longer diagonal with respect to $\{\frac{\partial}{\partial y^i}\}$ and even appears to depend on the point y according to the transformation rule

$$g_{ij}(y) \equiv g(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}) = \frac{\partial \Phi^l}{\partial y^i} \frac{\partial \Phi^r}{\partial y^j} g(\frac{\partial}{\partial x^l}, \frac{\partial}{\partial x^r}) = \delta_{lr} \frac{\partial \Phi^l}{\partial y^i}(y) \frac{\partial \Phi^r}{\partial y^j}(y)$$

The matrix $g_{ij}(y)$ above is always positive definite and symmetric.

An inner product $g_{ij}(y)$ on $T_y(\mathbb{R}^n)$ which varies smoothly in an open set $\Omega \subseteq \mathbb{R}^n$ (in the sense that the matrix $g_{ij}(y)$ varies smoothly in y) is called a **Riemannian metric** and the generalization of this concept to manifolds is the most basic structure in the study of Riemannian geometry. It is this structure which allows us to extend the local notions of length and higher dimensional volume determined by an inner product to global notions such as the length of a curve, the area of a surface, and so on.

Were this more general setting considered with a Riemannian metric $g_{ij}(y)$ on $T_y(K)$, we would have arrived instead at the more general formula

$$|D\Gamma(\partial_{t^1} \wedge \ldots \wedge \partial_{t^l})|(t) = \sqrt{\det \left(g_{jr}(\Gamma(t)) \frac{\partial \Gamma^j}{\partial t^i} \frac{\partial \Gamma^r}{\partial t^m} \right)_{i,m=1\ldots l}}$$

We still need to verify that the value of the integral defined here would not change if one were to choose a different parameterization $\tilde{\Gamma} : \tilde{\mathcal{R}} \to L$. But this fact is a simple corollary of the change of variables formula proven at the end of this appendix, applied to the new parameterization $\tilde{\Gamma} = \Gamma \circ \Phi$, as well as the basic properties of exterior powers:

$$\begin{aligned}
&\int_{\mathcal{R}} u(\Gamma(t)) \; |D\Gamma(\partial_{t^1} \wedge \ldots \wedge \partial_{t^l})|(t) \; dt^1 \ldots dt^l \\
&\qquad = \int_{\tilde{\mathcal{R}}} u(\Gamma(\Phi(y))) \; |D\Gamma(\partial_{t^1} \wedge \ldots \wedge \partial_{t^l})|(\Phi(y)) \;\cdot\; |\det D\Phi| dy^1 \ldots dy^l \\
&\qquad = \int_{\tilde{\mathcal{R}}} u(\Gamma(\Phi(y))) \; |D\Gamma \circ \Phi(\partial_{y^1} \wedge \ldots \wedge \partial_{y^l})|(y) dy^1 \ldots dy^l
\end{aligned}$$

Finally, we cannot always find a single parameterization Γ which encompasses the entirety of the submanifold L of interest. For example, if $L = \{x \in \mathbb{R}^l : f(x) = 0\}$ arises as the solution set for a system of equations given by a smooth function $f : \mathbb{R}^n \to \mathbb{R}^{n'}$, then the implicit function theorem may guarantee that we can parameterize these level sets locally, but a single parameterization which covers almost all of L may be impossible to come by. It is therefore important to be able to decompose the submanifold L into parameterized pieces and integrate on each piece separately. Fortunately this goal can be accomplished by decomposing the function to be integrated in an almost arbitrary manner into localized pieces.

Definition 1.3. Let $L = \cup_\alpha \Gamma_\alpha(U_\alpha)$ be a finite union where U_α are open subsets of $\mathbb{R}^l$ and the $\Gamma_\alpha : U_\alpha \to L$ are smooth parameterizations. Assume also that the transition maps between pieces $\Gamma_\alpha^{-1} \circ \Gamma_\beta : U_\beta \cap \Gamma_\beta^{-1}(\Gamma(U_\alpha)) \to U_\alpha$ are smooth maps between open subsets of $\mathbb{R}^l$. For a continuous function u on L, we can decompose $u = \sum_\alpha u_\alpha$ on L, where each u_α is supported on the parameterized submanifold $\Gamma(U_\alpha)$. We then define

$$\int_L u \, dS_L = \sum_\alpha \int_{\Gamma_\alpha(U_\alpha)} u_\alpha \, dS_L$$

as long as each term on the right hand side is well-defined.

The condition on the transition maps ensures that the images of Γ_α overlap in a nice way, rather than intersecting in some transverse or irregular fashion – for a parameterized surface in $\mathbb{R}^3$, one could picture two pieces of tape which are made to overlap nicely. The strong regularity assumption of smoothness is here only to simplify language, but is not necessary. This definition would of course be useless were it not for the following

Theorem 1.4. *The value $\int_L u \, dS_L$ is well-defined independent of the decompositions of L and u.*

One of the main applications of this theory to the lecture notes has been to define the pullback of a distribution as the adjoint to the pushforward of a smooth, compactly supported measure $\phi \, dx$ by a smooth map f. Let us discuss the observations relevant to this computation.

Example.

Take $L = \{x^2 + y^2 < 1\}$ to be a disk in $\mathbb{R}^2$, and let $\Gamma : (0,1) \times (0, 2\pi) \to L$ be the map $\Gamma(r, \theta) = (r\cos\theta, r\sin\theta)$. The easiest way to compute the density factor (which generalizes to higher dimensions) is to notice the geometrically obvious fact that $D\Gamma(\partial_r)$ and $D\Gamma(\partial_\theta)$ are perpendicular, with respective lengths 1 and r. Therefore we have

$$||D\Gamma\partial_r \wedge \partial_\theta|| = ||D\Gamma\partial_r|| \cdot ||D\Gamma\partial_\theta|| = 1 \cdot r = \rho_\gamma$$

A more general way to obtain the same answer is to use the basic rules of calculation for Λ^2 as follows:

$$\begin{aligned} D\Gamma(\partial_r \wedge \partial_\theta) &= (\cos\theta\partial_x + \sin\theta\partial_y) \wedge (-r\sin\theta\partial_x + r\cos\theta\partial_y) \\ &= (\cos\theta\partial_x) \wedge (r\cos\theta\partial_y) + (\sin\theta\partial_y) \wedge (-r\sin\theta\partial_x) \\ &= (\cos\theta\partial_x) \wedge (r\cos\theta\partial_y) - (-r\sin\theta\partial_x) \wedge (\sin\theta\partial_y) \\ &= r\partial_x \wedge \partial_y \in \Lambda^2(T_{\Gamma(r,\theta)}(L)) \end{aligned}$$

which has norm r because $\{\partial_x, \partial_y\}$ form an orthonormal basis for $T_{\Gamma(r,\theta)}(L)$ under the Euclidean inner product.

Either way, we obtain a formula for integration in polar coordinates in the disk:

$$\int_L u \, dS_L = \int_0^{2\pi} \int_0^1 u(\Gamma(r,\theta)) r dr d\theta$$

There is another method, dual to the previous calculation, which is often used to perform the same sort of computation. We have already mentioned the fact that, at any point $p \in \mathbb{R}^n$, the tangent space to $\mathbb{R}^n$ at p consists of exactly the span of the vectors $\{\frac{\partial}{\partial x^i} : i = 1, \ldots, n\}$. A vector $c^j(p)\frac{\partial}{\partial x^j} \in T_p(\mathbb{R}^n)$ may be applied to any real-valued function f which is smooth and defined in a neighborhood of p. The linear functional on $T_p(\mathbb{R}^n)$ so obtained, $c^j(p)\frac{\partial}{\partial x^j} \to c^j(p)\frac{\partial f}{\partial x^j}$ is sometimes denoted $df_p \in T_p(\mathbb{R}^n)^*$, or simply by df, and is referred to as the "differential of f". If we consider the coordinates $x^1, \ldots, x^n$ themselves as smooth functions on $\mathbb{R}^n$, we find that the linear maps $dx^1, \ldots, dx^n$ form the basis for $T_p(\mathbb{R}^n)^*$ which is dual to the basis $\frac{\partial}{\partial x^i}$ – in other words, $dx^i(\frac{\partial}{\partial x^j}) = \frac{\partial x^i}{\partial x^j} = \delta_{ij}$. Any element of $T_p(\mathbb{R}^n)^*$ may be written as a linear combination of these elements, in particular $df = \frac{\partial f}{\partial x^i}(p)dx^i$ – the dual space itself is called the **cotangent space** to $\mathbb{R}^n$ at the point p.

To repeat the earlier computation, one can proceed as follows: writing $x = r\cos\theta$ and $y = r\sin\theta$ as functions on the rectangle $\mathcal{R} = (0,1) \times (0, 2\pi)$, one may consider the volume element $dx \wedge dy$ which is dual to $\partial_x \wedge \partial_y$, and compute as follows:

$$\begin{aligned}
|dx \wedge dy| &= |(\cos\theta dr - r\sin\theta d\theta) \wedge (\sin\theta dr + r\cos\theta d\theta)| \\
&= r|dr \wedge d\theta|
\end{aligned}$$

We obtain the same density factor r, although there is no orthogonality to exploit using this method since the inverse image of an infinitesimal rectangle by Γ does not necessarily have edges which are orthogonal.

We must justify that such a calculation results in the density factor ρ_Γ we originally defined, since our original definition for computing the density required us to compute the norm of some 2-parallelogram in the one-dimensional vector space $\Lambda^2(T_{(x,y)}(L))$ on the image of the polar coordinate map $\Gamma : \mathcal{R} \to L$. To be more formal, the volume element $dx \wedge dy$ belongs to the dual space $\Lambda^2(T_{(x,y)}(L)^*)$ of alternating bilinear forms and therefore the object we just computed (using the chain rule) $dx \circ D\Gamma \wedge dy \circ D\Gamma = d(x \circ \Gamma) \wedge d(y \circ \Gamma)$ is an element of the one-dimensional vector space $\Lambda^2(T_{(r,\theta)}(\mathcal{R})^*)$. Hence, $d(x \circ \Gamma) \wedge d(y \circ \Gamma) = c_\Gamma dr \wedge d\theta$ for some scalar $c_\Gamma = c_\Gamma(r, \theta)$. The number c_Γ is exactly what was computed above, and the question in general is why $|c_\Gamma| = \rho_\Gamma$. This equality follows more or less from the definition of ρ_Γ, the fact that $dx \wedge dy$ is a volume element, the definition of the wedge product and the chain rule as follows:

By definition, $\rho_\Gamma \equiv ||D\Gamma(\partial_r \wedge \partial_\theta)||_{\Lambda^2(T_{\Gamma(r,\theta)}(L)}$ at each fixed point (r, θ).

But $D\Gamma(\partial_r \wedge \partial_\theta)$ is of the form $t_\Gamma \partial_x \wedge \partial_y$ for some constant t_Γ because the the tangent space to L is two-dimensional and hence has a one-dimensional second exterior power. Since $\partial_x \wedge \partial_y$ has unit norm, we have $\rho_\Gamma = |t_\Gamma|$. But then from the relation $dx \wedge dy(\partial_x \wedge \partial_y) = 1$ we may also calculate.

$$\begin{aligned}|t_\Gamma| &= |dx \wedge dy(D\Gamma\partial_r, D\Gamma\partial_\theta)| \\ &= |dx \circ D\Gamma \wedge dy \circ D\Gamma(\partial_r, \partial_\theta)| \\ &= |d(x \circ \Gamma) \wedge d(y \circ \Gamma)(\partial_r, \partial_\theta)| \\ &= |c_\Gamma dr \wedge d\theta(\partial_r, \partial_\theta)| \\ &= |c_\Gamma|\end{aligned}$$

It should be easy to generalize our previous methods of calculation to the lifted disk $L' \subseteq \mathbb{R}^3$ given by $L' = \{(x, y, 57) \ : \ x^2 + y^2 < 1\}$. Indeed, it is clear that no modification or additional concept is necessary to apply the first technique of parameterization by $\Gamma(r, \theta) = (r\cos\theta, r\sin\theta, 57)$ and computing the density factor $||D\Gamma(\partial_r, \partial_\theta)||$. To extend the second method of computation, however, requires another concept since we must compute the volume element generalizing $dx \wedge dy$ on $L' \subseteq \{z = 57\}$. In general, when we encounter sets $K = \{x \ : \ f^1(x) = \ldots = f^m(x) = 0\} \subseteq \mathbb{R}^n$ which are defined as the set of solutions to a system of equations, we should be able to compute the appropriate volume element generalizing $dx \wedge dy$ in terms of defining functions.

