Gregory Naber

Classical Klein-Gordon Fields

Lagrangian and Hamiltonian Formulations

This is for my grandchildren Amber, Emily, Garrett, and Lacey and my buddy Vinnie.

Preface

In these two chapters we take the next step toward our eventual goal of describing a mathematically rigorous model of the quantized Klein-Gordon field. It will come as no surprise that one does not construct the quantization of a classical physical system without understanding that classical system rather well and that, if the quantization is to lay any claims to mathematical rigor, the same must be true of the classical system. Classical mechanical systems admit rigorous descriptions within the context of symplectic geometry and here we will extend this Hamiltonian view of mechanics to classical scalar field theory and, more particularly, to the Klein-Gordon field. This extension requires a rather substantial amping-up of the mathematical pre-requisites, due in large measure to the fact that the phase space is necessarily infinite-dimensional. We will try to address these pre-requisites in Chapter 1 by briefly reviewing the differential calculus on Banach spaces and the Euler-Lagrange equations for real-valued functions on Banach spaces and also in two Appendices. Appendix A provides a synopsis of tempered distributions, Sobolev spaces, and Fourier transforms, while Appendix B contains a brief summary of those parts of the Hille-Yosida theory of semigroups of operators that we will need to call upon.

The heart of the material is in Chapter 2. Here we view the Klein-Gordon equation as the Euler-Lagrange equation for a certain Lagrangian on Minkowski spacetime and carefully discuss the sense in which it is relativistically invariant (Section 2.1.1). A version of Noether's Theorem is derived with which one can write out a number of associated conservation laws for Klein-Gordon fields (Section 2.1.2). In Section 2.2 we go to some lengths to carefully derive both smooth and distributional solutions to the Klein-Gordon equation and, in particular, to trace the origin of formulas one finds in the physics literature such as

$$\varphi(t,\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} \frac{1}{2\omega_{\mathbf{p}}} \left(e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\, a(\mathbf{p}) + e^{-i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\, \overline{a}(\mathbf{p}) \right) d^3\mathbf{p}.$$

Real and complex solutions are treated separately because they have different physical interpretations and we try to explain why this is the case in Section 2.3. Finally, in Section 2.4 we motivate and then define infinite-dimensional Hamiltonian systems and show that Klein-Gordon can be interpreted as such a system.

Contents

Chapter 1
Classical Scalar Fields

1.1 Differential Calculus on Banach Spaces

In this section we will lay the foundation required to extend the Lagrangian and Hamiltonian pictures of classical mechanics (Sections 2.2 and 2.3 of [Nab5] or Appendices A.2 and A.3 of [Nab6]) to field theory. Our basic reference for this material is Chapters 2 and 6 of [AMR], but there are many other sources as well, for example, [Lang4] and [Car2]. Since most of the results and many of the proofs are very much like their finite-dimensional counterparts (as in, say, [Sp1]), we will be relatively brief. A concise review of the required material on Banach spaces is available in Sections 2.1 and 2.2 of [AMR].

We begin by establishing some notation and terminology. $\mathcal{E}$ and $\mathcal{F}$ are Banach spaces (both over $\mathbb{R}$ or both over $\mathbb{C}$), $U \subseteq \mathcal{E}$ is an open set, $f : U \subseteq \mathcal{E} \to \mathcal{F}$ a mapping, and $u_0 \in U$. Norms on the Banach spaces will be denoted $\| \ \|_{\mathcal{E}}, \| \ \|_{\mathcal{F}}$, or simply $\| \ \|$ if this will cause no confusion. Then f is *(Fréchet) differentiable at* u_0 if there is a necessarily unique (see page 68 of [AMR]) bounded linear map $Df(u_0) : \mathcal{E} \to \mathcal{F}$ such that, for every $\epsilon > 0$ there exists a $\delta > 0$ such that $0 < \|u - u_0\|_{\mathcal{E}} < \delta$ implies

$$\frac{\| f(u) - f(u_0) - Df(u_0) \cdot (u - u_0) \|_{\mathcal{F}}}{\| u - u_0 \|_{\mathcal{E}}} < \epsilon,$$

where we use $Df(u_0) \cdot (u - u_0)$ to indicate the value of $Df(u_0)$ at $u - u_0$. In other words,

$$\lim_{u \to u_0} \frac{f(u) - f(u_0) - Df(u_0) \cdot (u - u_0)}{\| u - u_0 \|_{\mathcal{E}}} = 0.$$

In this case, $Df(u_0)$ is called the *(Fréchet) derivative of* f *at* u_0. Not unexpectedly, the Fréchet derivative of a bounded linear operator is that same linear operator at each point.

Exercise 1.1.1. Let $f : \mathcal{E} \to \mathcal{F}$ be a bounded linear operator and let $u_0 \in \mathcal{E}$ be arbitrary. Prove that $Df(u_0) = f$.

To get something a bit more interesting we will compute the Fréchet derivative of a nonlinear integral operator.

Example 1.1.1. We let $\mathcal{E} = \mathcal{F} = C^0[a,b]$ be the Banach space of continuous (real- or complex-valued) functions on $[a,b]$ with the sup-norm $\|\ \| = \|\ \|_\infty$ so that, for any $u \in C^0[a,b]$, $\|u\| = \max_{a\leq x\leq b}|u(x)|$. Let $K : [a,b]\times[a,b] \to \mathbb{R}$ be an arbitrary continuous function. Define $f : C^0[a,b] \to C^0[a,b]$ by

$$f(u)(x) = u(x)\int_a^b K(x,t)\,u(t)\,dt$$

for all $u \in C[a,b]$ and all $x \in [a,b]$. Thus, for any fixed $u_0 \in C^0[a,b]$,

$$f(u)(x) - f(u_0)(x) = u(x)\int_a^b K(x,t)\,u(t)\,dt - u_0(x)\int_a^b K(x,t)\,u_0(t)\,dt.$$

Now write this as

$$\begin{aligned} f(u)(x) - f(u_0)(x) - \Big[u_0(x)\int_a^b K(x,t)\,(u(t)-u_0(t))\,dt \\ + (u(x)-u_0(x))\int_a^b K(x,t)\,u_0(t)\,dt\Big] \\ = (u(x)-u_0(x))\int_a^b K(x,t)\,(u(t)-u_0(t))\,dt. \end{aligned} \tag{1.1}$$

But if we let M be the maximum value of $K(x,t)$ on $[a,b]\times[a,b]$ then

$$\left|(u(x)-u_0(x))\int_a^b K(x,t)\,(u(t)-u_0(t))\,dt\right| \leq M\,(b-a)\,\|u-u_0\|^2.$$

Since

$$\frac{M\,(b-a)\,\|u-u_0\|^2}{\|u-u_0\|} = M\,(b-a)\,\|u-u_0\| \to 0$$

as $u \to u_0$ in $C^0[a,b]$ and since

$$u_0(x)\int_a^b K(x,t)\,v(t)\,dt + v(x)\int_a^b K(x,t)\,u_0(t)\,dt$$

is linear in v, we conclude from (1.1) that the value of $Df(u_0)$ at any displacement vector $u - u_0$ is given, at each $x \in [a,b]$, by

$$
\begin{aligned}
[Df(u_0)\cdot(u-u_0)](x) =& u_0(x)\int_a^b K(x,t)\,(u(t)-u_0(t))\,dt \\
&+ (u(x)-u_0(x))\int_a^b K(x,t)\,u_0(t)\,dt. \qquad (1.2)
\end{aligned}
$$

Now let's get back to the general development. If f is differentiable at every $u \in U$, then it defines a map

$$Df : U \to L(\mathcal{E}, \mathcal{F})$$

from U to the Banach space $L(\mathcal{E}, \mathcal{F})$ of bounded linear maps from $\mathcal{E}$ to $\mathcal{F}$ given by

$$u \to Df(u).$$

Df is then called the *(Fréchet) derivative of f on U* and f is said to be *(Fréchet) differentiable on U*.

Remark 1.1.1. Recall that the norm of an element $A \in L(\mathcal{E}, \mathcal{F})$ is defined by

$$\|A\|_{L(\mathcal{E},\mathcal{F})} = \sup\left\{ \frac{\|Ae\|_{\mathcal{F}}}{\|e\|_{\mathcal{E}}} \; : e \in \mathcal{E}, e \neq 0 \right\} = \sup\{\, \|Ae\|_{\mathcal{F}} : e \in \mathcal{E}, \|e\|_{\mathcal{E}} = 1\}$$

(Proposition 2.2.4 of [AMR]).

In this case, f is necessarily continuous and, in fact, locally Lipschitz on U (that is, for each u_0 in U there exist constants $M_0 > 0$ and $\delta_0 > 0$ such that $u \in U$ and $\|u - u_0\|_{\mathcal{E}} < \delta_0$ implies $\|f(u) - f(u_0)\|_{\mathcal{F}} \leq M_0 \|u - u_0\|_{\mathcal{E}}$); see Proposition 2.4.1 of [AMR]. If the space $L(\mathcal{E}, \mathcal{F})$ is given its norm topology and if Df is a continuous map, then f is said to be *continuously differentiable* or C^1 on U.

Remark 1.1.2. Unless it is likely to cause some confusion we will henceforth tend to drop the subscripts on $\| \, \|$ and leave it to the context to indicate which Banach space is intended.

One defines higher order (Fréchet) derivatives in the following way. First recall that if $\mathcal{E}_1$ and $\mathcal{E}_2$ are two Banach spaces, then the *product* $\mathcal{E}_1 \times \mathcal{E}_2$ of $\mathcal{E}_1$ and $\mathcal{E}_2$ is the Banach space consisting of all ordered pairs (e_1, e_2) of elements $e_i \in \mathcal{E}_i, i = 1, 2$, with norm $\|(e_1, e_2)\|^2 = \|e_1\|^2 + \|e_2\|^2$. $\mathcal{E}_1 \times \mathcal{E}_2$ is also often called the *direct sum* of $\mathcal{E}_1$ and $\mathcal{E}_2$ and denoted $\mathcal{E}_1 \oplus \mathcal{E}_2$. Larger (finite) products/sums are defined in the obvious way by induction. A map $M : \mathcal{E}_1 \times \cdots \times \mathcal{E}_k \to \mathcal{F}$ is *k-multilinear* if it is linear in each variable separately and the linear space of all such is a Banach space with norm defined by

$$
\begin{aligned}
\|M\| &= \sup\left\{ \frac{\|M(e_1, \ldots, e_k)\|}{\|e_1\| \cdots \|e_k\|} : e_1, \ldots, e_k \neq 0 \right\} \\
&= \sup\{\, \|M(e_1, \cdots, e_k)\| : \|e_1\| = \cdots \|e_k\| = 1\,\}.
\end{aligned}
$$

If $\mathcal{E}_1 = \cdots = \mathcal{E}_k = \mathcal{E}$ we will write this Banach space $L^k(\mathcal{E}, \mathcal{F})$.