We therefore seek the object analogous to $dx \wedge dy$ to associate to a surface (or submanifold) $S \subseteq \mathbb{R}^n$, which may be pulled back by a parameterization in order to compute integrals on S in the same way that one computes the norm of $dx \wedge dy = r dr \wedge d\theta$ in order to integrate in polar coordinates. This example evidently associates to each tangent space $T_p(\mathbb{R}^3)$ an alternating, bilinear function $r dr \wedge d\theta$ which varies smoothly in the point p [8]. Such an object is called a **two-form**, and the general two-form on $\mathbb{R}^3$ may be written as $P dx \wedge dy + Q dy \wedge dz + R dz \wedge dx$ for some (smooth) functions P, Q, R on $\mathbb{R}^3$. Concretely, we would like to write down and solve equations for (P, Q, R) which determine a volume element on the surface $\{f(x, y, z) = 0\}$, however we will need to develop some more algebraic concepts for this task.

2. The Volume Element on $\{f^1(x) = \ldots = f^m(x) = 0\}$

Our goal in this section is to construct the analog of the volume element $dx^1 \wedge \ldots \wedge dx^k$ on the k-dimensional submanifold $x^{k+1} = \ldots = x^n = 0$ in $\mathbb{R}^n$ for more general embedded submanifolds which are given as solutions to equations.

Let us consider a nondegenerate solution $x_0 \in \mathbb{R}^n$ to the system of equations $K = \{x \ : \ f^1(x) = \ldots = f^m(x) = 0\}$ for smooth functions f^i. By nondegenerate, we mean any of the following equivalent conditions:

- The derivative of the map $x \to (f^1(x), \ldots, f^m(x))$ is a surjective linear map $df : T_{x_0}(\mathbb{R}^n) \to \mathbb{R}^m$
- The null space of df has dimension $k = n - m$ in $T_{x_0}(\mathbb{R}^n)$

[8] although the (x, y) coordinates give the illusion that $dx \wedge dy$ remains "constant" as p varies, we see from the polar expression that this phenomenon is only an artifact of coordinates

- The alternating m-linear map $df^1 \wedge \ldots \wedge df^m \in \Lambda^m(T_{x_0}(\mathbb{R}^n)^*)$ is nonzero

or, to put it geometrically, the hypersurfaces $\{f^1(x) = 0\}, \ldots, \{f^m(x) = 0\}$ intersect transversely at the point x_0. In other words, the situation is generic, so the implicit function theorem applies. We have already mentioned that the tangent space to K at x_0, defined as the set of velocities realized at x_0 by some trajectory in K, consists exactly of those directions which leave the f^i unchanged; succinctly $T_{x_0}(N) = \bigcap_{i=1}^n \ker df^i(x_0)$ where $df^i : T_{x_0}(\mathbb{R}^n) \to \mathbb{R}$ is the derivative of f^i at x_0.

As a subspace of $T_{x_0}(\mathbb{R}^n)$, $T_{x_0}(K)$ inherits an inner product by restriction, and therefore possesses its own well-defined notion of k-dimensional volume – a function $|dS|(v_1, \ldots, v_k)$, which may be represented by some alternating k-linear function η in the sense that $|dS|(v_1, \ldots, v_k) = |\eta(v_1, \ldots, v_k)|$ for vectors v_i tangent to K at x_0. Our main goal in this section is to express [9] η or $|dS|$ in terms of $df^1, \ldots, df^m$ in a computable fashion.

In our motivating example of the $z = 57$ plane in $\mathbb{R}^3$, however, we observe that the function $\eta = dx \wedge dy$ is defined as an alternating bilinear form on the whole ambient tangent space to $\mathbb{R}^3$. However, $dx \wedge dy$ is certainly not the only alternating bilinear form on $T_{x_0}(\mathbb{R}^3)$ which restricts to a volume element on $T_{x_0}(z = 57) = \ker dz$ – any two-form of the type $dx \wedge dy + Qdy \wedge dz + Rdz \wedge dx$ obtains the same restriction for arbitrary Q and R to $\{z - 57 = 0\}$. An attribute which uniquely distinguishes $dx \wedge dy$ from these other forms is that $[dx \wedge dy](v_1, v_2) = 0$ if either v_1 or v_2 is normal to $z = 57$; equivalently, $dx \wedge dy$ remains unchanged when pulled back by orthogonal projection onto $\ker dz$.

With these examples in mind, we state a linear algebraic formulation of our problem to be solved on each tangent space.

Problem Version I: Suppose we are given an ambient, inner product space $W, g(\cdot, \cdot)$ of dimension n, and a set of linear functionals $f^1, \ldots, f^m \in W^*$ which are linearly independent so that $K = \cap_i \ker f^i$ has dimension $k = n - m$ in W. Let $|S|$ be the k-dimensional volume density on K. How can we compute an alternating, k-linear function η so that $|\eta(v_1, \ldots, v_k)| = |S|(v_1, \ldots, v_k)$ for $v_1, \ldots, v_k \in K$?

To ensure a unique solution to the problem, we may demand that $\eta(P_K v_1, \ldots, P_K v_k) = \eta(v_1, \ldots, v_k)$ where P_K denotes the orthogonal projection onto K, but this requirement will allow for exactly two solutions. To distinguish between η and $-\eta$, we choose an orientation ω on W, and require that the alternating n-linear function $(f^1 \wedge \ldots \wedge f^m) \wedge \eta$ be a *positive* multiple of ω (for example, when $f^1 = dz$ and $\eta = dx \wedge dy$, we notice $dz \wedge (dx \wedge dy) = dx \wedge dy \wedge dz$ gives the correct orientation on $\mathbb{R}^3$). In this way, the unique solution to our problem will depend on both the inner product and the ambient orientation.

One minor issue we must settle before proceeding is that the " $\wedge$ " in the formula

$$(f^1 \wedge \ldots \wedge f^m) \text{“} \wedge \text{”} \eta$$

[9] For the present purpose, $-\eta$ would also suffice. The precise issue of orientation is unimportant at the moment.

has not yet been defined to extend the wedge product to a product on alternating, multilinear functions. This definition can be accomplished by recalling our identification of alternating k-linear functions on W with the exterior power $\Lambda^k(W^*)$, and completing the construction of the **exterior algebra** of W^*. For our purposes, we can assume the existence of a well-defined, associative and bilinear extension of the wedge product, so that $\eta \wedge \tau \in \Lambda^{m+k}(W^*)$ whenever $\eta \in \Lambda^m(W^*)$ and $\tau \in \Lambda^k(W^*)$ and so that $(f^1 \wedge \ldots \wedge f^m) = 0$ whenever $f^1, \ldots, f^m$ are linearly dependent vectors in W^*. These properties of the wedge product suffice for computations and determine the wedge product uniquely.

To construct an alternating multilinear map η on W satisfying all of the above requirements, one begins by taking one of the two volume elements η on K and extending η to an alternating multilinear map on W via the orthogonal projection onto K. As this is clearly the only way to construct such a volume element, we have the following

Proposition 2.1. *Let W be an inner product space with a volume element ω. Then for every linearly independent set of functionals $f^1, \ldots, f^m \in W^*$ there exists a unique alternating m-linear function $\eta \in \Lambda^m(W^*)$ with the following properties*

- $\eta(v_1, \ldots, v_m) = \eta(P_K v_1, \ldots, P_K v_m)$ *where P_K denotes orthogonal projection onto the null space $K = \cap_i \ker f^i$*
- $(f^1 \wedge \ldots \wedge f^m) \wedge \eta = c\omega$ *for some $c > 0$.*
- *The restriction of η to K is a volume element on K*

Exercise 2.2. Show that the value c in the above proposition is the norm of $f^1 \wedge \ldots \wedge f^m$ as an element of $\Lambda^m(W^*)$ with the induced inner product. Note that it suffices to consider the case where $f^1, \ldots, f^m$ orthonormal.

Because the proposition characterizes the volume element uniquely in terms of basic properties, we can use these properties to compute the volume element η given $f^1, \ldots, f^m$. Let us denote by $\star(f^1, \ldots, f^m)$ the unique solution $\rho \in \Lambda^k(W^*)$ to $(f^1 \wedge \ldots \wedge f^m) \wedge \rho = g(f^1 \wedge \ldots \wedge f^m, f^1 \wedge \ldots \wedge f^m)\omega = ||f^1 \wedge \ldots \wedge f^m||^2\omega$ which is fixed by orthogonal projection onto K. The existence and uniqueness of $\star$ is guaranteed by the above proposition. Observe that $\star(f^1, \ldots, f^m)$ remains unchanged under parallel shifts of $f^1, \ldots, f^m$, and that replacing any f^i by αf^i results in a rescaling $\star(f^1, \ldots, \alpha f^i, \ldots f^m) = \alpha \star (f^1, \ldots, f^m)$ for any $\alpha \in \mathbb{R}$. Observe also that $\star(f^1, \ldots, f^m) = ||f^1 \wedge \ldots \wedge f^m||\eta$ where η is the volume element guaranteed by the above proposition. The upshot of what follows will be that $\star$ will extend to a **linear** map $* : \Lambda^m(W^*) \to \Lambda^{n-m}(W^*) = \Lambda^k(W^*)$ which we now define.

Definition 2.3. Given an inner product space $W^*, g(\cdot, \cdot)$ with volume element ω, we define the **Hodge star** $* : \Lambda^m(W^*) \to \Lambda^{n-m}(W^*)$ as the unique solution to

$$\xi \wedge *\rho = g(\xi, \rho) \cdot \omega$$

for all $\xi, \rho \in \Lambda^m(W^*)$, where $g(\xi, \rho)$ is the induced inner product on $\Lambda^m(W^*)$.

It is easy to verify that the above map is well-defined and linear using the universal property of Λ^m, the definition of the wedge product, and the linearity and nondegeneracy of the inner product $g(\cdot,\cdot)$ on $\Lambda^m(W^*)$ (**Exercise**). Now let us show that $*$ and $\star$ coincide as claimed.

Proposition 2.4. *The volume element η in Proposition (2.1) is given by*

$$\eta = \frac{1}{||f^1 \wedge \ldots \wedge f^m||} * (f^1 \wedge \ldots \wedge f^m)$$

Proof

We need only verify that $*(f^1 \wedge \ldots \wedge f^m) = \star(f^1, \ldots, f^m)$ by Exercise (2.2). We need only consider orthonormal vectors $f^1, \ldots, f^m$ since both $\star$ and $*$ are invariant under parallel shifts and are homogeneous in each f^i. Let $g^1, \ldots, g^k$ be an orthonormal basis of K^* (extended to W by orthogonal projection onto K) which is correctly oriented so that $f^1 \wedge \ldots \wedge f^m \wedge g^1 \wedge \ldots \wedge g^k = c\omega$ for a positive number c. Clearly $g^1 \wedge \ldots \wedge g^k = \star(f^1, \ldots, f^m)$. But, in fact, $c = 1$; indeed, $|g^1 \wedge \ldots \wedge g^k|$ serves as a volume function on K and $|f^1 \wedge \ldots \wedge f^m|$ serves as a volume function on the orthogonal complement of K, so the value $c = 1$ follows from the basic property of volume proven in Appendix Exercise (0.9). Having computed the c above, we may now expand any $\lambda \in \Lambda^m(W^*)$ relative to wedges of the basis elements $g^1, \ldots, g^k, f^1, \ldots, f^m$ for W^* in order to see that, in fact, $g^1 \wedge \ldots \wedge g^k = *(f^1 \wedge \ldots \wedge f^m)$ as desired.

■

Example To compute the length element $|dl|$ on the circle $C = \{r^2 = x^2 + y^2 = R^2\}$ of radius R in the (x, y) plane, one computes $d(r^2) = 2xdx + 2ydy$, so that $||d(r^2)|| = 2\sqrt{x^2 + y^2} = 2R$ on C, and $*d(r^2) = 2x * dx + 2y * dy = 2(xdy - ydx)$. Hence $|dl| = \frac{|xdy - ydx|}{R} = |Rd\theta|$

2.5. Pushforward of a Measure, Pullback of Distributions, and the Coarea Formula. Recall that the pushforward of a measure ϕdx on $\mathbb{R}^n$ by a map f is the measure $f_{\#}\phi$ with the property that

$$\int u(f(x))\phi(x)dx = \int u(t)df_{\#}\phi(t)$$

for all continuous, compactly supported functions u. When ϕ is the density for a random variable X, $f_{\#}\phi$ is the density function for the new random variable $f(X)$. This equality also allows us to define the pullback of a distribution u by f when the pushforward measure $f_{\#}\phi$ has a smooth, compactly supported density – the smoothness of the pushforward measure will be evident through the calculation of its density function that follows.

For a smooth map $f : \mathbb{R}^n \to \mathbb{R}^m$ whose derivative at every point is surjective, and a measure ϕdx on $\mathbb{R}^n$, we can apply the preceding results to compute the density on $\mathbb{R}^m$ of the pushforward $f_{\#}\phi dx$ of the measure ϕdx by f. In fact, we have the following formula, whose proof consists of decomposing the volume form $dx^1 \wedge \ldots \wedge dx^n = \frac{f_*\omega \wedge * f_*\omega}{||f_*\omega||^2}$:

Theorem 2.6. The Coarea Formula *Let $f = (f^1, \ldots, f^m)$ be a smooth map $f : \mathbb{R}^n \to \mathbb{R}^m$, let $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. Suppose that the derivative of f is at every point a surjective linear map – equivalently, suppose that the pullback $f^*\omega = df^1 \wedge \ldots \wedge df^m$ of the volume form $\omega = dy^1 \wedge \ldots \wedge dy^m$ is nowhere vanishing. The pushforward measure $f_\#\phi$ has density at $y \in \mathbb{R}^n$ given by*

$$\int_{f(x)=y} \phi(x) \frac{d\sigma}{||f_*\omega||} = \int_{f(x)=y} \phi(x) \frac{|* f_*\omega|}{||f_*\omega||^2}$$

Or, equivalently,

$$\int u(f(x))\phi(x)dx = \int_{\mathbb{R}^n} u(y) \left[\int_{f(x)=y} \phi(x) \frac{d\sigma}{||f_*\omega||} \right] dy$$

for continuous, compactly supported u. Here $d\sigma$ refers to the induced surface measure on the level sets of f.

Proof Taking a partition of unity if necessary, we can assume that $f^1, \ldots, f^m$ can be completed to a coordinate system $(f, h) = (f^1, \ldots, f^m, h^1, \ldots h^k)$ on an open neighborhood U of the support of ϕ in $\mathbb{R}^n$. Let $\Gamma : (f^1, \ldots, f^m, h^1, \ldots h^k) \to (x^1, \ldots, x^n)$ be the inverse map.

To avoid confusion, let us rename the variables $(f^1, \ldots, f^m) \mapsto (y^1, \ldots, y^m)$ in the domain of Γ. Then the map $\Gamma(y, h)$ gives a parameterization of U, and for fixed $(y^1, \ldots, y^m)$. We can assume the domain of Γ to be a rectangle. We also observe that $\Gamma(y, h)$ gives a parameterization of the level set $f(x) = y$ as the variables $h^1, \ldots, h^k$ vary. We can decompose the volume form $|dx^1 \wedge \ldots \wedge dx^n| = \frac{f_*\omega \wedge * f_*\omega}{||f_*\omega||^2}$, and pull this volume form back via Γ to integrate.

$$\begin{aligned}
\int u(f(x))\phi(x)dx &= \int u(f(\Gamma(y,h))\phi(\Gamma(y,h))\rho_\Gamma dy^1 \ldots dy^m dh^1 \ldots dh^k \\
&= \int u(y)\phi(\Gamma(y,h))\rho_\Gamma(y,h) dh^1 \ldots dh^k dy^1 \ldots dy^m
\end{aligned}$$

The local content of the Coarea Formula is that the density factor ρ_Γ when multiplied by $||f_*\omega||(\Gamma(y,h))$ is the same density factor that one would obtain when computing the k-dimensional surface integral of ϕ over the level set $f(x) = y$. After computing this density factor, the Coarea Formula follows from performing the integration in the h variables. To compute the density factor here, we use basic properties of the wedge product to see that

$$\begin{aligned}
\rho_\Gamma ||f_*\omega|| \cdot |dy^1 \wedge \ldots \wedge dy^m \wedge dh^1 \wedge \ldots \wedge dh^k| &= ||f_*\omega|| |\Gamma^* dx^1 \wedge \ldots \wedge dx^n| \\
&= \left| \Gamma^*[(df^1 \wedge \ldots \wedge df^m) \wedge \frac{(*f_*\omega)}{||f_*\omega||}] \right| \\
&= \left| \Gamma^*(df^1 \wedge \ldots \wedge df^m) \wedge \Gamma^* \frac{*f_*\omega}{||f_*\omega||} \right|
\end{aligned}$$

From the identity $f^i(\Gamma(y,h)) = y^i$ and the chain rule, we see that $df^i \circ D\Gamma = dy^i$ and hence $\Gamma^*(df^1 \wedge \ldots \wedge df^m) = dy^1 \wedge \ldots \wedge dy^m$. On the other hand, $\Gamma^* \frac{*f_*\omega}{||f_*\omega||} = c_\Gamma dh^1 \wedge \ldots \wedge dh^k$ for the same scalar c_Γ we use to compute integrals over the level sets $f(x) = y$. Inserting this information into the equality above, we conclude that $c_\Gamma = \rho_\Gamma ||f_*\omega||$. ∎

In this proof of the Coarea formula, and in the proof that integration over submanifolds is well-defined independent of parameterization, we have presupposed the change of variables formula for integrals. Since the change of variables formula is at the heart of these calculations, we indicate a proof of this formula in the following section.

3. The Change of Variables Formula

The Change of Variables Formula states that

Theorem 3.1. Change of Variables *Let $\Phi : U \to \Phi(U) \subseteq \mathbb{R}^n$ be a $\mathcal{C}^2$-diffeomorphism with everywhere bijective derivative $D\Phi$, and let f be a compactly supported, continuous function on $\Phi(U)$. Then,*

$$\int f(y)dy = \int_U f(\Phi(x)) |\det D\Phi(x)| dx$$

The regularity conditions on f and Φ are not optimal. The formula for more general f and Φ can be deduced from an approximation argument, or from taking more care with the following proof, but at the moment we are only concerned with the proof of the above form. The formula is intuitively clear when one understands the fact that $|\det D\Phi(x)|$ is the factor by which Φ distorts volumes of small neighborhoods of a point x, since $|\det D\Phi(x)|$ is exactly the factor by which the linearization of Φ distorts volumes. Thus, we must first understand the case where $\Phi = D\Phi = T$ is an invertible linear map.

Lemma 3.2. Change of Variables for Linear Maps *For any invertible, linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ and continuous, compactly supported $f \in \mathcal{C}_0(\mathbb{R}^n)$*

$$\int f(y)dy = \int f(Tx)) |\det T| dx$$

This fact essentially boils down to the following fundamental observations: the pushforward of the Lebesgue measure by an invertible linear map is also a translation-invariant measure (which endows compact sets with finite measure and open sets with positive measure), and, up to a positive scalar multiple, there is only one such measure on $\mathbb{R}^n$ with these properties – namely, the Lebesgue measure, dx. These observations imply that $T_{\#}dx = \rho_T dx$ for a positive, scalar ρ_T, which we need to check is actually $|\det T|^{-1}$.

The uniqueness of the translation invariant measure alluded to above can be proved with one's bare hands. Indeed, it is clear on $\mathbb{R}^2$ that once one assigns a measure $\lambda > 0$ to the unit square $[0,1]\times[0,1]$, then translation invariance forces the measure of $[0,1/2]\times[0,1/2]$ to be $\frac{\lambda}{4}$, and so on for smaller dyadic cubes. But then any set one can imagine can be well-approximated by a net of disjoint, dyadic cubes (as in the Whitney decomposition of an open set). Here we can take a slicker route, and use essentially the same idea to prove the preliminary lemma that

Lemma 3.3. Pre-Change of Variables for Linear maps *For any invertible, linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ and continuous, compactly supported $f \in \mathcal{C}_0(\mathbb{R}^n)$, there exists a positive constant $\Delta(T) > 0$ so that*

$$\int f(y)dy = \int f(Tx)\Delta(T)dx$$

Proof The proof will involve several changes of variable by translation. Let $g \in \mathcal{C}_0(\mathbb{R}^n)$ be a continuous, compactly supported function normalized so that $\int g(y)dy = 1$. Then for an arbitrary continuous, compactly supported function f, we have

$$\begin{aligned}
\int f(Tx)dx &= \int g(y)dy \int f(Tx)dx \\
&= \int g(y)[\int f(Tx+y)dx]dy \\
&= \int\int g(y-Tx)f(y)dydx \\
&= \int g(-Tx)dx \cdot \int f(y)dy
\end{aligned}$$

which proves the lemma for $\frac{1}{\Delta(T)} = \int g(-Tx)dx$. Note that if g were chosen to be localized on a dyadic cube of small support, and f is the indicator function of a set, then the proof resembles the direct geometric argument outlined before. A similar idea will be used later in the proof of the general change of variables formula. ■

For now, let us postpone the proof that the $\Delta(T)$ guaranteed above is equal to $|\det T|$ and let us take this fact as granted to prove the general change of variables formula.

A typical proof of the change of variables formula may proceed by decomposing the region $\Phi(U)$ containing the support of f in the integral $\int f(y)dy$ into small pieces. If $y_0 = \Phi(x_0)$, then each y within some ϵ-neighborhood of y_0, has a unique preimage x so that $\Phi(x) = y$, and the volume of this preimage is proportional to the volume of the ϵ-neighborhood by a factor $|\det D\Phi(x_0)|^{-1}$ as ϵ tends to zero. Summing over disjoint ϵ-pieces, we would obtain the right hand side of the change of variables formula in the limit as ϵ tends to 0. This is essentially how we will proceed, but there is a nicer way to iron out the details [10]

Proof [Change of Variables Formula]

We now assume the change of variables formula for linear maps, and the translation invariance of the integral. Let $\Phi : U \to \Phi(U)$ be a $\mathcal{C}^2$ diffeomorphism with $\mathcal{C}^2$ inverse Ψ, and let f be a continuous function whose support is a compact subset of $\Phi(U) \subseteq \mathbb{R}^n$. Let $f_\epsilon(y) = \int f(t)\eta_\epsilon(y-t)dt$ be a smooth mollifier for f, where $\eta_\epsilon(y) = \epsilon^{-n}\eta(\frac{y}{\epsilon})$ is a smooth function supported in a ball of radius ϵ and normalized so that $\int \eta_\epsilon(y)dy = \int \eta dy = 1$ (by the change of variables formula for dilations). It is easily shown (by uniform continuity of f) that the sequence f_ϵ converges to f uniformly because the value $f_\epsilon(y)$ is an average of the values of f over a small neighborhood of the point y.

We can therefore compute

$$\int f(\Phi(x))|\det D\Phi(x)|dx = \lim_{\epsilon\to 0}\int_{\Phi(U)} f(t)\int_U \eta_\epsilon(\Phi(x)-t)|\det D\Phi(x)|dxdt$$

Since Φ is onto $\Phi(U)$, we can write

$$\int f(\Phi(x))|\det D\Phi(x)|dx = \lim_{\epsilon\to 0}\int_{\Phi(U)} f(t)\int_U \eta_\epsilon\left(\Phi(x)-\Phi(\Psi(t))\right)|\det D\Phi(x)|dxdt$$

Because Φ is one-to one, the region over which the inner integral takes place (namely, the image under Ψ of the ϵ-ball about t) is small – in fact, it is contained in a ball of radius $O(\epsilon)$ because Ψ is Lipschitz. We therefore wish to replace $|\det D\Phi(x)|$ with the constant value $|\det D\Phi(\Psi(t))|$ (the difference being of size $O(\epsilon)$). We also should linearize the argument of η_ϵ since the difference $\Phi(x) - \Phi(\Psi(t))$ is equal to $D\Phi(\Psi(t))(x-\Psi(t))$ up to some $O(\epsilon^2)$ error when Φ is $\mathcal{C}^2$. It is a simple exercise (involving the change of variables for dilations and differentiating η_ϵ) to show that these errors are acceptable, leaving us with.

$$\int f(\Phi(x))|\det D\Phi(x)|dx$$
$$= \lim_{\epsilon\to 0}\int_{\Phi(U)} f(t)\int_U \eta_\epsilon\left(D\Phi(\Psi(t))(x-\Psi(t))\right)|\det D\Phi(\Psi(t))|dxdt$$

Now we apply the change of variables for linear maps to the dx integral (giving the value 1) to obtain the general change of variables formula

$$\int f(\Phi(x))|\det D\Phi(x)|dx = \int f(t)dt$$

■

Let us complete the proof by showing that the constant $\Delta(T)$ in Lemma (3.3) is equal to $|\det T|$. We have already shown that the application of a linear map to

[10] From the point of view of these lecture notes, this proof is related to interpreting the change of variables as a decomposition $f(x) = \int f(t)\delta(x-t)dt$.

the Lebesgue measure results in a proportional measure $T_{\#}dx = \frac{1}{\Delta(T)}dx$. Applying two successive transformations yields a proportional measure $\frac{1}{\Delta(T_1 \circ T_2)}dx = (T_1 \circ T_2)_{\#}dx = (T_t)_{\#}(T_2)_{\#}dx = \frac{1}{\Delta(T_1)\Delta(T_2)}dx$; so we conclude that $\Delta(T_1 \circ T_2) = \Delta(T_1)\Delta(T_2)$ in the same way that $|\det(T_1 \circ T_2)| = |\det T_1||\det T_2|$. This homomorphism property reduces the task of showing $\Delta(T) = |\det T|$ to proving equality for any collection of T which generates the group of invertible matrices.

Thus, since every linear map may be written as a composition of maps of the following three types, it suffices to consider the following three cases:

Case I: If T permutes the standard basis vectors $e_1 = (1, 0, \dots)$, $e_2 = (0, 1, \dots)$, $\dots$, then $|\det T| = |\pm 1| = 1$, and $\Delta(T) = 1$ since $f(Tx) = f(x)$ for any function on $\mathbb{R}^n$ which is symmetric in the variables $(x_1, \dots, x_n)$.