Now observe that there is a natural isometric isomorphism from the Banach space $L(\mathcal{E}, L(\mathcal{E}, \mathcal{F}))$ of bounded linear maps from $\mathcal{E}$ to $L(\mathcal{E}, \mathcal{F})$ onto the Banach space $L^2(\mathcal{E}, \mathcal{F})$ of bounded bilinear maps $\mathcal{E} \times \mathcal{E} \to \mathcal{F}$ given by

$$A \in L(\mathcal{E}, L(\mathcal{E}, \mathcal{F})) \to \tilde{A} \in L^2(\mathcal{E}, \mathcal{F}),$$

where

$$\tilde{A}(e_1, e_2) = (A(e_1))(e_2)$$

(Proposition 2.2.9 of [AMR]).

Now suppose that $Df : U \subseteq \mathcal{E} \to L(\mathcal{E}, \mathcal{F})$ is differentiable at every $u \in U$. Then $D(Df) : U \to L(\mathcal{E}, L(\mathcal{E}, \mathcal{F}))$. Identifying $L(\mathcal{E}, L(\mathcal{E}, \mathcal{F}))$ with $L^2(\mathcal{E}, \mathcal{F})$ we obtain the *second (Fréchet) derivative of f*, denoted

$$D^2 f : U \subseteq \mathcal{E} \to L^2(\mathcal{E}, \mathcal{F}).$$

Exercise 1.1.2. Let $f : \mathcal{E} \to \mathcal{F}$ be a bounded linear operator. Show that

$$(D^2 f(e))(e_1, e_2) = f(e_2)$$

for all $e \in \mathcal{E}$ and all $(e_1, e_2) \in \mathcal{E} \times \mathcal{E}$.

If $D^2 f$ exists and is norm continuous, then f is *twice continuously differentiable* or C^2 on U. Continuing inductively we obtain, for any $k \geq 2$,

$$D^k f = D(D^{k-1} f) : U \subseteq \mathcal{E} \to L^k(\mathcal{E}, \mathcal{F})$$

if it exists (here $D^1 = D$). If $D^k f$ exists and is norm continuous, then f is said to be *k-times continuously differentiable* or C^k on U. By convention, $f : U \subseteq \mathcal{E} \to \mathcal{F}$ is C^0 if it is continuous on U. If f is C^k for every $k \geq 0$, then it is C^∞, or *smooth*.

As in the finite-dimensional case, Fréchet derivatives are often more conveniently thought of in terms of directional derivatives. The following is Proposition 2.4.6 of [AMR].

Theorem 1.1.1. *If $f : U \subseteq \mathcal{E} \to \mathcal{F}$ is differentiable at $u \in U$, then, for any $e \in \mathcal{E}$, the* ***directional derivative of*** *f* ***at*** *u* ***in the direction*** *e, defined by*

$$\frac{d}{d\varepsilon} f(u + \varepsilon e)\Big|_{\varepsilon=0} = \lim_{\varepsilon \to 0} \frac{f(u + \varepsilon e) - f(u)}{\varepsilon}, \tag{1.3}$$

exists and is given by

$$\frac{d}{d\varepsilon} f(u + \varepsilon e)\Big|_{\varepsilon=0} = Df(u) \cdot e. \tag{1.4}$$

Exercise 1.1.3. Compute $\frac{d}{d\varepsilon} f(u_0+\varepsilon(u-u_0))\Big|_{\varepsilon=0}$ for the integral operator in Example 1.1.1 and thereby obtain another proof of (1.2).

Thus, a function that is differentiable at u has directional derivatives in every direction at u. The converse is not true, however. Counterexamples exist even in elementary calculus. Exercise 2.4-9 of [AMR] offers the following. The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \frac{2x^2y}{x^4+y^2} & \text{if } (x,y) \neq (0,0); \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is continuous everywhere except the origin so it is not Fréchet differentiable at $(0,0)$, but all of its directional derivatives exist at $(0,0)$.

Exercise 1.1.4. If you have never checked anything like this before, do so now.

A function $f : U \subseteq \mathcal{E} \to \mathcal{F}$ for which the directional derivative $\frac{d}{d\varepsilon} f(u+\varepsilon e)\Big|_{\varepsilon=0}$ exists for every $e \in \mathcal{E}$ is said to be *Gâteaux differentiable at u*. If f is differentiable at u, then it is Gâteaux differentiable at u, but not conversely.

Let $\mathcal{E}_1, \mathcal{E}_2$ and $\mathcal{F}$ be Banach spaces (all real or all complex), U an open set in $\mathcal{E}_1 \times \mathcal{E}_2$, $f : U \subseteq \mathcal{E}_1 \times \mathcal{E}_2 \to \mathcal{F}$ a mapping and $u_0 = (u_{01}, u_{02})$ a point in U. Then the *partial derivatives* $D_1 f(u_0) \in L(\mathcal{E}_1, \mathcal{F})$ and $D_2 f(u_0) \in L(\mathcal{E}_2, \mathcal{F})$ are the derivatives of the maps $x \mapsto f(x, u_{02})$ and $y \mapsto f(u_{01}, y)$ at u_{01} and u_{02}, respectively, provided they exist. The following is Proposition 2.4.12 of [AMR].

Theorem 1.1.2. *If $f : U \subseteq \mathcal{E}_1 \times \mathcal{E}_2 \to \mathcal{F}$ is differentiable on U, then, for every $u \in U$, the partial derivatives $D_1 f(u)$ and $D_2 f(u)$ exist. Furthermore*

1. *For every $(e_1, e_2) \in \mathcal{E}_1 \times \mathcal{E}_2$ and every $u \in U$,*

$$D_1 f(u) \cdot e_1 = Df(u) \cdot (e_1, 0),$$

$$D_2 f(u) \cdot e_2 = Df(u) \cdot (0, e_2),$$

and

$$Df(u) \cdot (e_1, e_2) = D_1 f(u) \cdot e_1 + D_2 f(u) \cdot e_2.$$

2. *f is C^k on U if and only if $D_i f$, $i = 1, 2$, both exist and are C^{k-1} on their domains.*

More variables and higher orders are handled by induction. Next we will list a few of the expected properties of derivatives, all of which are proved in Sections 2.4 and 2.5 of [AMR].

Theorem 1.1.3. *Suppose $f, g : U \subseteq \mathcal{E} \to \mathcal{F}$ are both k-times differentiable and a is in $\mathbb{R}$ (or $\mathbb{C}$ if the Banach spaces are complex). Then $af, f+g : U \subseteq \mathcal{E} \to \mathcal{F}$ are k times differentiable with $D^k(af) = aD^k f$ and $D^k(f+g) = D^k f + D^k g$.*

Theorem 1.1.4. *Suppose* $f_i : U \subseteq \mathcal{E} \to \mathcal{F}_i, i = 1, \ldots, n$, *are all k times differentiable. Then* $f = f_1 \times \cdots \times f_n : U \subseteq \mathcal{E} \to \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, *defined by* $f(u) = (f_1(u), \ldots, f_n(u))$ *for all* $u \in U$, *is k times differentiable and*

$$D^k f = D^k f_1 \times \cdots \times D^k f_n.$$

Theorem 1.1.5. *(Chain Rule) Let* $f : U \subseteq \mathcal{E} \to V \subseteq \mathcal{F}$ *and* $g : V \subseteq \mathcal{F} \to \mathcal{G}$ *be differentiable on the open subsets U and V of* $\mathcal{E}$ *and* $\mathcal{F}$, *respectively. Then* $g \circ f : U \subseteq \mathcal{E} \to \mathcal{G}$ *is differentiable on U and* $D(g \circ f)(u) = Dg(f(u)) \circ Df(u)$.

Remark 1.1.3. For the statement of the next two results (Theorems 2.5.2 and 2.5.7 of [AMR]) we recall that an *isomorphism* of Banach spaces is a linear isomorphism that is also a homeomorphism; a linear isomorphism that preserves norms will be called an *isometric isomorphism*. According to the Open Mapping Theorem (Theorem 2.2.15 of [AMR] or Theorem 4.6.1 of [Fried]) a continuous, injective, linear map from one Banach space $\mathcal{E}$ to another $\mathcal{F}$ is a homeomorphism if and only if it is surjective.

Theorem 1.1.6. *(Inverse Function Theorem) Suppose* $f : U \subseteq \mathcal{E} \to \mathcal{F}$ *is* $C^k, k \geq 1, u_0 \in U$ *and* $Df(u_0)$ *is a Banach space isomorphism of* $\mathcal{E}$ *onto* $\mathcal{F}$. *Then f is a* C^k*-diffeomorphism (*C^k*-bijection with a* C^k*-inverse) of some neighborhood* U_0 *of* u_0 *onto the neighborhood* $f(U_0)$ *of* $f(u_0)$ *and, moreover, the derivative of the inverse map is given by*

$$Df^{-1}(y) = [Df(f^{-1}(y))]^{-1}.$$

for every $y \in f(U_0)$.

Theorem 1.1.7. *(Implicit Function Theorem) Let* $U \subseteq \mathcal{E}$ *and* $V \subseteq \mathcal{F}$ *be open sets and* $f : U \times V \subseteq \mathcal{E} \times \mathcal{F} \to \mathcal{G}$ *a* C^k *map for some* $k \geq 1$. *Suppose* $(u_0, v_0) \in U \times V$ *and* $D_2 f(u_0, v_0) : \mathcal{F} \to \mathcal{G}$ *is a Banach space isomorphism. Then there exist neighborhoods* U_0 *of* u_0 *and* W_0 *of* $f(u_0, v_0)$ *and a unique* C^k *map* $g : U_0 \times W_0 \to V$ *such that, for all* $(u, w) \in U_0 \times W_0$,

$$f(u, g(u, w)) = w.$$

Remark 1.1.4. We mention in passing that there is a much more subtle version of the Inverse Function Theorem for (certain) Fréchet manifolds due to Nash which he used to prove his famous Isometric Embedding Theorem for Riemannian manifolds. There is a very detailed exposition of this result together with all of the prerequisite material on Fréchet spaces and Fréchet manifolds in [Ham].