Case II: If $Te_1 = \alpha e_1$, $\alpha \in \mathbb{R}$, and $Te_j = e_j$ for $j \neq 1$, then $|\det T| = |\alpha|$, and furthermore $\Delta(T) = |\alpha|$ also, since $\int f(Tx)|\alpha|dx = \int f(y)dy$ when f is the characteristic function of the unit cube $0 < x_i < 1$.[11]

Case III: If $Te_j = e_j$ for $j \geq 2$ and $Te_1 = e_1 + e_2$ (so that T is a parallel shift), then $|\det T| = \det T = 1$. On the other hand, the image of the cube $Q = [0, 1)^n$ under the parallel shift T has the same measure as the unit cube. To see this equality of measures, one can decompose the image of Q into two parts: $Q \cap T(Q)$ and $T(Q)\backslash Q$. One can then reconstruct Q by translating $T(Q)\backslash Q$ by $-e_2$; hence, Q and $T(Q)$ have the same measure, and $\Delta(T) = 1$ as desired.

[11] The characteristic function of the unit cube is not a continuous function, but it is a monotone limit of continuous functions of compact support, so the change of variables formula for linear maps still applies by the monotone convergence theorem.

APPENDIX B

Basic Concepts in Riemannian and Lorentzian Geometry

In what follows we give a short overview of the basic notions in Riemannian and Lorentzian geometry. These will allow us to extend some of the basic facts about the standard Laplace, Heat and Wave equations, to manifolds. It will also allow us later to discuss more complicated nonlinear geometric equations.

1. Introduction

A pseudo-riemannian manifold [1], or simply a spacetime, consist of a pair $(\mathbf{M}, \mathbf{g})$ where $\mathbf{M}$ is an orientable $p+q$-dimensional manifold and $\mathbf{g}$ is a pseudo-riemannian metric defined on it, that is a smooth, a non degenerate, 2-covariant symmetric tensor field of signature (p, q). This means that at each point $p \in \mathbf{M}$ one can choose a basis of $p+q$ vectors, $\{e_{(\alpha)}\}$, belonging to the tangent space $T\mathbf{M}_p$, such that

$$\mathbf{g}(e_{(\alpha)}, e_{(\beta)}) \quad = \quad \eta_{\alpha\beta} \tag{417}$$

for all $\alpha, \beta = 0, 1, ..., n$, where η is the diagonal matrix with -1 in the first p entries and $+1$ in the last q entries. If X is an arbitrary vector at p expressed, in terms of the basis $\{e_{(\alpha)}\}$, as $X = X^\alpha e_{(\alpha)}$, we have

$$\mathbf{g}(X, X) = -(X^1)^2 - \ldots - (X^p)^2 + (X^{p+1})^2 + ... + (X^{p+q})^2 \tag{418}$$

The case when $p = 0$ and $q = n$ corresponds to Riemannian manifolds of dimension n. The other case of interest for us is $p = 1, q = n$ which corresponds to a Lorentzian manifolds of dimension $n+1$. The primary example of Riemannian manifold is the Euclidean space $\mathbb{R}^n$. Any other Riemannian manifold looks, locally, like $\mathbb{R}^n$. Similarly, the primary example of a Lorentzian manifold is the Minkowski spacetime, the spacetime of Special Relativity. It plays the same role, in Lorentzian geometry, as the Euclidean space in Riemannian geometry. In this case the manifold $\mathbf{M}$ is diffeomorphic to $\mathbb{R}^{n+1}$ and there exists globally defined systems of coordinates, x^α, relative to which the metric takes the diagonal form $-1, 1, ..., 1$. All such systems are related through Lorentz transformations and are called inertial. We shall denote the Minkowski spacetime of dimension $n+1$ by $(\mathbb{R}^{n+1}, \mathbf{m})$.

[1] We assume that our reader is already familiar with the basics concepts of differential geometry such as manifolds, tensor fields, covariant, Lie and exterior differentiation. For a short introduction to these concepts see Chapter 2 of Hawking and Ellis, "The large scale structure of space-time", [**?**]

Relative to a given coordinate system x^μ, the components of a pseudo-riemannian metric take the form

$$g_{\mu\nu} = \mathbf{g}(\partial_\mu, \partial_\nu)$$

where $\partial_\mu = \frac{\partial}{\partial x^\mu}$ are the associated coordinate vectorfields. We denote by $g^{\mu\nu}$ the components of the inverse metric g^{-1} relative to the same coordinates x, and by $|g|$ the determinant of the matrix $g_{\mu\nu}$. The volume element $dv_{\mathbf{M}}$ of $\mathbf{M}$ is expressed, in local coordinates, by $\sqrt{|g|}dx = \sqrt{|g|}dx^1 \dots dx^n$. Thus the integral $\int_{\mathbf{M}} f dv_{\mathbf{M}}$ of a function f, supported in coordinate chart $U \subset \mathbf{M}$ is defined by $\int_U f(x)\sqrt{|g(x)|}dx$. The integral on $\mathbf{M}$ of an arbitrary function f is defined by making a partition of unity subordinated to a covering of $\mathbf{M}$ by coordinate charts. One can easily check that the definition is independent of the particular system of local coordinates.

In view of (418) we see that a Lorentzian metric divides the vectors in the tangent space $T\mathbf{M}_p$ at each p, into timelike, null or spacelike according to whether the quadratic form

$$(X, X) = g_{\mu\nu}X^\mu X^\nu \tag{419}$$

is, respectively, negative, zero or positive. The set of null vectors N_p forms a double cone, called the null cone of the corresponding point p. The set of timelike vectors I_p forms the interior of this cone. The vectors in the union of I_p and N_p are called causal. The set S_p of spacelike vectors is the complement of $I_p \cup N_p$.

A frame $e_{(\alpha)}$ verifying (417) is said to be orthonormal. In the case of Lorentzian manifolds it makes sense to consider, in addition to orthonormal frames, null frames. These are collections of vectorfields[2] e_α consisting of two null vectors e_{n+1}, e_n and orthonormal spacelike vectors $(e_a)_{a=1,\dots,n-1}$ which verify,

$$\mathbf{g}(e_n, e_n) = \mathbf{g}(e_{n+1}, e_{n+1}) = 0\ ,\ \mathbf{g}(e_n, e_{n+1}) = -2$$
$$\mathbf{g}(e_n, e_a) = \mathbf{g}(e_{n+1}, e_a) = 0\ ,\ \mathbf{g}(e_a, e_b) = \delta_{ab}$$

One-forms $A = A_\alpha dx^\alpha$ are sections of the cotangent bundle of $\mathbf{M}$. We denote by $A(X)$ the natural pairing between A and a vectorfield X. We can raise the indices of A by $A^\alpha = \mathbf{g}^{\alpha\beta}\mathbf{A}_\beta$. $A' = A^\alpha\partial_\alpha$ defines a vectorfield on $\mathbf{M}$ and we have, $A(X) = \mathbf{g}(A', X)$. Covariant tensors A of order k are k-multilinear forms on $T\mathbf{M}$.

Notation: We will use the following notational conventions: We shall use boldface characters to denote important tensors such as the metric $\mathbf{g}$, and the Riemann curvature tensor $\mathbf{R}$. Their components relative to arbitrary frames will also be denoted by boldface characters. Thus, given a frame $\{e_{(\alpha)}\}$ we write $\mathbf{g}_{\alpha\beta} = \mathbf{g}(e_\alpha, e_\beta)$, $\mathbf{R}_{\alpha\beta\gamma\delta} = \mathbf{R}(e_\alpha, e_\beta, e_\gamma, e_\delta)$ and, for an arbitrary tensor T,

$$T_{\alpha\beta\gamma\delta\dots} \equiv T(e_\alpha, e_\beta, e_\gamma, e_\delta, \dots)$$

We shall not use boldface characters for the components of tensors, relative to a fixed system of coordinates. Thus, for instance, in (419) $g_{\mu\nu} = \mathbf{g}(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu})$. In the

[2]We write e_α instead of $e_{(\alpha)}$ to simplify the notation, whenever there can be no confusion.

case of a Riemannian manifold we use latin letters $i, j, k, l, \dots$ to denote indices of coordinates $x^1, x^2, \dots, x^n$ or tensors. For a Lorentzian manifold we use greek letters $\alpha, \beta, \gamma, \dots$ to denote indices $0, 1, \dots, n$.

We will review the following topics below:

1.) Lie brackets of vectorfields. Frobenius theorem

2.) Lie derivative of a tensorfield

3.) Multilinear forms and exterior differentiation

4.) Connections and covariant derivatives

5.) Pseudo-riemannian metrics. Riemannian and Lorentzian geometry.

6.) Levi-Civita connection associated to a pseudo-riemannian metric.

7.) Parallel transport, geodesics, exponential map, completeness

8.) Curvature tensor of a pseudo-riemannian manifold. Symmetries. First and second Bianchi identities.

9.) Isometries and conformal isometries. Killing and conformal Killing vectorfields.

2. Various notions of differentiation

We recall here the three fundamental operators of the differential geometry on a Riemann or Lorentz manifold: the exterior derivative, the Lie derivative, and the Levi-Civita connection with its associated covariant derivative.

2.1. The exterior derivative. Given a scalar function f its differential df is the 1-form defined by

$$df(X) = X(f)$$

for any vector field X. This definition can be extended for all differential forms on $\mathbf{M}$ in the following way:

i) d is a linear operator defined from the space of all k-forms to that of $k+1$-forms on $\mathbf{M}$. Thus for all k-forms A,B and real numbers λ, μ

$$d(\lambda A + \mu B) = \lambda dA + \mu dB$$

ii) For any k-form A and arbitrary form B

$$d(A \wedge B) = dA \wedge B + (-1)^k A \wedge dB$$

iii) For any form A,

$$d^2 A = 0 \ .$$

We recall that, if Φ is a smooth map defined from $\mathbf{M}$ to another manifold $\mathbf{M}'$, then

$$d(\Phi^* A) = \Phi^*(dA) \ .$$

Finally if A is a one form and X, Y arbitrary vector fields, we have the equation

$$dA(X,Y) = \frac{1}{2}\Big(X(A(Y)) - Y(A(X)) - A([X,Y]) \Big)$$

where $[X,Y]$ is the commutator $X(Y) - Y(X)$. This can be easily generalised to arbitrary k forms, see Spivak's book, Vol.I, Chapter 7, Theorem 13. [**?**]

2.2. Lie derivative. Consider an arbitrary vector field X. In local coordinates x^μ, the flow of X is given by the system of differential equations

$$\frac{dx^\mu}{dt} = X^\mu(x^1(t), ..., x^{p+q}(t)) \ .$$

The corresponding curves, $x^\mu(t)$, are the integral curves of X. For each point $p \in \mathbf{M}$ there exists an open neighborhood $\mathcal{U}$, a small $\epsilon > 0$ and a family of diffeomorphism $\Phi_t : \mathcal{U} \to \mathbf{M}$, $|t| \le \epsilon$, obtained by taking each point in $\mathcal{U}$ to a parameter distance t, along the integral curves of X. We use these diffeomorphisms to construct, for any given tensor T at p, the family of tensors $(\Phi_t)_* T$ at $\Phi_t(p)$.

The Lie derivative $\mathcal{L}_X T$ of a tensor field T, with respect to X, is:

$$\mathcal{L}_X T|_p \equiv \lim_{t \to 0} \frac{1}{t} \left(T|_p - (\Phi_t)_* T|_p \right) \ .$$

It has the following properties:

i) $\mathcal{L}_X$ linearly maps (p,q)-tensor fields into tensor fields of the same type.

ii) $\mathcal{L}_X$ commutes with contractions.

iii) For any tensor fields S, T,

$$\mathcal{L}_X(S \otimes T) = \mathcal{L}_X S \otimes T + S \otimes \mathcal{L}_X T \ .$$

If X is a vector field we easily check that

$$\mathcal{L}_X Y = [X, Y]$$

by writing $(\mathcal{L}_X Y)^i = -\frac{d}{dt}((\Phi_t)_* Y)^i \Big|_{t=0}$ and expressing $(\Phi_t)_* Y)^i \Big|_p = \frac{\partial x^i(\Phi_t(q))}{\partial x^j(q)} Y^j \Big|_q$, where $q = \Phi_{-t}(p)$. (See [**?**], Hawking and Ellis, section 2.4 for details.)

If A is a k-form we have, as a consequence of the commutation formula of the exterior derivative with the pull-back Φ^*,

$$d(\mathcal{L}_X A) = \mathcal{L}_X(dA) \ .$$

For a given k-covariant tensorfield T we have,

$$\mathcal{L}_X T(Y_1, \dots, Y_k) = X T(Y_1, \dots, Y_k) - \sum_{i=1}^{k} T(Y_1, \dots, \mathcal{L}_X Y_i, \dots, Y_k)$$

We remark that the Lie bracket of two coordinate vector fields vanishes,

$$\left[\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right] = 0.$$

The converse is also true, namely, see Spivak, [**?**], Vol.I, Chapter 5,

Proposition 2.3. *If $X_{(0)}, ..., X_{(k)}$ are linearly independent vector fields in a neighbourhood of a point p and the Lie bracket of any two of them is zero then there exists a coordinate system x^μ, around p such that $X_{(\rho)} = \frac{\partial}{\partial x^\rho}$ for each $\rho = 0, ..., k$.*

The above proposition is the main step in the proof of Frobenius Theorem. To state the theorem we recall the definition of a k-distribution in $\mathbf{M}$. This is an arbitrary smooth assignment of a k-dimensional plane π_p at every point in a domain $\mathcal{U}$ of $\mathbf{M}$. The distribution is said to be involute if, for any vector fields X, Y on $\mathcal{U}$ with $X|_p, Y|_p \in \pi_p$, for any $p \in \mathcal{U}$, we have $[X, Y]|_p \in \pi_p$. This is clearly the case for integrable distributions[3]. Indeed if $X|_p, Y|_p \in T\mathcal{N}_p$ for all $p \in \mathcal{N}$, then X, Y are tangent to $\mathcal{N}$ and so is also their commutator $[X, Y]$. The Frobenius Theorem establishes that the converse is also true[4], that is being in involution is also a sufficient condition for the distribution to be integrable,

Theorem 2.4. *(Frobenius Theorem) A necessary and sufficient condition for a distribution $(\pi_p)_{p\in\mathcal{U}}$ to be integrable is that it is involute.*

2.5. The connection and the covariant derivative. A connection $\mathbf{D}$ is a rule which assigns to each vectorfield X a differential operator $\mathbf{D}_X$. This operator maps vector fields Y into vector fields $\mathbf{D}_X Y$ in such a way that, with $\alpha, \beta \in \mathbb{R}$ and f, g scalar functions on $\mathbf{M}$,

$$\begin{aligned} &a)\ \ \mathbf{D}_{fX+gY} Z = f\mathbf{D}_X Z + g\mathbf{D}_Y Z \\ &b)\ \ \mathbf{D}_X(\alpha Y + \beta Z) = \alpha\mathbf{D}_X Y + \beta\mathbf{D}_X Z \\ &c)\ \ \mathbf{D}_X fY = X(f)Y + f\mathbf{D}_X Y \end{aligned} \tag{420}$$

Therefore, at a point p,

$$\mathbf{D}Y \equiv Y^\alpha_{;\beta}\, \theta^{(\beta)} \otimes e_{(\alpha)} \tag{421}$$

where the $\theta^{(\beta)}$ are the one-forms of the dual basis respect to the orthonormal frame $e_{(\beta)}$. Observe that $Y^\alpha_{;\beta} = \theta^{(\alpha)}(\mathbf{D}_{e_{(\beta)}} Y)$. On the other side, from c),

$$\mathbf{D}fY = df \otimes Y + f\mathbf{D}Y$$

so that

$$\mathbf{D}Y = \mathbf{D}(Y^\alpha e_{(\alpha)}) = dY^\alpha \otimes e_{(\alpha)} + Y^\alpha \mathbf{D}e_{(\alpha)}$$

and finally, using $df(\cdot) = e_{(\alpha)}(f)\theta^{(\alpha)}(\cdot)$,

$$\mathbf{D}Y = \Big(e_{(\beta)}(Y^\alpha) + Y^\gamma \theta^{(\alpha)}(\mathbf{D}_{e_{(\beta)}} e_{(\gamma)})\Big)\, \theta^{(\beta)} \otimes e_{(\alpha)} \tag{422}$$

[3]Recall that a distribution π on $\mathcal{U}$ is said to be integrable if through every point $p \in \mathcal{U}$ there passes a unique submanifold $\mathcal{N}$, of dimension k, such that $\pi_p = T\mathcal{N}_p$.

[4]For a proof see Spivak, citeSpivak, Vol.I, Chapter 6.

Therefore

$$Y^{\alpha}_{;\beta} = e_{(\beta)}(Y^{\alpha}) + \Gamma^{\alpha}_{\beta\gamma} Y^{\gamma}$$

and the connection is, therefore, determined by its connection coefficients,

$$\Gamma^{\alpha}_{\beta\gamma} = \theta^{(\alpha)}(\mathbf{D}_{e_{(\beta)}} e_{(\gamma)}) \tag{423}$$

which, in a coordinate basis, are the usual Christoffel symbols and have the expression

$$\Gamma^{\mu}_{\rho\nu} = dx^{\mu}(\mathbf{D}_{\frac{\partial}{\partial x^{\rho}}} \frac{\partial}{\partial x^{\nu}})$$

Finally

$$\mathbf{D}_X Y = \left(X(Y^{\alpha}) + \Gamma^{\alpha}_{\beta\gamma} X^{\beta} Y^{\gamma} \right) e_{(\alpha)} \tag{424}$$

In the particular case of a coordinate frame we have

$$\mathbf{D}_X Y = \left(X^{\mu} \frac{\partial Y^{\nu}}{\partial x^{\mu}} + \Gamma^{\nu}_{\rho\sigma} X^{\rho} Y^{\sigma} \right) \frac{\partial}{\partial x^{\nu}}$$

A connection is said to be a Levi-civita connection if $\mathbf{Dg} = 0$. That is, for any three vector fields X, Y, Z,

$$Z(\mathbf{g}(X,Y)) = \mathbf{g}(\mathbf{D}_Z X, Y) + \mathbf{g}(X, \mathbf{D}_Z Y) \tag{425}$$

A very simple and basic result of differential geometry asserts that for any given metric there exists a unique affine connection associated to it.

Proposition 2.6. *There exists a unique connection on* $\mathbf{M}$*, called the Levi-Civita connection, which satisfies* $\mathbf{D}\,\mathbf{g} = 0$*. The connection is torsion free, that is,*

$$\mathbf{D}_X Y - \mathbf{D}_Y X = [X, Y] \ .$$

Moreover, relative to a system of coordinates, x^{μ}*, the Christoffel symbol of the connection is given by the standard formula*

$$\Gamma^{\mu}_{\rho\nu} = \frac{1}{2} g^{\mu\tau} \left(\partial_{\rho} g_{\nu\tau} + \partial_{\nu} g_{\tau\rho} - \partial_{\tau} g_{\nu\rho} \right) \ .$$

Exercise: Prove the proposition yourself, without looking in a book.

So far we have only defined the covariant derivative of a a vector field. We can easily extend the definition to one forms $A = A_{\alpha} dx^a$ by the requirement that,

$$X(A(Y)) = \mathbf{D}_X A(Y) + A(D_X Y),$$

for all vectorfields X, Y. Given a k-covariant tensor field T we define its covariant derivative $\mathbf{D}_X T$ by the rule,

$$\mathbf{D}_X T(Y_1, \dots, Y_k) = XT(Y_1, \dots, Y_k) - \sum_{i=1}^{k} T(Y_1, \dots, \mathbf{D}_X Y_i, \dots, Y_k)$$

We can talk about $\mathbf{D}T$ as a covariant tensor of rank $k+1$ defined by,

$$\mathbf{D}T(X, Y_1, \dots, Y_k) = \mathbf{D}_X T(Y_1, \dots, Y_k).$$

Given a frame e_α we denote by $T_{\alpha_1\dots,\alpha_k;\beta} = \mathbf{D}T(e_\beta, e_{a_1}, \dots, e_{\alpha_k})$ the components of $\mathbf{D}T$ relative to the frame. By repeated covariant differentiation we can define $\mathbf{D}^2T, \dots \mathbf{D}^m\mathbf{T}$. Relative to a frame e_α we write,

$$\mathbf{D}_{\beta_1}\dots\mathbf{D}_{\beta_m}T_{\alpha_1\dots\alpha_k} = T_{\alpha_1\dots\alpha_k;\beta_1\dots\beta_m} = \mathbf{D}^mT(e_{\beta_1}\dots, e_{\beta_m}, e_{\alpha_1}, ..., e_{\alpha_k}).$$

The fact that the Levi-Civita connection is torsion free allows us to connect covariant differentiation to the Lie derivative. Thus, if T is a k-covariant tensor we have, in a coordinate basis,

$$(\mathcal{L}_XT)_{\sigma_1\dots\sigma_k} = X^\mu T_{\sigma_1\dots\sigma_k;\mu} + X^\mu{}_{;\sigma_1}T_{\mu\sigma_2\dots\sigma_k} + ... + X^\mu{}_{;\sigma_k}T_{\sigma_1\dots\sigma_{k-1}\mu}\ .$$

The covariant derivative is also connected to the exterior derivative according to the following simple formula. If A is a k-form, we have[5] $A_{[\sigma_1\dots\sigma_k;\mu]} = A_{[\sigma_1\dots\sigma_k,\mu]}$ and

$$dA = \sum A_{\sigma_1\dots\sigma_k;\mu}dx^\mu \wedge dx^{\sigma_1} \wedge dx^{\sigma_2} \wedge .... \wedge dx^{\sigma_k}\ .$$

Given a smooth curve $\mathbf{x} : [0,1] \to \mathbf{M}$, parametrized by t, let $T = \left(\frac{\partial}{\partial t}\right)_{\mathbf{x}}$ be the corresponding tangent vector field along the curve. A vector field X, defined on the curve, is said to be parallelly transported along it if $\mathbf{D}_TX = 0$. If the curve has the parametric equations $x^\nu = x^\nu(t)$, relative to a system of coordinates, then $T^\mu = \frac{dx^\mu}{dt}$ and the components $X^\mu = X^\mu(\mathbf{x}(t))$ satisfy the ordinary differential system of equations

$$\frac{\mathbf{D}}{dt}X^\mu \equiv \frac{dX^\mu}{dt} + \Gamma^\mu_{\rho\sigma}(\mathbf{x}(t))\frac{dx^\rho}{dt}X^\sigma = 0\ .$$

The curve is said to be geodesic if, at every point of the curve, $\mathbf{D}_TT$ is tangent to the curve, $\mathbf{D}_TT = \lambda T$. In this case one can reparametrize the curve such that, relative to the new parameter s, the tangent vector $S = \left(\frac{\partial}{\partial s}\right)_{\mathbf{x}}$ satisfies $\mathbf{D}_SS = 0$. Such a parameter is called an “affine parameter”. The affine parameter is defined up to a transformation $s = as' + b$ for a, b constants. Relative to an affine parameter s and arbitrary coordinates x^μ the geodesic curves satisfy the equations

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{ds}\frac{dx^\sigma}{ds} = 0\ .$$

A geodesic curve parametrized by an affine parameter is simply called a geodesic. In Lorentzian geometry timelike geodesics correspond to world lines of particles freely falling in the gravitational field represented by the connection coefficients. In this case the affine parameter s is called the proper time of the particle.

Given a point $p \in \mathbf{M}$ and a vector X in the tangent space $T_p\mathbf{M}$, let $\mathbf{x}(t)$ be the unique geodesic starting at p with “velocity” X. We define the exponential map:

$$\exp_p : T_p\mathbf{M} \to \mathbf{M}\ .$$

This map may not be defined for all $X \in T_p\mathbf{M}$. The theorem of existence and uniqueness for systems of ordinary differential equations implies that the exponential map is defined in a neighbourhood of the origin in $T_p\mathbf{M}$. If the exponential

[5] $[\sigma_1...\sigma_k;\mu]$ indicates the antisymmetrization with respect to all indices (i.e. $\frac{1}{k!}$(alternating sum of the tensor over all permutations of the indices)) and “$,\mu$” indicates the ordinary derivative with respect to x^μ.

map is defined for all $T_p\mathbf{M}$, for every point p the manifold $\mathbf{M}$ is said geodesically complete. In general if the connection is a C^r connection[6] there exists an open neighbourhood $\mathcal{U}_0$ of the origin in $T_p\mathbf{M}$ and an open neighbourhood of the point p in $\mathbf{M}$, $\mathcal{V}_p$, such that the map $\exp_p$ is a C^r diffeomorphism of $\mathcal{U}_0$ onto $\mathcal{V}_p$. The neighbourhood $\mathcal{V}_p$ is called a normal neighbourhood of p.