The theory of differential forms on Banach spaces is virtually identical to the more familiar theory on finite-dimensional vector spaces. Because we restrict our attention to linear problems we will not require the, also fairly routine, extension to Banach manifolds. Our basic references for this material are Chapter 7 of [AMR] and the book [Car1] of Cartan. We will provide just a brief synopsis to establish our notation; needless to say, $\mathbb{R}^n$ is a real Banach space so everything we have to say is equally true in the finite-dimensional case.

We let $\mathcal{E}$ denote a real Banach space. For each $e \in \mathcal{E}$ we will canonically identify the tangent space $T_e(\mathcal{E})$ to $\mathcal{E}$ at e with $\mathcal{E}$ itself; specifically, each $v_e \in T_e(\mathcal{E})$ is uniquely expressible as $\frac{d}{dt}(e + tv)|_{t=0}$ for some $v \in \mathcal{E}$ and we identify $T_e(\mathcal{E})$ and $\mathcal{E}$ by the isomorphism $v_e \leftrightarrow v$. Consequently, a vector field on an open subset U of $\mathcal{E}$ can be identified with a map from U to $\mathcal{E}$; the vector field is C^k for some $k = 0, 1, \ldots, \infty$ if this map is C^k. $\mathcal{E}^*$ will denote the dual (or conjugate) of $\mathcal{E}$, that is, the Banach space of all bounded linear functionals $\alpha : \mathcal{E} \to \mathbb{R}$ on $\mathcal{E}$ with its usual norm $\|\alpha\| = \sup_{\|e\|=1} |\alpha(e)|$. For $k \geq 2$ we denote by $\mathcal{A}^k(\mathcal{E})$ the Banach space (closed subspace of $L^k(\mathcal{E}, \mathbb{R})$) of all bounded k-multilinear real-valued functions $\omega : \mathcal{E}^k = \mathcal{E} \times \overset{k}{\cdots} \times \mathcal{E} \to \mathbb{R}$ that are alternating, that is, satisfy

$$\omega(e_{\sigma(1)} \ldots, e_{\sigma(k)}) = sgn(\sigma)\, \omega(e_1, \ldots, e_k)$$

for every σ in the symmetric group S_k of permutations on $\{1, \ldots, k\}$; here $sgn(\sigma)$ is 1 if σ is an even permutation and -1 if σ is odd. It is convenient to take $\mathcal{A}^0(\mathcal{E}) = \mathbb{R}$ and $\mathcal{A}^1(\mathcal{E}) = \mathcal{E}^*$ as well.

Now, if U is an open subset of $\mathcal{E}$, then a *(differential) k-form* on U is a mapping

$$\omega : U \to \mathcal{A}^k(\mathcal{E})$$

that assigns to each $p \in U$ an element $\omega(p) = \omega_p$ of $\mathcal{A}^k(\mathcal{E})$. One can consider k-forms with any degree of differentiability, but unless otherwise specified, we will generally restrict attention to those that are smooth (C^∞). Thus, for example, a 0-form on U is just a smooth, real-valued function on U. The vector space of all smooth k-forms on U will be denoted $\Omega^k(U)$. For any $k, l = 0, 1, \ldots$ we define a bilinear map

$$\wedge : \Omega^k(U) \times \Omega^l(U) \to \Omega^{k+l}(U),$$

called the *wedge product* (or *exterior product*) as follows. If $\omega \in \Omega^k(U)$ and $\eta \in \Omega^l(U)$ with $k, l \geq 1$, then $\omega \wedge \eta \in \Omega^{k+l}(U)$ is defined, at each $p \in U$, by

$$(\omega \wedge \eta)_p(e_1, \ldots, e_{k+l}) = \frac{1}{k!l!} \sum_{\sigma \in S_{k+l}} sgn(\sigma)\omega_p(e_{\sigma(1)}, \ldots, e_{\sigma(k)})\, \eta_p(e_{\sigma(k+1)} \ldots, e_{\sigma(k+l)}).$$

If $\omega = f \in \Omega^0(U)$ and $\eta \in \Omega^l(U)$, then $f \wedge \eta \in \Omega^l(U)$ is defined to be the pointwise product of f and η, that is,

$$(f \wedge \eta)_p(e_1, \ldots, e_l) = f(p)\eta_p(e_1, \ldots, e_l)$$

and similarly if $\omega \in \Omega^k(U)$ and $\eta = f \in \Omega^0(U)$. If ω and η are both 0-forms, then there wedge products is just the pointwise product of these smooth, real-valued functions.

Remark 1.1.5. In (7.1.1) of [AMR] one finds $(\omega \wedge \eta)_p(e_1, \dots, e_{k+l})$ expressed without the factor $\frac{1}{k!l!}$ and with the sum over only the *shuffle permutations* of $\{1, \dots, k, k+1 \dots, k+l\}$. This is sometimes more convenient for explicit computations because there are fewer shuffle permutations, but it is equivalent to the somewhat more common definition we have given above.

Note that, if ω and η are both 1-forms on U, then

$$(\omega \wedge \eta)_p(e_1, e_2) = \omega_p(e_1)\eta_p(e_2) - \omega_p(e_2)\eta_p(e_1).$$

Just as in the finite-dimensional case one shows that the wedge product is associative

$$(\omega \wedge \eta) \wedge \tau = \omega \wedge (\eta \wedge \tau)$$

and graded commutative, that is, if $\omega \in \Omega^k(U)$ and $\eta \in \Omega^l(U)$, then

$$\eta \wedge \omega = (-1)^{kl} \omega \wedge \eta$$

(see Proposition 7.1.5 of [AMR]).

If $F : U \to V$ is a smooth map from the open set U in the Banach space $\mathcal{E}$ to the open set V in the Banach space $\mathcal{F}$ and ω is a k-form on V, then we define a k-form $F^*\omega$ on U, called the *pullback* of ω by F, by

$$(F^*\omega)_p(e_1, \dots, e_k) = \omega_{F(p)}(DF(p) \cdot e_1, \dots, DF(p) \cdot e_k)$$

for each $p \in U$ and $e_1, \dots, e_k \in \mathcal{E}$. Thus, $F^* : \Omega^k(V) \to \Omega^k(U)$ carries k-forms to k-forms. On the other hand, given a vector field X on U and a k-form ω on U with $k \geq 1$ we define a $(k-1)$-form $\iota_X \omega$ on U, called the *contraction* of ω with X, by

$$(\iota_X \omega)_p(e_1, \dots, e_{k-1}) = \omega_p(X(p), e_1, \dots, e_{k-1})$$

for each $p \in U$ and $e_1, \dots, e_{k-1} \in \mathcal{E}$. It is customary to define $\Omega^{-1}(U)$ to be the real vector space consisting of only the zero element and to define the contraction $\iota_X f$ of a 0-form f to be zero. Pullback commutes with the wedge product (Proposition 7.3.10 (v) of [AMR])

$$F^*(\omega \wedge \eta) = F^*\omega \wedge F^*\eta,$$

and ι_X satisfies a graded product rule with respect to the wedge product (Proposition 7.4.8 (i) of [AMR]) , that is,

$$\iota_X(\omega \wedge \eta) = \iota_X \omega \wedge \eta + (-1)^k \omega \wedge \iota_X \eta,$$

where $\omega \in \Omega^k(U)$.

Just as for a finite-dimensional manifold, a Banach space has an intrinsic notion of differentiation for forms defined on it. This is called *exterior differentiation*, it carries k-forms to $(k+1)$-forms and is denoted

$$d : \Omega^k(U) \to \Omega^{k+1}(U).$$

Remark 1.1.6. It would be more proper to write $d^k : \Omega^k(U) \to \Omega^{k+1}(U)$ since the maps do depend on the degree of the form to which they are being applied, but it has become customary to drop the k and call them all d.

It is defined in the following way. Let $\omega \in \Omega^k(U)$. Regarding ω as a smooth map from the open set U in the Banach space $\mathcal{E}$ to the Banach space $\mathcal{A}^k(\mathcal{E})$, it has a Fréchet derivative $D\omega$. At each $p \in U$, $D\omega(p) : \mathcal{E} \to \mathcal{A}^k(\mathcal{E})$ so each $D\omega(p) \cdot e$ is an alternating k-multilinear form on $\mathcal{E}$. The *exterior derivative* of ω is the $(k+1)$-form $d\omega$ defined by

$$d\omega_p(e_0, e_1, \ldots, e_k) = \sum_{i=0}^{k} (-1)^i (D\omega(p) \cdot e_i)(e_0, \ldots, \hat{e}_i, \ldots, e_k),$$

where the hat ˆ indicates that e_i is missing.

Example 1.1.2. If f is a 0-form on U (that is, a smooth, real-valued function on U), then, for each $p \in U$, $Df(p) : \mathcal{E} \to \mathbb{R}$ is given by

$$Df(p) \cdot e = \frac{d}{d\varepsilon} f(p + \varepsilon e)|_{\varepsilon=0}.$$

Since there is only one term in the sum defining it, $df_p(e_0) = Df(p) \cdot e_0$. Thus, df is the 1-form on U which, at any $p \in U$, sends $e \in \mathcal{E}$ to the directional derivative of f at p in the direction e.

$$df_p(e) = Df(p) \cdot e = \frac{d}{d\varepsilon} f(p + \varepsilon e)|_{\varepsilon=0}$$

Example 1.1.3. Let ω be a 1-form on U. Then $\omega : U \to \mathcal{A}^1(\mathcal{E}) = \mathcal{E}^*$ so, for each $p \in U$, $D\omega(p) : \mathcal{E} \to \mathcal{E}^*$. Then

$$\begin{aligned} d\omega_p(e_0, e_1) &= \sum_{i=0}^{1} (-1)^i (D\omega(p) \cdot e_i)(e_0, \ldots, \hat{e}_i, \ldots, e_1) \\ &= (D\omega(p) \cdot e_0)(e_1) - (D\omega(p) \cdot e_1)(e_0). \end{aligned}$$

Just as in the finite-dimensional case one shows that the exterior differentiation operator d has the following properties and is, in fact, characterized by them.