3. Riemann curvature tensor, Ricci tensor, Bianchi identities

Riemann curvature tensor, Ricci tensor, Bianchi identities

In the flat spacetime if we parallel transport a vector along any closed curve we obtain the vector we have started with. This fails in general because the second covariant derivatives of a vector field do not commute. This lack of commutation is measured by the Riemann curvature tensor,

$$\mathbf{R}(X,Y)Z = \mathbf{D}_X(\mathbf{D}_Y Z) - \mathbf{D}_Y(\mathbf{D}_X Z) - \mathbf{D}_{[X,Y]}Z \tag{426}$$

or written in components relative to an arbitrary frame,

$$\mathbf{R}^{\alpha}{}_{\beta\gamma\delta} = \theta^{(\alpha)}\big((\mathbf{D}_\gamma\mathbf{D}_\delta - \mathbf{D}_\delta\mathbf{D}_\gamma)e_{(\beta)}\big) \tag{427}$$

Relative to a coordinate system x^μ and written in terms of the $g_{\mu\nu}$ components, the Riemann components have the expression

$$R^{\mu}{}_{\nu\rho\sigma} = \frac{\partial\Gamma^{\mu}_{\sigma\nu}}{\partial x^\rho} - \frac{\partial\Gamma^{\mu}_{\rho\nu}}{\partial x^\sigma} + \Gamma^{\mu}_{\rho\tau}\Gamma^{\tau}_{\sigma\nu} - \Gamma^{\mu}_{\sigma\tau}\Gamma^{\tau}_{\rho\nu} \tag{428}$$

The fundamental property of the curvature tensor, first proved by Riemann, states that if $\mathbf{R}$ vanishes identically in a neighbourhood of a point p one can find families of local coordinates such that, in a neighbourhood of p, $g_{\mu\nu} = \eta_{\mu\nu}$ [7].

The trace of the curvature tensor, relative to the metric $\mathbf{g}$, is a symmetric tensor called the Ricci tensor,

$$\mathbf{R}_{\alpha\beta} = \mathbf{g}^{\gamma\delta}\mathbf{R}_{\alpha\gamma\beta\delta}$$

The scalar curvature is the trace of the Ricci tensor

$$\mathbf{R} = \mathbf{g}^{\alpha\beta}\mathbf{R}_{\alpha\beta}\ .$$

The Riemann curvature tensor of an arbitrary spacetime $(\mathbf{M},\mathbf{g})$ has the following symmetry properties,

$$\begin{aligned} &\mathbf{R}_{\alpha\beta\gamma\delta} = -\mathbf{R}_{\beta\alpha\gamma\delta} = -\mathbf{R}_{\alpha\beta\delta\gamma} = \mathbf{R}_{\gamma\delta\alpha\beta} \\ &\mathbf{R}_{\alpha\beta\gamma\delta} + \mathbf{R}_{\alpha\gamma\delta\beta} + \mathbf{R}_{\alpha\delta\beta\gamma} = 0 \end{aligned} \tag{429}$$

The second identity in (429) is called the first Bianchi identity.

It also satisfies the second Bianchi identities, which we refer to here as the *Bianchi equations* and, in a generic frame, have the form:

$$\mathbf{D}_{[\epsilon}\mathbf{R}_{\gamma\delta]\alpha\beta} = 0 \tag{430}$$

---

[6] A C^r connection is such that if Y is a C^{r+1} vector field then $\mathbf{D}Y$ is a C^r vector field.

[7] For a thorough discussion and proof of this fact, refer to Spivak, [**?**], Vol. II.

The traceless part of the curvature tensor, $\mathbf{C}$ is called the Weyl tensor, and has the following expression in an arbitrary frame,

$$\begin{aligned}\mathbf{C}_{\alpha\beta\gamma\delta} &= \mathbf{R}_{\alpha\beta\gamma\delta} - \frac{1}{n-1}\left(\mathbf{g}_{\alpha\gamma}\mathbf{R}_{\beta\delta} + \mathbf{g}_{\beta\delta}\mathbf{R}_{\alpha\gamma} - \mathbf{g}_{\beta\gamma}\mathbf{R}_{\alpha\delta} - \mathbf{g}_{\alpha\delta}\mathbf{R}_{\beta\gamma}\right)\\ &+ \frac{1}{n(n-1)}(\mathbf{g}_{\alpha\gamma}\mathbf{g}_{\beta\delta} - \mathbf{g}_{\alpha\delta}\mathbf{g}_{\beta\gamma})\mathbf{R} \end{aligned} \tag{431}$$

Observe that $\mathbf{C}$ verifies all the symmetry properties of the Riemann tensor:

$$\begin{aligned}&\mathbf{C}_{\alpha\beta\gamma\delta} = -\mathbf{C}_{\beta\alpha\gamma\delta} = -\mathbf{C}_{\alpha\beta\delta\gamma} = \mathbf{C}_{\gamma\delta\alpha\beta}\\ &\mathbf{C}_{\alpha\beta\gamma\delta} + \mathbf{C}_{\alpha\gamma\delta\beta} + \mathbf{C}_{\alpha\delta\beta\gamma} = 0\end{aligned} \tag{432}$$

and, in addition, $\quad \mathbf{g}^{\alpha\gamma}\mathbf{C}_{\alpha\beta\gamma\delta} = 0$.

We say that two metrics $\mathbf{g}$ and $\hat{\mathbf{g}}$ are conformal if $\hat{\mathbf{g}} = \lambda^2\mathbf{g}$ for some non zero differentiable function λ. Then the following theorem holds (see Hawking- Ellis, [**?**], chapter 2, section 2.6):

Theorem 3.1. *Let* $\hat{\mathbf{g}} = \lambda^2\mathbf{g}$*,* $\hat{\mathbf{C}}$ *the Weyl tensor relative to* $\hat{\mathbf{g}}$ *and* $\mathbf{C}$ *the Weyl tensor relative to* $\mathbf{g}$*. Then*

$$\hat{\mathbf{C}}^{\alpha}_{\beta\gamma\delta} = \mathbf{C}^{\alpha}_{\beta\gamma\delta} \ .$$

Thus $\mathbf{C}$ *is conformally invariant.*

3.2. Isometries and conformal isometries, Killing and conformal Killing vector fields. Definition. A diffeomorphism $\Phi : \mathcal{U} \subset \mathbf{M} \to \mathbf{M}$ is said to be a conformal isometry if, at every point p, $\Phi_*\mathbf{g} = \Lambda^2\mathbf{g}$, that is,

$$(\Phi^*\mathbf{g})(X,Y)|_p = \mathbf{g}(\Phi_*X, \Phi_*Y)|_{\Phi(p)} = \Lambda^2\mathbf{g}(X,Y)|_p$$

with $\Lambda \neq 0$. If $\Lambda = 1$, Φ is called an isometry of $\mathbf{M}$.

Definition. A vector field K which generates a one parameter group of isometries (respectively, conformal isometries) is called a Killing (respectively, conformal Killing) vector field.

Let K be such a vector field and Φ_t the corresponding one parameter group. Since the $(\Phi_t)_*$ are conformal isometries, we infer that $\mathcal{L}_K\mathbf{g}$ must be proportional to the metric $\mathbf{g}$. Moreover $\mathcal{L}_K\mathbf{g} = 0$ if K is a Killing vector field.

Definition. Given an arbitrary vector field X we denote ${}^{(X)}\pi$ the deformation tensor of X defined by the formula

$${}^{(X)}\pi_{\alpha\beta} = (\mathcal{L}_X g)_{\alpha\beta} = \mathbf{D}_\alpha X_\beta + \mathbf{D}_\beta X_\alpha \ .$$

The tensor ${}^{(X)}\pi$ measures, in a precise sense, how much the diffeomorphism generated by X differs from an isometry or a conformal isometry. The following Proposition holds, (see Hawking-Ellis, citeHawkEll, chapter 2, section 2.6):

Proposition 3.3. *The vector field* X *is Killing if and only if* ${}^{(X)}\pi = 0$*. It is conformal Killing if and only if* ${}^{(X)}\pi$ *is proportional to* $\mathbf{g}$*.*

Remark: One can choose local coordinates such that $X = \frac{\partial}{\partial x^\mu}$. It then immediately follows that, relative to these coordinates the metric $\mathbf{g}$ is independent of the component x^μ.

Proposition 3.4. *On any pseudo-riemannian spacetime* $\mathbf{M}$*, of dimension* $n = p+q$*, there can be no more than* $\frac{1}{2}(p+q)(p+q+1)$ *linearly independent Killing vector fields.*

Proof: Proposition 3.4 is an easy consequence of the following relation, valid for an arbitrary vector field X, obtained by a straightforward computation and the use of the symmetries of $\mathbf{R}$.

$$\mathbf{D}_\beta \mathbf{D}_\alpha X_\lambda = \mathbf{R}_{\lambda\alpha\beta\delta} X^\delta + {}^{(X)}\Gamma_{\alpha\beta\lambda} \tag{433}$$

where

$${}^{(X)}\Gamma_{\alpha\beta\lambda} = \frac{1}{2}\left(\mathbf{D}_\beta \pi_{\alpha\lambda} + \mathbf{D}_\alpha \pi_{\beta\lambda} - \mathbf{D}_\lambda \pi_{\alpha\beta}\right) \tag{434}$$

and $\pi \equiv {}^{(X)}\pi$ is the X deformation tensor.

If X is a Killing vector field equation (433) becomes

$$\mathbf{D}_\beta(\mathbf{D}_\alpha X_\lambda) = \mathbf{R}_{\lambda\alpha\beta\delta} X^\delta \tag{435}$$

and this implies, in view of the theorem of existence and uniqueness for ordinary differential equations, that any Killing vector field is completely determined by the $\frac{1}{2}(n+1)(n+2)$ values of X and $\mathbf{D}X$ at a given point. Indeed let p, q be two points connected by a curve $x(t)$ with tangent vector T. Let $L_{\alpha\beta} \equiv \mathbf{D}_\alpha X_\beta$, Observe that along $x(t)$, X, L verify the system of differential equations

$$\frac{\mathbf{D}}{dt} X = T \cdot L \quad , \quad \frac{\mathbf{D}}{dt} L = \mathbf{R}(\cdot, \cdot, X, T)$$

therefore the values of X, L along the curve are uniquely determined by their values at p.

The n-dimensional Riemannian manifold which possesses the maximum number of Killing vector fields is the Euclidean space $\mathbb{R}^n$. Simmilarily the Minkowski spacetime $\mathbb{R}^{n+1}$ is the Lorentzian manifold with the maximum numbers of Killing vectorfields.

3.5. Laplace-Beltrami operator. The scalar Laplace-Beltrami operator on a pseudo-riemannian manifold $\mathbf{M}$ is defined by,

$$\Delta_{\mathbf{M}} u(x) = g^{\mu\nu} \mathbf{D}_\mu \mathbf{D}_\nu u \tag{436}$$

where u is a scalar function on $\mathbf{M}$. Or, in local coordinates,

$$\Delta_{\mathbf{M}} u(x) = \frac{1}{\sqrt{|g(x)|}} \partial_\mu (g^{\mu\nu} \sqrt{|g(x)|} \partial_\nu) u(x) \tag{437}$$

The Laplace-Beltrami operator is called D'Alembertian in the particular case of a Lorentzian manifold, and is then denoted by $\Box_{\mathbf{M}}$. On any pseudo-riemannian manifold, $\Delta_{\mathbf{M}}$ is symmetric relative to the following scalar product for scalar functions

u, v:

$$(u, v)_{\mathbf{M}} = \int u(x)v(x)dv_{\mathbf{M}}$$

Indeed the following identities are easily established by integration by parts, for any two smooth, compactly supported[8] functions u, v,

$$(-\Delta u, v)_{\mathbf{M}} = \int_{\mathbf{M}} \nabla u \cdot \nabla v \, dv_{\mathbf{M}} = (u, -\Delta v)_{\mathbf{M}} \tag{438}$$

where $\nabla u \cdot \nabla v = g^{ij}\partial_i u \partial_j v$. In the particular case when $u = v$ we derive, $(-\Delta u, v)_{\mathbf{M}} = \int_{\mathbf{M}} |\nabla u|^2$, with $|\nabla u|^2 = \nabla u \cdot \nabla u$. Thus, $-\Delta = -\Delta_{\mathbf{M}}$ is symmetric for functions $u \in \mathcal{C}_0^\infty(\mathbf{M})$. It is positive definite if the manifold $\mathbf{M}$ is Riemannian. This is not the case for Lorentzian manifolds: $\square_{\mathbf{M}}$ is non positive definite.

4. Minkowski space

4.1. Basic definitions. The n+1 dimensional Minkowski space, which we denote by $\mathbb{R}^{n+1}$, consists of the manifold $\mathbb{R}^{n+1}$ together with a Lorentz metric $\mathbf{m}$ and a distinguished system of coordinates x^α, $\alpha = 0, 1, \ldots n$, called inertial, relative to which the metric has the diagonal form $\mathbf{m}_{\alpha\beta} = \mathrm{diag}(-1, 1, \ldots, 1)$. Two inertial systems of coordinates are connected to each other by translations or Lorentz transformations. We use standard geometric conventions of lowering and raising indices relative to $\mathbf{m}$, and its inverse $\mathbf{m}^{-1} = \mathbf{m}$, as well as the usual summation convention over repeated indices. The coordinate vectorfields $\frac{\partial}{\partial x^\alpha}$ are denoted by ∂_α, an arbitrary vectorfield is denoted by $X = X^\alpha \partial_\alpha$ with $X^\alpha = X^\alpha(x^0, \ldots, x^n)$. Observe that by lowering indices relative to $\mathbf{m}$ we get $X_0 = -X^0$ and $X_i = X^i$ for all $i = 1, \ldots, n$. We denote by D the flat covariant derivative of $\mathbb{R}^{n+1}$, that is $D_\alpha \omega_\beta = \partial_\alpha \omega_\beta$ for an arbitrary 1- form $w = \omega_\alpha dx^\alpha$. We also split the spacetime coordinates x^α into the time component $x^0 = t$ and space components $x = x^i, \ldots x^n$. Note that $t_0 = -t$ and $x^i = x_i$ for $i = 1, \ldots, n$.

A vector X is said to be timelike, null or spacelike according to whether $\mathbf{m}(X, X)$ is < 0, $= 0$ or > 0. Accordingly a smooth curve $x^\alpha(s)$ is said to be timelike, null or spacelike if its tangent vector $\frac{dx^\alpha}{ds}$ is timelike, null or spacelike at every one of its points. A causal curve may be timelike or null. Similarly a hypersurface $u(x^0, \ldots x^n) = 0$ is said to be spacelike, null or timelike if its normal $N^\alpha = -\mathbf{m}^{\alpha\beta}\partial_\beta u$ is, respectively, timelike, null or spacelike. The metric induced by $\mathbf{m}$ on a spacelike hypersurface is necessarily positive definite, that is Riemannian. A function $\mathbf{t}(x^0, x^1, \ldots, x^n)$ is said to be a time function if its level hypersurfaces $\mathbf{t} = t$ are spacelike. On a null hypersurface the induced metric is degenerate relative to the normal direction, i.e. $\mathbf{m}(N, N) = 0$. In particular function $\mathbf{u} = \mathbf{u}(x^0, \ldots x^n)$ whose level surfaces $\mathbf{u} = u$ are null must verify the Eikonal equation

$$\mathbf{m}^{\alpha\beta}\partial_\alpha \mathbf{u} \partial_\beta \mathbf{u} = 0 \tag{439}$$

Equation (439) can also be written in the form $D_N N = 0$. We call N a *null geodesic generator* of the level hypersurfaces of $\mathbf{u}$.

[8]This is automatically satisfied if the manifold $\mathbf{M}$ is compact.

A causal curve can be either timelike and null at any of its points. The canonical time orientation of $\mathbb{R}^{n+1}$ is given by the vectorfield $T_0 = \partial_0$. A timelike vector X is said to be *future oriented* if $\mathbf{m}(X, T_0) < 0$ and *past oriented* if $\mathbf{m}(X, T_0) > 0$. The causal future $J^+(S)$ of a set S consists of all points in $\mathbb{R}^{n+1}$ which can be connected to S by a future directed causal curve. The causal past $\mathcal{J}^-(S)$ is defined in the same way. Thus, for a point $p = (t, x)$, $\mathcal{J}^+(p) = \{(t \geq t_0, x)/|x - x_0| \leq t - t_0\}$. Given a smooth domain D, its future set $\mathcal{J}^+(D)$ may, in general, have a nonsmooth boundary, due to caustics.

We consider *conservative domains* $\mathcal{J}^+(D_1) \cap \mathcal{J}^-(D_2)$ with $D_1 \subset \Sigma_1$, $D_2 \subset \Sigma_2$, spacelike hypersurfaces. The domain is *regular* if both D_1, D_2 are regular and its non- spacelike boundaries $\mathcal{N}_1 \subset \partial(\mathcal{J}^+(D_1)) \setminus D_1$ and $\mathcal{N}_2 \subset \partial(\mathcal{J}^-(D_2)) \setminus D_2$ are smooth. In the particular case, when $D_1 = \Sigma_1$ and $D = D_2 \subset \Sigma_2$, we obtain $\mathcal{J}^+(\Sigma_1) \cap \mathcal{J}^-(D)$, called *domain of dependence* of D relative to Σ_1, consisting of all points in the causal past of $D \subset \Sigma_2$, to the future of Σ_1. Similarily $\mathcal{J}^+(D) \cap \mathcal{J}^-(\Sigma_2)$, with $D \subset \Sigma_1$ is called the domain of *dependence of influence* of D relative to Σ_2. Particularly useful examples are given in terms of a time function $\mathbf{t}$ with $\Sigma_1 = \{(t, x)/\mathbf{t}(t, x) = t_1\}$, $\Sigma_2 = \{(t, x)/\mathbf{t}(t, x) = t_1\}$ two, nonintersecting, level hypersurfaces, Σ_2 lying in the future of Σ_1.

A pair of null vectorfields $L, \underline{L}$ form a *null pair* if $\mathbf{m}(L, \underline{L}) = -2$. A null pair $e_n = L, e_{n+1} = \underline{L}$ together with vectorfields $e_1, \ldots e_{n-1}$ such that $\mathbf{m}(L, e_a) = \mathbf{m}(\underline{L}, e_a) = 0$ and $\mathbf{m}(e_a, e_b) = \delta_{ab}$, for all $a, b = 1, \ldots, n-1$, is called a *null frame*. The null pair,

$$\mathbf{L} = \partial_t + \partial_r, \quad \underline{\mathbf{L}} = \partial_t - \partial_r, \tag{440}$$

with $r = |x|$ and $\partial_r = x^i/r\partial_i$, is called *canonical*. Simmilarly a null frame $e_1, \ldots e_{n+1}$ with $e_n = \mathbf{L}, e_{n+1} = \underline{\mathbf{L}}$ is called a canonical null frame. In that case $e_1, \ldots, e_{n-1}$ form, at any point, an orthonormal basis for the the sphere $S_{t,r}$, of constant t and r, passing through that point. Observe also that $\mathbf{L}$ is the null geodesic generator associated to $\mathbf{u} = t - r$ while $\underline{\mathbf{L}}$ the null geodesic of $\underline{u} = t + r$.

4.2. Conformal Killing vectorfields. Let x^μ be an inertial coordinate system of Minkowski space $\mathbb{R}^{n+1}$. The following are all the isometries and conformal isometries of $\mathbb{R}^{n+1}$.

1. Translations: for any given vector $a = (a^0, a^1, ..., a^n) \in \mathbb{R}^{n+1}$,

$$x^\mu \to x^\mu + a^\mu$$

2. Lorentz rotations: Given any $\Lambda = \Lambda^\rho_\sigma \in \mathbf{O}(1, n)$,

$$x^\mu \to \Lambda^\mu_\nu \, x^\nu$$

3. Scalings: Given any real number $\lambda \neq 0$,

$$x^\mu \to \lambda x^\mu$$

4. Inversion: Consider the transformation $x^\mu \to I(x^\mu)$, where

$$I(x^\mu) = \frac{x^\mu}{(x,x)}$$

defined for all points $x \in \mathbb{R}^{n+1}$ such that $(x,x) \neq 0$.

The first two sets of transformations are isometries of $\mathbb{R}^{n+1}$, the group generated by them is called the Poincarè group. The last two type of transformations are conformal isometries. the group generated by all the above transformations is called the Conformal group. In fact the Liouville theorem, whose infinitesimal version will be proved later on, states that it is the group of all the conformal isometries of $\mathbb{R}^{n+1}$.

We next list the Killing and conformal Killing vector fields which generate the above transformations.

i. The generators of translations in the x^μ directions, $\mu = 0, 1, ..., n$:

$$\mathbf{T}_\mu = \frac{\partial}{\partial x^\mu}$$

ii. The generators of the Lorentz rotations in the (μ, ν) plane:

$$\mathbf{L}_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu$$

iii. The generators of the scaling transformations:

$$\mathbf{S} = x^\mu \partial_\mu$$

iv. The generators of the inverted translations [9]:

$$\mathbf{K}_\mu = 2x_\mu x^\rho \frac{\partial}{\partial x^\rho} - (x^\rho x_\rho) \frac{\partial}{\partial x^\mu}$$

We also list below the commutator relations between these vector fields,

$$\begin{aligned}
&[\mathbf{L}_{\alpha\beta}, \mathbf{L}_{\gamma\delta}] = \eta_{\alpha\gamma}\mathbf{L}_{\beta\delta} - \eta_{\beta\gamma}\mathbf{L}_{\alpha\delta} + \eta_{\beta\delta}\mathbf{L}_{\alpha\gamma} - \eta_{\alpha\delta}\mathbf{L}_{\beta\gamma} \\
&[\mathbf{L}_{\alpha\beta}, \mathbf{T}_\gamma] = \eta_{\alpha\gamma}\mathbf{T}_\beta - \eta_{\beta\gamma}\mathbf{T}_\alpha \\
&[\mathbf{T}_\alpha, \mathbf{T}_\beta] = 0 \\
&[\mathbf{T}_\alpha, \mathbf{S}] = \mathbf{T}_\alpha \\
&[\mathbf{T}_\alpha, \mathbf{K}_\beta] = 2(\eta_{\alpha\beta}\mathbf{S} + \mathbf{L}_{\alpha\beta}) \\
&[\mathbf{L}_{\alpha\beta}, \mathbf{S}] = [\mathbf{K}_\alpha, \mathbf{K}_\beta] = 0 \\
&[\mathbf{L}_{\alpha\beta}, \mathbf{K}_\gamma] = \eta_{\alpha\gamma}\mathbf{K}_\beta - \eta_{\beta\gamma}\mathbf{K}_\alpha
\end{aligned} \tag{441}$$

Denoting $\mathcal{P}(1,n)$ the Lie algebra generated by the vector fields $\mathbf{T}_\alpha, \mathbf{L}_{\beta\gamma}$ and $\underline{\mathcal{K}}(1,n)$ the Lie algebra generated by all the vector fields $\mathbf{T}_\alpha, \mathbf{L}_{\beta\gamma}, \mathbf{S}, \mathbf{K}_\delta$ we state the following version of the Liouville theorem,

[9] Observe that the vector fields $\mathbf{K}_\mu$ can be obtained applying I_* to the vector fields $\mathbf{T}_\mu$.

Theorem 4.3. *The following statements hold true.*

1) $\mathcal{P}(1,n)$ is the Lie algebra of all Killing vector fields in $\mathbb{R}^{n+1}$.

2) If $n>1$, $\underline{K}(1,n)$ is the Lie algebra of all conformal Killing vector fields in $\mathbb{R}^{n+1}$.

3) If $n=1$, the set of all conformal Killing vector fields in $\mathbb{R}^{1+1}$ is given by the following expression

$$f(x^0+x^1)(\partial_0+\partial_1)+g(x^0-x^1)(\partial_0-\partial_1)$$

where f,g are arbitrary smooth functions of one variable.

Proof: The proof for part 1 of the theorem follows immediately, as a particular case, from Proposition (3.4). From (433) as $\mathbf{R}=0$ and X is Killing we have

$$D_\mu D_\nu X_\lambda = 0\ .$$

Therefore, there exist constants $a_{\mu\nu}, b_\mu$ such that $X^\mu = a_{\mu\nu}x^\nu + b_\mu$. Since X is Killing $D_\mu X_\nu = -D_\nu X_\mu$ which implies $a_{\mu\nu} = -a_{\nu\mu}$. Consequently X can be written as a linear combination, with real coefficients, of the vector fields $T_\alpha, L_{\beta\gamma}$.

Let now X be a conformal Killing vector field. There exists a function Ω such that

$$^{(X)}\pi_{\rho\sigma} = \Omega\eta_{\rho\sigma} \tag{442}$$

From (433) and (434) it follows that

$$D_\mu D_\nu X_\lambda = \frac{1}{2}\left(\Omega_{,\mu}\eta_{\nu\lambda} + \Omega_{,\nu}\eta_{\mu\lambda} - \Omega_{,\lambda}\eta_{\nu\mu}\right) \tag{443}$$

Taking the trace with respect to μ,ν, on both sides of (443) we infer that

$$\begin{aligned} \Box X_\lambda &= -\frac{n-1}{2}\Omega_\lambda \\ D^\mu X_\mu &= \frac{n+1}{2}\Omega \end{aligned} \tag{444}$$

and applying D^λ to the first equation, $\Box$ to the second one and subtracting we obtain

$$\Box\Omega = 0 \tag{445}$$

Applying D_μ to the first equation of (444) and using (445) we obtain

$$\begin{aligned} (n-1)D_\mu D_\lambda\Omega &= \frac{n-1}{2}(D_\mu D_\lambda\Omega + D_\lambda D_\mu\Omega) = -\Box(D_\mu X_\lambda + D_\lambda X_\mu) \\ &= -(\Box\Omega)\eta_{\mu\lambda} = 0 \end{aligned} \tag{446}$$

Hence for $n\neq 1$, $D_\mu D_\lambda\Omega = 0$. This implies that Ω must be a linear function of x^μ. We can therefore find a linear combination, with constant coefficients, $cS + d^\alpha K_\alpha$ such that the deformation tensor of $X-(cS+d^\alpha K_\alpha)$ must be zero. This is the case because $^{(S)}\pi = 2\eta$ and $^{(K_\mu)}\pi = 4x_\mu\eta$. Therefore $X-(cS+d^\alpha K_\alpha)$ is Killing which, in view of the first part of the theorem, proves the result.