1. $d : \Omega^k(U) \to \Omega^{k+1}(U)$ is linear for each $k = 0, 1, \ldots$.
2. $d^2 = 0$, that is, the composition

$$\Omega^k(U) \xrightarrow{d} \Omega^{k+1}(U) \xrightarrow{d} \Omega^{k+2}(U)$$

   is identically zero for every $k = 0, 1, \ldots$.
3. If $\omega \in \Omega^k(U)$ and $\eta \in \Omega^l(U)$ for any $k, l = 0, 1, \ldots$, then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

Exercise 1.1.5. Verify the following special cases of these properties.

1. Show that $d^2 = 0$ on 0-forms, that is, $d(df) = 0$ for every $f \in \Omega^0(U)$.
2. Write out explicitly the exterior derivative $d\eta$ of a 2-form η and then show that $d^2 = 0$ on 1-forms.
3. Show that, if ω is a 1-form and η is a 2-form, then

$$d(\omega \wedge \eta) = d\omega \wedge \eta - \omega \wedge d\eta.$$

The *Lie derivative* of a k-form ω with respect to a vector field X is another k-form denoted $L_X\omega$. In finite dimensions it measures the rate of change of ω along the integral curves of X and is defined by

$$(L_X\omega)_p(v_1, \ldots, v_k) = \frac{d}{dt}\,(\varphi_t^*\omega)_p\,(v_1, \ldots, v_k)\,|_{t=0}$$

where $\{\varphi_t\}$ is the (local) 1-parameter group of diffeomorphisms (local flow) induced by X. One can show that it is given in terms of the exterior derivative and contraction with X by Cartan's "magic" formula

$$L_X\omega = d(\iota_X\omega) + \iota_X(d\omega) \tag{1.5}$$

(for a proof of this see Theorem 4.4.1 of [BG]). It is (1.5) that we will adopt as the definition of the Lie derivative of forms on a Banach space. Lie differentiation satisfies the product rule

$$L_X(\omega \wedge \eta) = L_X\omega \wedge \eta + \omega \wedge L_X\eta$$

with respect to the wedge product.

Exercise 1.1.6. Show that the Lie derivative commutes with the exterior derivative, that is,

$$d(L_X\omega) = L_X(d\omega).$$

Exercise 1.1.7. Show that L_X commutes ι_X, that is,

$$L_X\,(\iota_X\omega) = \iota_X\,(L_X\omega).$$

1.2 Euler-Lagrange Equations for Real Scalar Fields

Our principal references for this section are [AMR], [LL] and [Evans]. We will let $\mathcal{H}$ denote a *real*, separable Hilbert space with inner product denoted $\langle\ ,\ \rangle_{\mathcal{H}}$ or simply $\langle\ ,\ \rangle$ if this will cause no confusion.

Remark 1.2.1. The reason we consider the real and complex cases separately is as follows. The functions for which we need to write down a variational principle and derive the Euler-Lagrange equations are *Lagrangians* and these are *always real-valued* functionals on a Hilbert space of fields. If the fields are real-valued, then such Lagrangians are maps from a real Banach space to a real Banach space and therefore have a (real-linear) Fréchet derivative.

Furthermore, the quantization of complex-valued fields (such as complex Klein-Gordon fields) often presupposes the quantization of the corresponding real field and the resulting quantum fields have different physical interpretations in the real and complex cases. For the time being we will focus on real fields and will return to the complex case in Section 2.3.

Let $f : U \subseteq \mathcal{H} \to \mathbb{R}$ be a real-valued function that is (Fréchet) C^1 on the open set U. For each $\psi \in \mathcal{H}$, $Df(\psi) : \mathcal{H} \to \mathbb{R}$ is a continuous linear functional on $\mathcal{H}$ so, by the Riesz Representation Theorem (Theorem 6.2.4 of [Fried]), there exists a unique element of $\mathcal{H}$ denoted

$$\frac{\delta f}{\delta\psi} \in \mathcal{H}$$

such that, for every $\phi \in \mathcal{H}$,

$$Df(\psi)\cdot\phi = \frac{d}{d\varepsilon}f(\psi+\varepsilon\phi)\Big|_{\varepsilon=0} = \langle\,\frac{\delta f}{\delta\psi},\phi\,\rangle_{\mathcal{H}}. \tag{1.6}$$

$\frac{\delta f}{\delta\psi}$ is generally called the *functional derivative* of f with respect to ψ; it is simply the analogue of the finite-dimensional gradient for a real-valued function on a real Hilbert space.

Theorem 1.2.1. *Let $\mathcal{H}$ be a real Hilbert space, U an open subset of $\mathcal{H}$ and $f : U \subseteq \mathcal{H} \to \mathbb{R}$ a C^1 real-valued function on U. Then a necessary condition for f to have a local extremum at $\psi \in U$ is that $\frac{\delta f}{\delta\psi} = 0$.*

Remark 1.2.2. Points $\psi \in \mathcal{H}$ at which $\frac{\delta f}{\delta \psi} = 0$ are called *critical points*, or *stationary points* of f and, as in elementary calculus, they need not be points at which f actually has a local extremum.

Proof. Notice that, by the nondegeneracy of $\langle\ ,\ \rangle_{\mathcal{H}}$, $\frac{\delta f}{\delta \psi} = 0$ if and only if $\frac{d}{d\varepsilon} f(\psi + \varepsilon\phi)\big|_{\varepsilon=0} = 0$ for every $\phi \in \mathcal{H}$. But, if f has a local extremum at $\psi \in \mathcal{H}$, then, for every $\phi \in \mathcal{H}$, the differentiable, real-valued function of the real variable ε given by $\varepsilon \mapsto f(\psi + \varepsilon\phi)$ has a local extremum at $\varepsilon = 0$ and therefore $\frac{d}{d\varepsilon} f(\psi + \varepsilon\phi)\big|_{\varepsilon=0} = 0$. □

Now we turn to the action functionals whose critical points are classical real scalar fields. To construct an action functional one begins with a smooth real-valued function $\mathcal{L} : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ which we will write as

$$\mathcal{L}(u, v_1, \ldots, v_n, x^1, \ldots, x^n) = \mathcal{L}(u, v, x).$$

Remark 1.2.3. In principle this smooth function could be quite general, but until further notice we will restrict our attention to those of the form

$$\mathcal{L}(u, v_1, \ldots, v_n) = K(v_1, \ldots, v_n) - V(u),$$

where K and V, called the *kinetic* and *potential terms*, respectively, are at most *quadratic* in $v_1, \ldots, v_n$ and in u, respectively. In particular, $\mathcal{L}$ will not depend on x (see Section 8.1.2 of [Evans] for a more general discussion).

Example 1.2.1. The two examples we will focus on are as follows.

1. (*Dirichlet*) $\mathcal{L}(u, v) = \frac{1}{2}\|v\|^2 = \frac{1}{2}(v_1^2 + \cdots + v_n^2)$

2. (*Klein-Gordon*) In this case, $n = 4$ and

$$\mathcal{L}(u, v) = \mathcal{L}(u, v_1, v_2, v_3, v_4) = \frac{1}{2}\left(\frac{1}{c^2} v_1^2 - v_2^2 - v_3^2 - v_4^2 - \frac{m^2c^2}{\hbar^2} u^2\right),$$

   where c is the speed of light, $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant, and m is a positive constant.

We would like $\mathcal{L}(u, v_1, \ldots, v_n)$ to determine a *Lagrangian density* on some real Hilbert space $\mathcal{H}$ of fields $\varphi(x)$ by substituting $u = \varphi(x)$ and $v_i = \partial_i\varphi(x)$, $i = 1 \ldots, n$. That is, we want to define

$$\mathcal{L}(\varphi(x), \nabla\varphi(x)) = \mathcal{L}(\varphi(x), \partial_1\varphi(x), \ldots, \partial_n\varphi(x))$$

for $\varphi \in \mathcal{H}$ and from this obtain an *action functional* $S : \mathcal{H} \to \mathbb{R}$ of the form

$$S[\varphi] = \int_{\mathbb{R}^n} \mathcal{L}(\varphi(x), \nabla\varphi(x))\, d^n x. \tag{1.7}$$

From $S[\varphi]$ and the Principle of Least Action we will arrive at differential equations defining the dynamics of the fields. Of course, one must specify the Hilbert space $\mathcal{H}$ of fields in such a way that the integral in (1.7) exists for every $\varphi \in \mathcal{H}$ and the choice of this Hilbert space will be dictated by the nature of $\mathcal{L}$. For quadratic $\mathcal{L}$ one can take $\mathcal{H}$ to be the Sobolev space $H^1(\mathbb{R}^n;\mathbb{R})$ since then φ and all of its first order distributional derivatives $\partial_j\varphi = \frac{\partial\varphi}{\partial x^j}$, $j = 1, \ldots, n$, are in $L^2(\mathbb{R}^n)$ (see Appendix A).

Example 1.2.2. For the Dirichlet and Klein-Gordon examples one obtains the following action functionals.

1. (*Dirichlet*) $\mathcal{L}(u, v_1, \ldots, v_n) = \frac{1}{2}(v_1^2 + \cdots + v_n^2) \Rightarrow$

$$S[\varphi] = \frac{1}{2}\int_{\mathbb{R}^n} \left(\frac{\partial\varphi}{\partial x^1}\right)^2 + \cdots + \left(\frac{\partial\varphi}{\partial x^n}\right)^2 d^n x = \frac{1}{2}\int_{\mathbb{R}^n} \|\nabla\varphi\|^2\, d^n x.$$

   The Hilbert space of fields is taken to be $\mathcal{H} = H^1(\mathbb{R}^n;\mathbb{R})$. In particular, $S[\varphi]$ is well-defined and finite for all smooth L^2 functions with L^2 first partial derivatives.

2. (*Klein-Gordon*) $\mathcal{L}(u, v_1, v_2, v_3, v_4) = \frac{1}{2}\left(\frac{1}{c^2}v_1^2 - v_2^2 - v_3^2 - v_4^2 - \frac{m^2c^2}{\hbar^2}u^2\right) \Rightarrow$

$$S[\varphi] = \frac{1}{2}\int_{\mathbb{R}^4} \left[\frac{1}{c^2}\left(\frac{\partial\varphi}{\partial t}\right)^2 - \sum_{k=1}^{3}\left(\frac{\partial\varphi}{\partial x^k}\right)^2 - \frac{m^2c^2}{\hbar^2}\varphi^2\right] d^4 x,$$

   where we have denoted the coordinates on $\mathbb{R}^4$ by (t, x^1, x^2, x^3). As in Example (1) above we will take $\mathcal{H} = H^1(\mathbb{R}^4;\mathbb{R})$.

   Remark 1.2.4. In Section 2.1 we will show that the Klein-Gordon action $S[\varphi]$ is invariant under proper, orthochronous Lorentz transformations so that it is more natural to regard the fields φ as defined on Minkowski spacetime $\mathbb{R}^{1,3}$. Needless to say, the Lebesgue measure does not care if we choose to supply $\mathbb{R}^4$ with a Minkowski inner product and regard it as $\mathbb{R}^{1,3}$ so the Sobolev spaces $H^1(\mathbb{R}^4;\mathbb{R})$ and $H^1(\mathbb{R}^{1,3};\mathbb{R})$ are precisely the same.

For a quadratic Lagrangian density the action functional is therefore a real-valued function on the real Hilbert space $H^1(\mathbb{R}^n;\mathbb{R})$ so that a necessary condition for S to have a local extremum at $\psi \in H^1(\mathbb{R}^n;\mathbb{R})$ is that the functional derivative $\frac{\delta S}{\delta\psi}$ vanish (Theorem 1.2.1). We would like to write this condition out more explicitly. We will do this first for a smooth element ψ of $H^1(\mathbb{R}^n;\mathbb{R})$ and arrive at a differential equation that ψ must satisfy and then we will discuss more general stationary points of S on $H^1(\mathbb{R}^n;\mathbb{R})$. Let $\psi \in C^\infty(\mathbb{R}^n;\mathbb{R}) \cap H^1(\mathbb{R}^n;\mathbb{R})$ be fixed. The condition $\frac{\delta S}{\delta\psi} = 0$

implies, in particular, that if $\phi \in \mathcal{S}(\mathbb{R}^n;\mathbb{R})$ is arbitrary, then the smooth function $\varepsilon \mapsto S[\psi + \varepsilon\phi]$ satisfies

$$\frac{d}{d\varepsilon} S[\psi + \varepsilon\phi]\Big|_{\varepsilon=0} = 0.$$

But

$$\begin{aligned}\frac{d}{d\varepsilon} S[\psi + \varepsilon\phi]\Big|_{\varepsilon=0} &= \frac{d}{d\varepsilon} \int_{\mathbb{R}^n} \mathcal{L}\big(\,\psi(x) + \varepsilon\phi(x),\ \nabla\psi(x) + \varepsilon\nabla\phi(x)\,\big)\, d^n x\Big|_{\varepsilon=0} \\ &= \int_{\mathbb{R}^n} \frac{d}{d\varepsilon} \mathcal{L}\big(\,\psi(x) + \varepsilon\phi(x),\ \nabla\psi(x) + \varepsilon\nabla\phi(x)\,\big)\Big|_{\varepsilon=0} d^n x \\ &= \int_{\mathbb{R}^n} \Big[\, \frac{\partial\mathcal{L}}{\partial u}(\psi(x), \nabla\psi(x))\, \phi(x) + \sum_{k=1}^{n} \frac{\partial\mathcal{L}}{\partial v_k}(\psi(x), \nabla\psi(x))\, \frac{\partial\phi}{\partial x^k}\,\Big]\, d^n x.\end{aligned}$$

Notice that we have used the smoothness of $\mathcal{L}, \psi$ and ϕ to justify differentiating under the integral sign. Since $\mathcal{L}$ is quadratic and ϕ is a Schwartz function, each term in the sum can be integrated by parts to give

$$\frac{d}{d\varepsilon} S[\psi + \varepsilon\phi]\Big|_{\varepsilon=0} = \int_{\mathbb{R}^n} \Big[\, \frac{\partial\mathcal{L}}{\partial u}(\psi(x), \nabla\psi(x)) + \sum_{k=1}^{n} -\frac{\partial}{\partial x^k}\Big(\frac{\partial\mathcal{L}}{\partial v_k}(\psi(x), \nabla\psi(x))\Big)\Big]\, \phi(x)\, d^n x.$$

Since $\mathcal{S}(\mathbb{R}^n;\mathbb{R})$ is dense in $H^1(\mathbb{R}^n;\mathbb{R})$ a limiting argument shows that this equality is satisfied for every ϕ in $H^1(\mathbb{R}^n;\mathbb{R})$ so we conclude that

$$\frac{\delta S}{\delta\psi} = \frac{\partial\mathcal{L}}{\partial u}(\psi(x), \nabla\psi(x)) - \sum_{k=1}^{n} \frac{\partial}{\partial x^k}\Big(\frac{\partial\mathcal{L}}{\partial v_k}(\psi(x), \nabla\psi(x))\Big).$$

The condition $\frac{\delta S}{\delta\psi} = 0$ for a smooth element ψ of $H^1(\mathbb{R}^n;\mathbb{R})$ therefore becomes

$$\frac{\partial\mathcal{L}}{\partial u}(\psi(x), \nabla\psi(x)) - \sum_{k=1}^{n} \frac{\partial}{\partial x^k}\Big(\frac{\partial\mathcal{L}}{\partial v_k}(\psi(x), \nabla\psi(x))\Big) = 0. \tag{1.8}$$

This is the *Euler-Lagrange equation*. It is satisfied by any smooth stationary point of the action functional $S[\phi]$. One generally suppresses the variables $(u, v_1, \ldots, v_n)$ in favor of $(\psi, \partial\psi/\partial x^1, \ldots, \partial\psi/\partial x^n)$ and omits the argument $(\psi(x), \nabla\psi(x))$ to write the Euler-Lagrange equation in its traditional form as

$$\frac{\partial\mathcal{L}}{\partial\psi} - \sum_{k=1}^{n} \frac{\partial}{\partial x^k}\Big(\frac{\partial\mathcal{L}}{\partial(\partial\psi/\partial x^k)}\Big) = 0, \tag{1.9}$$

where it is understood that everything is evaluated at $(\psi(x), \nabla\psi(x))$.

Example 1.2.3. The Dirichlet action is

$$S[\psi] = \frac{1}{2}\int_{\mathbb{R}^n} \left(\frac{\partial\psi}{\partial x^1}\right)^2 + \cdots + \left(\frac{\partial\psi}{\partial x^n}\right)^2 d^n x = \frac{1}{2}\int_{\mathbb{R}^n} \|\nabla\psi\|^2 \, d^n x.$$

Since $\frac{\partial\mathcal{L}}{\partial\psi} = 0$ and $\frac{\partial\mathcal{L}}{\partial(\partial\psi/\partial x^k)} = \frac{\partial\psi}{\partial x^k}$, the Euler-Lagrange equation is just the Laplace equation

$$\varDelta\psi = \sum_{k=1}^{n} \frac{\partial^2\psi}{(\partial x^k)^2} = 0.$$

The Klein-Gordon action is

$$S[\psi] = \frac{1}{2}\int_{\mathbb{R}^4} \left[\frac{1}{c^2}\left(\frac{\partial\psi}{\partial t}\right)^2 - \sum_{k=1}^{3}\left(\frac{\partial\psi}{\partial x^k}\right)^2 - \frac{m^2c^2}{\hbar^2}\psi^2\right] d^4x.$$

Exercise 1.2.1. Show that the Euler-Lagrange equation can be written

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \varDelta\psi + \frac{m^2c^2}{\hbar^2}\psi = 0,$$

where $\varDelta\psi = \sum_{k=1}^{3} \frac{\partial^2\psi}{(\partial x^k)^2}$ is the spatial Laplacian. This is the *Klein-Gordon equation*. Letting

$$\Box_c = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \varDelta$$

and

$$\mu = \frac{mc}{\hbar}$$

one can write it as

$$(\Box_c + \mu^2)\psi = 0.$$

In units for which $c = 1$, $\Box_c$ is just the ordinary d'Alembertian and we will write it simply $\Box$.

Smooth stationary points of the action functional $S[\varphi]$ are solutions to the Euler-Lagrange equation, but need not correspond to extrema and, even when extrema exist, one cannot expect them to be smooth in general. To carry out a systematic study of the solutions to the Euler-Lagrange equation one needs to enlarge the space of functions in which the search for them takes place. If the Euler-Lagrange equation happens to be linear (as it is for the Dirichlet and Klein-Gordon examples), then there is a weaker notion of "solution" that we will briefly try to motivate here and then exploit in the next section. Let $D = \sum_{|\alpha|\leq m} a_\alpha(x)\partial_\alpha$, where $a_\alpha \in C^\infty(\mathbb{R}^n)$ for each multi-index α, be a linear differential operator (the functions can be real or complex). If f is smooth, then one can seek smooth solutions to the differential equation

$$D\psi = f. \tag{1.10}$$

On the other hand, if f is a distribution (such as a Dirac delta or some L^2 function), then one can generally not expect to find a smooth function ψ satisfying (1.10). One can, however, interpret each ∂_α as a distributional derivative and seek distributions ψ for which (1.10) is satisfied. Such a ψ is called a *distributional solution* to $D\psi = f$. Suppose one can prove that distributional solutions exist (we will discuss how this might be done in a moment). One can then contend with the issue of whether or not these distributions are regular and have some degree of smoothness. For elliptic equations (such as the Laplace equation) one can appeal to what are called *elliptic regularity theorems*. However, we will be interested primarily in wave equations, which are hyperbolic, and for these there are no such general results because distributional solutions simply need not be smooth.

We will gain some experience finding distributional solutions to the Klein-Gordon equation in Section 2.2. Briefly, the idea is this. The Klein-Gordon equation (with $c = 1$) is $(\square + \mu^2)\psi = 0$, where the differential operator $\square + \mu^2$ is defined in the ordinary or the distributional sense depending on the nature of ψ. Applying the Fourier transform to both sides gives an algebraic equation for the Fourier transform $\hat{\psi}$. Specifically, one obtains $(p^2 - \mu^2)\hat{\psi} = 0$, where $p \in \mathbb{R}^4$ and $p^2 = p_0^2 - p_1^2 - p_2^2 - p_3^2$. One then finds all of the solutions to this algebraic equation. These are generally distributions. Next one applies the inverse Fourier transform to each of these to obtain the Klein-Gordon solutions. These may be either classical solutions or distributional solutions depending on the distribution $\hat{\psi}$ from which they arose. The purpose of the next section is to carry all of this out in more detail.