Part 3 can be easily derived by solving (442). Indeed posing $X = a\partial_0 + b\partial_1$, we obtain $2D_0X_0 = -\Omega$, $2D_1X_1 = \Omega$ and $D_0X_1 + D_1X_0 = 0$. Hence a, b verify the system

$$\frac{\partial a}{\partial x^0} = \frac{\partial b}{\partial x^1}\ ,\ \frac{\partial b}{\partial x^0} = \frac{\partial a}{\partial x^1}\ .$$

Hence the one form $adx^0 + bdx^1$ is exact, $adx^0 + bdx^1 = d\phi$, and $\frac{\partial^2 a}{\partial x^{0\,2}} = \frac{\partial^2 b}{\partial x^{1\,2}}$, that is $\Box\phi = 0$. In conclusion

$$X = \frac{1}{2}\left(\frac{\partial\phi}{\partial x^0} + \frac{\partial\phi}{\partial x^1}\right)(\partial_0 + \partial_1) + \frac{1}{2}\left(\frac{\partial\phi}{\partial x^0} - \frac{\partial\phi}{\partial x^1}\right)(\partial_0 - \partial_1)$$

which proves the result.

Remark. Expresse relative to the canonical null pair,

$$\mathbf{T}_0 = 2^{-1}(\mathbf{L} + \underline{\mathbf{L}}), \quad \mathbf{S} = 2^{-1}(\underline{u}\,\mathbf{L} + \mathbf{u}\,\underline{\mathbf{L}}), \quad \mathbf{K}_0 = 2^{-1}(\underline{u}^2\,\mathbf{L} + \mathbf{u}^2\,\underline{\mathbf{L}}). \tag{447}$$

Both $\mathbf{T}_0 = \partial_t$ and $\mathbf{K}_0 = (t^2 + |x|^2)\partial_t + 2tx^i\partial_i$ are causal. This makes them important in deriving energy estimates. Observe that $\mathbf{S}$ is causal only in $\mathcal{J}^+(0) \cup \mathcal{J}^-(0)$.

4.4. Null hypersurfaces. Null hypersurfaces are particularly important as they correspond to the propagation fronts of solutions to the wave or Maxwell equation in Minkowski space[10]. The simplest way to describe the geometry of a null hypersurfaces is to start with a codimension one hypersurface $S_0 \subset \Sigma_0$, where Σ_0 is a fixed spacelike hypersurface of $\mathbb{M}^{n+1}$. At every point $p \in S_0$ there are precisely two null directions ortogonal to the tangent space $T_p(S_0)$. Let L denote a smooth null vectorfield orthogonal to S_0 and consider the congruence of null geodesics[11] generated by the integral curves of L. As long as these null geodesics do not intersect the congruence forms a smooth null hypersurface $\mathcal{N}$. We can also extend L, by parallel transport, to all points of $\mathcal{N}$. Clearly $D_L L = 0$, $\mathbf{m}(L, L) = 0$, moreover $\mathbf{m}(L, X) = 0$ for every vector X tangent to $\mathcal{N}$. Observe also that L is uniquely defined up to multiplication by a conformal factor depending only on S_0. Define, for all vectorfields X, Y tangent to $\mathcal{N}$,

$$\gamma(X, Y) = \mathbf{m}(X, Y), \qquad \chi(X, Y) = \mathbf{m}(D_X L, Y) \tag{448}$$

They are both symmetric tensors, called, respectively, the first and second null fundamental forms of $\mathcal{N}$. Observe that χ is uniquely defined up to the same conformal factor associated to L. Clearly $\gamma(L, X) = \chi(L, X) = 0$ for all X tangent to $\mathcal{N}$, therefore they both depend, at a fixed $p \in \mathcal{N}$, only on a fixed hyperplane transversal to L_p. Define s, called affine parameter, by the condition $L(s) = 1$, $s = 0$ on S_0. Its level surfaces defines the *geodesic foliation* of $\mathcal{N}$. Given coordinates $w = (\omega^a)$, $a = 1, \dots n-1$ on S_0 we can parametrize points on S_s by the flow $x^\mu(s, \omega)$ defined by $\frac{dx^\mu}{ds} = L^\mu$ with $x^\mu(0, \omega)$ the point on S_0 of coordinates w. Let,

$$\gamma_{ab} = \gamma(\frac{\partial}{\partial\omega^a}, \frac{\partial}{\partial\omega^a}), \quad \chi_{ab} = \chi(\frac{\partial}{\partial\omega^a}, \frac{\partial}{\partial\omega^b})$$

---

[10]Or more generally on a Lorentz spacetime.

[11] These are in fact straight lines in Minkowski space.

denote the components of γ and χ relative to these coordinates. One can easily check that $\frac{d}{ds}\gamma_{ab} = 2\chi_{ab}$. The volume element of S_s is given by

$$da_{S_s} = \sqrt{|\gamma|}d\omega^1 \dots dw^{n-1}$$

with γ the determinant of the metric γ. Observe that $\frac{d}{ds}\log|\gamma| = \gamma^{ab}\frac{d}{ds}\gamma_{ab} = 2\mathrm{tr}\chi$, with $\mathrm{tr}\chi = \gamma^{ab}\chi_{ab}$ the *expansion coefficient* of the null hypersurface. Thus,

$$\frac{d}{ds}\sqrt{|\gamma|} = \mathrm{tr}\chi\sqrt{|\gamma|}.$$

The rate of change of the total volume $|S_s|$ is given by the following formula,

$$\frac{d}{ds}|S_s| = \int_{S_s}\mathrm{tr}\chi da_{S_s}. \tag{449}$$

We also remark that χ verifies the following Ricatti type equation,

$$\frac{d}{ds}\chi + \chi^2 = 0 \tag{450}$$

which can be explicitely integrated. Thus one can verify that $\mathrm{tr}\chi(s,\omega_0)$ may become $-\infty$ at a finite value of $s > 0$ if $\mathrm{tr}\chi(0,\omega_0) < 0$ at some point of S_0. This occurence corresponds to the formation of a *caustic*.

An arbitrary foliation S_v on $\mathcal{N}$ can be parametrized by $v(s,\omega)$ with (s,ω) the geodesic coordinates defined above. We call $\Omega = \frac{dv}{ds}$ the null lapse function of the foliation and denote by γ' and χ' the restiction of γ, χ to S_v. If X is a vectorfield tangent to the geodesic foliation S_s then $X' = X - \Omega^{-1}X(v)L$ is tangent to S_v. Thus, if X, Y are tangent to S_s then $\gamma(X,Y) = \gamma(X',Y')$ and $\chi(X',Y') = \chi(X,Y)$. Relative to the coordinates (v,ω) we have

$$\gamma'_{ab} = \gamma_{ab}, \qquad \chi'_{ab} = \chi_{ab}.$$

To define the volume element on a null hypersurface $\mathcal{N}$ we choose an arbitrary foliation v with null lapse function $\frac{dv}{ds} = \Omega$ and induced metric γ and set

$$da_{\mathcal{N}} = \Omega^{-1}da_{S_v}dv \tag{451}$$

where da_{S_v} denotes the area element of S_v induced by γ. The definition does not depend on the particular foliation.

4.5. Energy momentum tensor. An energy momentum tensor in $\mathbb{R}^{n+1}$ is a symmetric two tensor $Q_{\alpha\beta}$ verifying the *positive energy condition*,

$$Q(X,Y) \geq 0,$$

for all X, Y causal, future oriented. We say that Q is *divergenceless* if,

$$D^\beta Q_{\alpha\beta} = 0 \tag{452}$$

Given an arbitrary vectorfield X,

$$D^\alpha(Q_{\alpha\beta}X^\beta) = Q^{\alpha\beta}D_\alpha X_\beta = \frac{1}{2}Q^{\alpha\beta\ (X)}\pi_{\alpha\beta},$$

where ${}^{(X)}\pi = \mathcal{L}_X \mathbf{m}$ denotes the deformation tensor of X. Recall that ${}^{(X)}\pi_{\alpha\beta} = \partial_\alpha X_\beta + \partial_\beta X_\alpha$. In the particular case when X is a Killing vectorfield, that is ${}^{(X)}\pi = 0$, we derive

$$D^\alpha(Q_{\alpha\beta}X^\beta) = 0. \tag{453}$$

The same identity holds if X is conformal Killing and Q is *traceless*, that is $\mathbf{m}^{\alpha\beta}Q_{\alpha\beta} = 0$.

A typical conservation law is obtained when we integrate the latter identity, and apply Stokes theorem, on a *regular conservative* spacetime domain(see section 4.1) $\mathcal{J}^+(D_1) \cap \mathcal{J}^-(D_2)$ with smooth spacelike boundaries $D_i \subset \Sigma_i$ and null boundaries $\mathcal{N}_i$, $i = 1, 2$. We denote by T_1, T_2 the future unit normals to the spacelike hypersurfaces Σ_1, Σ_2 and chose the null normals L_1, L_2 such that $\mathbf{m}(L_i, T_i) = -1$ along the boundaries $D_i \subset \Sigma_i$, $i = 1, 2$. For simplicity we denote both timelike normals by T and both null normals by L whenever there is no possibility of confusion.

Proposition 4.6. *Assume that $Q_{\alpha\beta}$ is a divergenceless energy momentum tensor and X a Killing vectorfield in a neighborhood of the regular conservative domain $\mathcal{J}(D_1, D_2)$ as above. Then,*

$$\int_{\mathcal{N}_2} Q(X, L) + \int_{D_2} Q(X, T) = \int_{\mathcal{N}_1} Q(X, L) + \int_{D_1} Q(X, T) \tag{454}$$

The integrals are taken with respect to the area elements $da_{\mathcal{N}}$ along the null hypersurfaces $\mathcal{N}_1, \mathcal{N}_2$ and the area elements of the Riemannian metrics induced by $\mathbf{m}$ on Σ_1, Σ_2. Observe that all integrands are positive if X is causal. The identity (454) *remains valied if X is conformal Killing and Q is traceless.*

Proof : Let $P_\alpha = Q_{\alpha\beta}X^\beta$. According to eqrefeq:cons-law1 we have $\mathbf{D}^\alpha P_\alpha = 0$. ■

The result simplifies for domains of dependence $\mathcal{J}^+(\Sigma_1) \cap \mathcal{J}^-(D \subset \Sigma_2)$, or influence $\mathcal{J}^+(D \subset \Sigma_1) \cap \mathcal{J}^-(\Sigma_2)$, with Σ_2 in the future of Σ_1. We normalize L by the condition $\mathbf{m}(L, T) = -1$ on $\partial D \subset \Sigma_2$ where T denotes the unit normal to Σ_1, Σ_2.

Corollary 4.7. *If Q is divergenceless, X is Killing and $D \subset \Sigma_2$,*

$$\int_{\mathcal{N}} Q(X, L) + \int_{D \subset \Sigma_2} Q(X, T) = \int_{\mathcal{J}^-(D) \cap \Sigma_1} Q(X, T) \tag{455}$$

Similarily, if $D \subset \Sigma_1$,

$$\int_{\mathcal{N}} Q(X, L) + \int_{D \subset \Sigma_1} Q(X, T) = \int_{\mathcal{J}^+(D) \cap \Sigma_2} Q(X, T) \tag{456}$$

The identity remains true if X is conformal Killing and Q is traceless.

Bibliography

[E] L.C. Evans, *Partial Differential Equations*, Providence: American Mathematical Society, 1998

[Fr-Io] F.G. Friedlander and M. Ioshi *An introduction to the theory of distributions*, Cambridge University Press, second edition, 1998.

[G-T] D. Gilbarg and N. Trudinger, *Elliptic partial differential equations of second order*, Springer Verlag, second edition,1983.

[Ge-S] I.M. Gelfand and G.E. Shilov *Generalized Functions Volume I: Properties and Operations*. Academic Press, 1964.

[Ho] L. Hormander, *The Analysis of linear partial differential operators* vols 1-4, Springer Verlag 1983, 1985.

[H-E] S.W. Hawking, G.F.R. Ellis, *The Large Scale Structure of Space-Time* Cambridge University Press, 1973.

[J] F. John, *Partial Differential Equations*, Providence: American Mathematical Society, fourth edition, 1982.

[K-T] Markus Keel and Terrence Tao, "Endpoint Strichartz Estimates", *Am. J. Math.* 120 (1998), 955 980.

[Kl-Rodn2] S. Klainerman and I. Rodnianski, *Causal geometry of Einstein-vacuum spacetimes with finite curvature flux*, Invent. Math. **159** (2005), 437–529.

[Kl-Rodn2] S. Klainerman and I. Rodnianski, *A geometric approach to the Littlewood-Paley theory*, Geom. Funct. Anal. **16** (2006), 126–163.

[Kl-Rodn3] S. Klainerman and I. Rodnianski, *Sharp trace theorems for null hypersurfaces on Einstein metrics with finite curvature flux*, Geom. Funct. Anal. **16** (2006), 164–229.

[T] M.E. Taylor, *Partial Differential Equations*, Springer Verlag, texts in applied mathematics, vol 23, 1996.