Chapter 2
Klein-Gordon Fields

2.1 Real Klein-Gordon Fields: Lagrangian Formulation

2.1.1 Relativistic Invariance

We have seen in the previous section how the Klein-Gordon equation arises as the Euler-Lagrange equation for a certain action functional. We will begin this section by considering another means by which the classical Klein-Gordon equation can be "derived" from physical principles.

The derivation of the Klein-Gordon equation that one generally sees in the physics literature (for example, Section 2.2 of [Ryd]) goes something like this. One begins with the relativistic energy-momentum relation

$$E^2 = |\mathbf{p}|^2 c^2 + m^2 c^4 \tag{2.1}$$

(see (2.59) of [Nab6]). Now we quote from Section 2.2 of [Ryd]. "The wave equation is obtained from (2.1) by substituting differential operators for E and $\mathbf{p}$, in the fashion standard in quantum theory

$$E \to i\hbar \frac{\partial}{\partial t}, \quad p_k \to -i\hbar \frac{\partial}{\partial x^k}\,, \quad k = 1, 2, 3.\text{"} \tag{2.2}$$

The result is

$$-\hbar^2 \frac{\partial^2}{\partial t^2} = -\hbar^2 c^2 \Delta + m^2 c^4$$

and, from this,

$$\frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} - \Delta \varphi + \frac{m^2 c^2}{\hbar^2} \varphi = 0$$

for any smooth φ. In this sense, the Klein-Gordon equation is regarded as the quantized version of the relativistic energy-momentum relation.

The Klein-Gordon equation was originally proposed as a relativistic substitute for the Schrödinger equation, but certain issues were apparent from the outset. It is, for example, second order in t and therefore a well-posed initial value problem would require the specification of both ψ and $\frac{\partial \psi}{\partial t}$ at $t = 0$. Physicists would say that there is an extra degree of freedom in the Klein-Gordon equation that is not present in the Schrödinger equation. Because of such issues one should regard the Klein-Gordon equation as a classical field equation (analogous to Maxwell's equations for the electromagnetic field) and *not* as an equation describing the evolution of the quantum state of some particle.

It is not uncommon in the physics literature for an investigation to begin, not with equations describing the time evolution of the system of interest, but rather with a Lagrangian density. From this one obtains an action functional and then an appeal to the principle of stationary action dictates that the equations of motion are just the Euler-Lagrange equations. We have already seen how the Klein-Gordon equation can be arrived at in this way. The Lagrangian density

$$\mathcal{L}(\varphi(x), \nabla\varphi(x)) = \frac{1}{2}\Big[\frac{1}{c^2}\left(\frac{\partial \varphi}{\partial t}\right)^2 - \sum_{k=1}^{3}\left(\frac{\partial \varphi}{\partial x^k}\right)^2 - \frac{m^2c^2}{\hbar^2}\varphi^2\Big] \tag{2.3}$$

gives rise to the action

$$S[\varphi] = \frac{1}{2}\int_{\mathbb{R}^4}\Big[\frac{1}{c^2}\left(\frac{\partial \varphi}{\partial t}\right)^2 - \sum_{k=1}^{3}\left(\frac{\partial \varphi}{\partial x^k}\right)^2 - \frac{m^2c^2}{\hbar^2}\varphi^2\Big]d^4x \tag{2.4}$$

and then the Euler-Lagrange (Klein-Gordon) equation

$$\frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2} - \Delta\varphi + \frac{m^2c^2}{\hbar^2}\varphi = 0. \tag{2.5}$$

Whether or not this derivation is more fundamental than the quantization of the energy-momentum relation is a matter of taste, or philosophy, but it is certainly more typical.

Remark 2.1.1. Quantum field theory is an attempt to reconcile quantum mechanics and special relativity. As a result, the equations are rich in c s and $\hbar$ s and these can become something of an algebraic nuisance. To alleviate some of the bother it is customary to work in what are called *natural units* in which both c and $\hbar$ are dimensionless and equal to 1. In principle, this is not difficult to do. Any system of units is constructed in the same way. One chooses certain quantities to regard as fundamental units (in mechanics these would include length, time, and mass). For each of these one chooses (arbitrarily) some standard (such as the *international prototype kilogram*, which is just a hunk of metal stored somewhere safe), selects a name for the unit (like *kilogram*) and assigns a value of 1 to the standard. Other *derived units* are then determined by physical laws; for example, the unit of force in classical mechanics is determined by $\mathbf{F} = m\mathbf{A}$ to be $(kg)ms^{-2}$ which is then given a new name (*Newton*). One could equally well have begun with speed as a fundamental unit, taken the speed of light *in vacuo* to be the standard and assigned it the dimensionless value 1 (which would then dictate that length and time have the same units). This, in fact, is what we have done in Section 2.2 of [Nab6]. The advantages are clear; formulas are streamlined and calculations are simplified. However, there are equally obvious disadvantages. For instance, one loses any visual distinction between various orders of magnitude (c looks the same as c^2). Moreover, laboratory measurements are invariably recorded in more traditional units and one must eventually compare theoretical predictions with experimental results. All of this sounds straightforward and rather dull, but it is actually not so straightforward and not at all dull.

It seems like a pretty dull subject. However, in the realm of modern physics a careful examination of the choice of units leads to some useful (even profound) insights into the way the Universe works.

-R.L. Jaffe

We will not pursue this any further here, but will simply refer those interested in the matter to http://stuff.mit.edu/afs/athena/course/8/8.06/spring08/handouts/units.pdf and will adopt natural units in which

$$\boxed{c = \hbar = 1}$$

unless there is some reason to believe that this will obscure an essential point.

In particular, the Klein-Gordon Lagrangian density (2.3), action (2.4), and equation (2.5) can now be written

$$\mathcal{L}(\varphi(x), \nabla\varphi(x)) = \frac{1}{2}\Big[\Big(\frac{\partial\varphi}{\partial t}\Big)^2 - \sum_{k=1}^{3}\Big(\frac{\partial\varphi}{\partial x^k}\Big)^2 - m^2\varphi^2\Big],$$

$$S[\varphi] = \frac{1}{2}\int_{\mathbb{R}^4}\Big[\Big(\frac{\partial\varphi}{\partial t}\Big)^2 - \sum_{k=1}^{3}\Big(\frac{\partial\varphi}{\partial x^k}\Big)^2 - m^2\varphi^2\Big]d^4x$$

and

$$(\square + m^2)\varphi = 0,$$

where

$$\square = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{(\partial x^1)^2} - \frac{\partial^2}{(\partial x^2)^2} - \frac{\partial^2}{(\partial x^3)^2}$$

is the d'Alembertian.

Remark 2.1.2. The Lagrangian density $\mathcal{L}$ is the difference of a kinetic term $T = \frac{1}{2}\Big[\Big(\frac{\partial\varphi}{\partial t}\Big)^2 - \sum_{k=1}^3 \Big(\frac{\partial\varphi}{\partial x^k}\Big)^2\Big]$ and a potential term $V = \frac{1}{2}m^2\varphi^2$. Notice that the potential term is quadratic and therefore contributes a linear term to the Euler-Lagrange equation. This is generally called the *mass term*. Adding a quartic term $\frac{\lambda}{4}\varphi^4$, $\lambda > 0$, to the potential introduces a cubic nonlinearity into the Euler-Lagrange equations which then represents a self-interaction energy; λ is the *coupling constant* and measures the strength of the self-interaction. The resulting Lagrangian density

$$\mathcal{L}_{Higgs}(\varphi(x), \nabla\varphi(x)) = \frac{1}{2}\Big[\Big(\frac{\partial\varphi}{\partial t}\Big)^2 - \sum_{k=1}^{3}\Big(\frac{\partial\varphi}{\partial x^k}\Big)^2 - m^2\varphi^2 - \frac{\lambda}{2}\varphi^4\Big]$$

is called the *Higgs Lagrangian*.

Now that $c = 1$ we can, and will, adopt the relativistic notation introduced in Chapter 2 of [Nab6]. In particular, $t = x^0$ and the Klein-Gordon Lagrangian density, action and equation can now be written

$$\mathcal{L}(\varphi, \partial_\alpha\varphi) = \frac{1}{2}\left(\partial_\alpha\varphi\, \partial^\alpha\varphi - m^2\varphi^2\right),$$

$$S[\varphi] = \frac{1}{2}\int_{\mathbb{R}^{1,3}} (\partial_\alpha\varphi\, \partial^\alpha\varphi - m^2\varphi^2)\, d^4x,$$

and

$$(\partial_\alpha\partial^\alpha + m^2)\,\varphi = 0.$$

We have written the action as an integral over $\mathbb{R}^{1,3}$ rather than $\mathbb{R}^4$, but only because we are now making some explicit use of the Lorentz inner product; d^4x still refers to integration with respect to the usual Lebesgue measure. However, this splitting of $\mathbb{R}^{1,3}$ into $\mathbb{R}\times\mathbb{R}^3$ by a choice of Minkowski basis suggests writing the action as

$$S[\varphi] = \int_{\mathbb{R}}\left[\frac{1}{2}\int_{\mathbb{R}^3} (\partial_\alpha\varphi\, \partial^\alpha\varphi - m^2\varphi^2)\, d^3\mathbf{x}\right] dx^0,$$

where $d\mathbf{x} = dx^1 dx^2 dx^3$. The spatial integral of the Lagrangian density $\mathcal{L}$ is generally denoted

$$L = \int_{\mathbb{R}^3} \mathcal{L}\, d^3\mathbf{x} = \frac{1}{2}\int_{\mathbb{R}^3} (\partial_\alpha\varphi\, \partial^\alpha\varphi - m^2\varphi^2)\, d^3\mathbf{x}$$

and called simply the *Lagrangian*. The action is the time integral of the Lagrangian.

Remark 2.1.3. One should be aware, however, that the current literature very often uses the terms *Lagrangian* and *Lagrangian density* interchangeably.

Writing the Klein-Gordon equation in the form $(\partial_\alpha\partial^\alpha + m^2)\varphi(x) = 0$ is particularly convenient for the discussion of what is called its *Lorentz invariance*. We would like to be quite clear on what this is intended to mean. Suppose we have two inertial coordinate systems related by $\hat{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta,\ \alpha = 0,1,2,3$. The relativity principle requires that the "laws of physics" must be the same in these two frames. Although admittedly rather vague, the "laws of physics" should certainly include the equations governing the behavior of whatever physical system is under consideration at the moment. In particular, the Klein-Gordon equation in the hatted reference frame should have exactly the same form as in the unhatted frame, that is,

$$(\hat{\partial}_\alpha\hat{\partial}^\alpha + m^2)\,\hat{\varphi}(\hat{x}) = 0.$$

Here $\hat{\partial}_\alpha$ means $\partial/\partial\hat{x}^\alpha$, but the meaning of $\hat{\varphi}(\hat{x})$ requires some discussion. The Klein-Gordon field is intended to be a *scalar* field, that is, simply a real- (or complex-) valued *function* on Minkowski spacetime. Like any real- (or complex-) valued function on any manifold the Klein-Gordon field will have a coordinate representation in any chart and two such coordinate representations are related in the simplest possible way; one need only substitute the corresponding coordinate transformation map into one to obtain the other (in $\mathbb{R}^2$, for example, if the Cartesian coordinate representation is $x^2 + 3y^2$, then the polar coordinate representation is $r^2 + 2r^2 sin^2\theta$). For coordinates on Minkowski spacetime corresponding to two oriented, orthonormal bases, this takes the form $\hat{\varphi}(\hat{x}) = \varphi(\Lambda^{-1}\hat{x})$ for some $\Lambda \in \mathcal{L}_+^\uparrow$. More explicitly, if $\hat{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta$, then

$$\hat{\varphi}(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) = \varphi(\,\Lambda_\alpha{}^0\,\hat{x}^\alpha, \Lambda_\alpha{}^1\,\hat{x}^\alpha, \Lambda_\alpha{}^2\,\hat{x}^\alpha, \Lambda_\alpha{}^3\,\hat{x}^\alpha\,).$$

This transformation law is what it means for φ to be a *Lorentz scalar field*.

Having specified a transformation law for the coordinate representations $\varphi, \hat{\varphi}, \ldots$, we can now check to see if the Klein-Gordon equation is Lorentz invariant in the sense that $(\partial_\alpha\partial^\alpha + m^2)\varphi(x) = 0$ if and only if $(\hat{\partial}_\alpha\hat{\partial}^\alpha + m^2)\,\hat{\varphi}(\hat{x}) = 0$. In fact, we will show more. We claim that, for any smooth φ,

$$(\hat{\partial}_\alpha\hat{\partial}^\alpha + m^2)\,\hat{\varphi}(\hat{x}) = (\partial_\alpha\partial^\alpha + m^2)\,\varphi(x). \tag{2.6}$$

In more detail, we claim that

$$(\,\eta^{\alpha\beta}\frac{\partial}{\partial\hat{x}^\alpha}\frac{\partial}{\partial\hat{x}^\beta} + m^2\,)\,\hat{\varphi}(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) = (\,\eta^{\alpha\beta}\frac{\partial}{\partial x^\alpha}\frac{\partial}{\partial x^\beta} + m^2\,)\,\varphi(x^0, x^1, x^2, x^3), \tag{2.7}$$

where $x^\mu = \Lambda_\nu{}^\mu \hat{x}^\nu$. To see this note first that

$$m^2\hat{\varphi}(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) = m^2\varphi(\Lambda_\nu{}^0\hat{x}^\nu, \Lambda_\nu{}^1\hat{x}^\nu, \Lambda_\nu{}^2\hat{x}^\nu, \Lambda_\nu{}^3\hat{x}^\nu) \tag{2.8}$$

by definition. Then two applications of the chain rule give

$$\eta^{\alpha\beta}\frac{\partial}{\partial\hat{x}^\alpha}\frac{\partial}{\partial\hat{x}^\beta}\hat{\varphi}(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) = \eta^{\alpha\beta}\Lambda_\alpha{}^\gamma\Lambda_\beta{}^\delta\frac{\partial}{\partial x^\gamma}\frac{\partial}{\partial x^\delta}\varphi(\Lambda_\nu{}^0\hat{x}^\nu, \Lambda_\nu{}^1\hat{x}^\nu, \Lambda_\nu{}^2\hat{x}^\nu, \Lambda_\nu{}^3\hat{x}^\nu). \tag{2.9}$$

But $\eta^{\alpha\beta}\Lambda_\alpha{}^\gamma\Lambda_\beta{}^\delta = \eta^{\gamma\delta}$ (Exercise 2.3.2 (3) of [Nab6]) so

$$\begin{aligned}\eta^{\alpha\beta}\frac{\partial}{\partial\hat{x}^\alpha}\frac{\partial}{\partial\hat{x}^\beta}\hat{\varphi}(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3) &= \eta^{\gamma\delta}\frac{\partial}{\partial x^\gamma}\frac{\partial}{\partial x^\delta}\varphi(\Lambda_\nu{}^0\hat{x}^\nu, \Lambda_\nu{}^1\hat{x}^\nu, \Lambda_\nu{}^2\hat{x}^\nu, \Lambda_\nu{}^3\hat{x}^\nu) \\ &= \eta^{\alpha\beta}\frac{\partial}{\partial x^\alpha}\frac{\partial}{\partial x^\beta}\varphi(\Lambda_\nu{}^0\hat{x}^\nu, \Lambda_\nu{}^1\hat{x}^\nu, \Lambda_\nu{}^2\hat{x}^\nu, \Lambda_\nu{}^3\hat{x}^\nu).\end{aligned} \tag{2.10}$$

Adding (2.8) and (2.10) gives (2.7), that is, (2.6).

Shortly we will find it convenient to have this last calculation slightly rephrased and generalized. For this we would like to think of a $\Lambda \in \mathcal{L}_+^\uparrow$ as giving rise to, not a change of coordinates (*passive transformation*), but rather a diffeomorphism of $\mathbb{R}^{1,3}$ onto itself (*active transformation*). Then the defining left action (representation) of $\mathcal{L}_+^\uparrow$ on $\mathbb{R}^{1,3}$

$$x \in \mathbb{R}^{1,3} \mapsto \Lambda \cdot x = \Lambda x \in \mathbb{R}^{1,3}$$

induces a left action (representation) of $\mathcal{L}_+^\uparrow$ on the real- (or complex-) valued functions on $\mathbb{R}^{1,3}$

$$\varphi(x) \mapsto (\Lambda \cdot \varphi)(x) = \varphi(\Lambda^{-1}x).$$

What we have shown above is that *the set of smooth solutions to the Klein-Gordon equation is invariant under this action of* $\mathcal{L}_+^\uparrow$.

Exercise 2.1.1. Define analogous actions of the Poincaré group $\mathcal{P}_+^\uparrow$ on Minkowski spacetime and its real- (or complex-) functions and show that the set of smooth solutions to the Klein-Gordon equation is invariant under the action of $\mathcal{P}_+^\uparrow$. As a result the Klein-Gordon equation is said to be *Poincaré invariant*.

More fundamentally, this invariance of the solution space is a consequence of the fact that the Lagrangian density itself is invariant. To make this more precise let's think of the Klein-Gordon Lagrangian density $\mathcal{L}(\varphi, \partial_\alpha\varphi)$ as a mapping from $C^\infty(\mathbb{R}^{1,3};\mathbb{R})$ to $C^\infty(\mathbb{R}^{1,3};\mathbb{R})$. Specifically, $\varphi \in C^\infty(\mathbb{R}^{1,3};\mathbb{R}) \mapsto \mathcal{L}(\varphi) \in C^\infty(\mathbb{R}^{1,3};\mathbb{R})$, defined by

$$\mathcal{L}(\varphi)(x) = \mathcal{L}(\varphi(x), \partial_\alpha\varphi(x)) = \frac{1}{2}[\,(\partial_\alpha\varphi(x))\,(\partial^\alpha\varphi(x)) - m^2\varphi(x)^2\,].$$

For each $\Lambda \in \mathcal{L}_+^\uparrow$ define $\Lambda \cdot \mathcal{L}$ by

$$(\Lambda \cdot \mathcal{L})(\varphi)(x) = \mathcal{L}(\varphi \circ \Lambda)(\Lambda^{-1} \cdot x).$$

Exercise 2.1.2. Here you will prove the Lorentz and Poincaré invariance of the Klein-Gordon Lagrangian density and action.

1. By computing just as we did above for the proof of (2.6) show that

$$\Lambda \cdot \mathcal{L} = \mathcal{L}$$

   for every $\Lambda \in \mathcal{L}_+^\uparrow$. This is the sense in which the Klein-Gordon Lagrangian density is Lorentz invariant.
2. Use the fact that every element of $\mathcal{L}_+^\uparrow$ has determinant 1 to show that the Klein-Gordon action $S[\varphi]$ is invariant under $\mathcal{L}_+^\uparrow$ in the sense that $S[\Lambda \cdot \varphi] = S[\varphi]$.
3. Extend both of these results to $\mathcal{P}_+^\uparrow$.

Remark 2.1.4. The Klein-Gordon equation is the simplest example of a *relativistically invariant wave equation*. There are many others, the most famous of which is the *Dirac equation*. These are of fundamental significance to relativistic quantum mechanics and quantum field theory. For a careful and systematic introduction to the study of such equations we refer to pages 269-352 of [Gel]. In particular, we draw attention to the discussion of equations (such as Klein-Gordon) that arise as Euler-Lagrange equations for Lorentz invariant Lagrangian densities. With these one can, as in classical mechanics, associate invariantly defined conserved quantities. We will now have a brief look at how these conservation laws arise for Klein-Gordon.

2.1.2 Conservation Laws

We begin by generalizing just a bit. Since it costs no more effort to do so we will begin by considering a general Lagrangian density $\mathcal{L}(\varphi) = \mathcal{L}(\varphi, \partial_\alpha \varphi) = \mathcal{L}(\varphi, \partial_1 \varphi, \ldots, \partial_n \varphi)$ on $\mathbb{R}^n$. As usual the partial derivatives with respect to $x^1, \ldots, x^n$ are denoted $\partial_1, \ldots, \partial_n$, while the derivatives of $\mathcal{L}$ with respect to its $n+1$ variables are written

$$\frac{\partial \mathcal{L}}{\partial \varphi}, \frac{\partial \mathcal{L}}{\partial(\partial_1 \varphi)}, \ldots, \frac{\partial \mathcal{L}}{\partial(\partial_n \varphi)}.$$

The diffeomorphism group $\text{Diff}(\mathbb{R}^n)$ of $\mathbb{R}^n$ acts on $\mathbb{R}^n$ in the obvious way; specifically, if $g : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism of $\mathbb{R}^n$ onto itself, then g acts on $\mathbb{R}^n$ by $g \cdot x = g(x)$. Consequently, each such g acts on the smooth real- (or complex-) valued functions on $\mathbb{R}^n$ by $(g \cdot \varphi)(x) = \varphi(g^{-1} \cdot x)$. If $\mathcal{L}$ is any Lagrangian density on $\mathbb{R}^n$ we define $g \cdot \mathcal{L}$ by

$$(g \cdot \mathcal{L})(\varphi)(x) = \mathcal{L}(\varphi \circ g)(g^{-1} \cdot x).$$

We will say that $\mathcal{L}$ is *invariant under the diffeomorphism* g if

$$g \cdot \mathcal{L} = \mathcal{L},$$

that is,

$$\mathcal{L}(\varphi \circ g)(g^{-1} \cdot x) = \mathcal{L}(\varphi)(x)$$

or, equivalently,

$$\mathcal{L}(\varphi \circ g)(x) = \mathcal{L}(\varphi)(g \cdot x)$$

for all φ and all $x \in \mathbb{R}^n$. In this case g is said to be a *symmetry* of $\mathcal{L}$. We have shown that every element of $\mathcal{P}_+^\uparrow$ is a symmetry of the Klein-Gordon Lagrangian density so $\mathcal{P}_+^\uparrow$ is called a *symmetry group* of the Klein-Gordon Lagrangian density (or action).

Exercise 2.1.3. Show that if a Lagrangian density $\mathcal{L}$ is invariant under g and if the Jacobian determinant of g is 1 everywhere, then the corresponding action functional is invariant under g in the sense that $S[g \cdot \varphi] = S[\varphi]$ for all φ. Stated otherwise, if $\mathcal{L}$ is invariant under g and g preserves the Lebesgue measure on $\mathbb{R}^n$ (or, equivalently, the standard volume form $\omega = dx^1 \wedge \cdots \wedge dx^n$), then it also preserves the action S.

Exercise 2.1.4. Find a diffeomorphism of $\mathbb{R}^{1,3}$ onto itself for which the Klein-Gordon Lagrangian density is *not* invariant under g.

Exactly as we did in classical mechanics (Appendix A.2 of [Nab6]) we will say that a smooth vector field

$$X = X^i(x)\frac{\partial}{\partial x^i}$$

on $\mathbb{R}^n$ is an *infinitesimal symmetry* of $\mathcal{L}$ if its corresponding 1-parameter group $\{g_t : \mathbb{R}^n \to \mathbb{R}^n\}_{t \in \mathbb{R}}$ of diffeomorphisms has the property that each g_t is a symmetry of $\mathcal{L}$, that is, if $\mathcal{L}(\varphi)(g_t(x)) = \mathcal{L}(\varphi \circ g_t)(x)$ for all $t \in \mathbb{R}$ and all $x \in \mathbb{R}^n$.

Remark 2.1.5. To streamline the exposition *we will assume that the vector field X is complete*, but all that we will say can be localized if it is not. *We will also assume that each g_t preserves the Lebesgue measure on $\mathbb{R}^n$.*

For the purposes of the ensuing calculation we will write the coordinate functions of g_t as

$$g_t(x) = (g_t^1(x), \ldots, g_t^n(x))$$

so that, by definition of the 1-parameter group associated with X,

$$\frac{dg_t^i}{dt}(x)\Big|_{t=0} = X^i(x), \quad i = 1, \ldots, n.$$

We begin by computing the derivative of $\mathcal{L}(\varphi) = \mathcal{L}(\varphi, \partial_\alpha \varphi)$ along the integral curves of X (that is, the Lie derivative of $\mathcal{L}(\varphi)$ with respect to X). This is given at each x by

$$\begin{aligned} X\left[\mathcal{L}(\varphi)(x)\right] = X^\alpha(x)\partial_\alpha(\mathcal{L}(\varphi)(x)) &= \lim_{t\to 0}\frac{\mathcal{L}(\varphi)(g_t(x)) - \mathcal{L}(\varphi)(x)}{t} \\ &= \lim_{t\to 0}\frac{\mathcal{L}(\varphi\circ g_t)(x) - \mathcal{L}(\varphi\circ g_0)(x)}{t} \\ &= \frac{d}{dt}\mathcal{L}(\varphi\circ g_t)(x)\Big|_{t=0} \end{aligned}$$

and we need to write out this derivative explicitly. Since

$$\mathcal{L}(\varphi\circ g_t)(x) = \mathcal{L}(\,\varphi\circ g_t(x),\ \partial_1(\varphi\circ g_t(x)), \ldots, \partial_n(\varphi\circ g_t(x))\,),$$

and

$$\varphi\circ g_t(x) = \varphi(g_t^1(x), \ldots, g_t^n(x))$$

this amounts to a completely routine, although rather annoying sequence of applications of the chain rule. Here is the first term.

$$\begin{aligned} &\left[\frac{\partial\mathcal{L}}{\partial\varphi}(\varphi\circ g_t(x))\frac{d}{dt}\varphi(g_t^1(x), \ldots, g_t^n(x))\right]_{t=0} \\ &\qquad = \frac{\partial\mathcal{L}}{\partial\varphi}(\varphi(x))\left[\,\frac{\partial\varphi}{\partial x^1}(x)\frac{dg_t^1}{dt}(x)\Big|_{t=0} + \cdots + \frac{\partial\varphi}{\partial x^n}(x)\frac{dg_t^n}{dt}(x)\Big|_{t=0}\,\right] \\ &\qquad = \frac{\partial\mathcal{L}}{\partial\varphi}(\varphi(x))\left[X^1(x)\frac{\partial\varphi}{\partial x^1}(x) + \cdots + X^n(x)\frac{\partial\varphi}{\partial x^n}(x)\,\right] \\ &\qquad = \frac{\partial\mathcal{L}}{\partial\varphi}(\varphi(x))\,X^\beta(x)\partial_\beta\varphi(x) \end{aligned}$$

The remaining terms

$$\left[\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}(\varphi\circ g_t(x))\frac{d}{dt}\partial_\alpha(\varphi(g_t^1(x), \ldots, g_t^n(x)))\right]_{t=0}$$

are a bit messier, but you no doubt get the idea.

Exercise 2.1.5. Compute as much of this as you feel you need to in order to be convinced that the end result is

$$\begin{aligned} X\left[\mathcal{L}(\varphi)(x)\right] &= X^\alpha(x)\partial_\alpha(\mathcal{L}(\varphi)(x)) = \frac{d}{dt}\mathcal{L}(\varphi\circ g_t(x))\Big|_{t=0} \\ &= \frac{\partial\mathcal{L}}{\partial\varphi}(\varphi(x))X^\beta(x)\partial_\beta\varphi(x) + \frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}(\varphi(x))\Big[\partial_\alpha X^\beta(x)\partial_\beta\varphi(x) + X^\beta(x)\partial_\beta\partial_\alpha\varphi(x)\Big], \end{aligned}$$

or, with a bit less clutter,

$$X^\alpha \partial_\alpha \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \varphi} X^\beta \partial_\beta \varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} [\partial_\alpha X^\beta \partial_\beta \varphi + X^\beta \partial_\beta \partial_\alpha \varphi].$$

Now *we assume that φ satisfies the Euler-Lagrange equation (1.9).* Then

$$\frac{\partial \mathcal{L}}{\partial \varphi} = \partial_\alpha \left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \right)$$

so we obtain

$$X^\alpha \partial_\alpha \mathcal{L} = \partial_\alpha \left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \right) X^\beta \partial_\beta \varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} [\partial_\alpha X^\beta \partial_\beta \varphi + X^\beta \partial_\beta \partial_\alpha \varphi].$$

Next we invoke our assumption that each g_t preserves the standard volume form ω on $\mathbb{R}^{1,3}$. This implies that the Lie derivative of ω with respect to X vanishes. But this Lie derivative is just the divergence if X times ω (this is essentially the definition of the divergence of X) so we conclude that $\partial_\beta X^\beta = 0$. Consequently, $\partial_\alpha (X^\alpha \mathcal{L}) = X^\alpha \partial_\alpha \mathcal{L}$. Substituting this into the last equation and rearranging gives

$$\begin{aligned} 0 &= -\partial_\alpha (X^\alpha \mathcal{L}) + \partial_\alpha \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} X^\beta \partial_\beta \varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} [\partial_\alpha X^\beta \partial_\beta \varphi + X^\beta \partial_\beta \partial_\alpha \varphi] \\ &= \partial_\alpha \Big(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} X^\beta \partial_\beta \varphi - \delta^\alpha_\beta X^\beta \mathcal{L} \Big) \\ &= \partial_\alpha J^\alpha (X), \end{aligned}$$

where

$$J^\alpha (X) = X^\beta \left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \partial_\beta \varphi - \delta^\alpha_\beta \mathcal{L} \right), \quad \alpha = 1, \ldots, n.$$

The vector $J(X) = (J^1(X), \ldots, J^n(X))$ is called the *current* of the field φ with respect to the infinitesimal symmetry $X = X^\beta \partial_\beta$ and it satisfies the *conservation law*

$$div\, J(X) = \partial_\alpha J^\alpha (X) = 0.$$

We will write out some concrete examples shortly and, in the process, will explain why this qualifies as a "conservation law".

Remark 2.1.6. This constitutes one version of *Noether's Theorem* for field theory. There are many other versions and a through, rigorous presentation of all of these is available in Sections 4.4 and 5.3 of [Olv]. For future reference, we would like to record just one these other versions here which can be proved in much the same way, but is also a special case of Corollary 4.30 of [Olv]. Suppose that instead of one field we have several $\varphi = (\varphi^1, \ldots, \varphi^N)$ and that the Lagrangian density $\mathcal{L}$ depends on these and their first partial derivatives $\partial_\alpha \varphi = (\partial_\alpha \varphi^1, \ldots, \partial_\alpha \varphi^N)$, $\alpha = 1, \ldots, n$. The

Euler-Lagrange equations then take the form

$$\frac{\partial \mathcal{L}}{\partial \varphi^\beta} - \partial_\alpha\left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi^\beta)}\right) = 0, \ \ \beta = 1, \ldots, N.$$

Instead of the diffeomorphism group acting on $\mathbb{R}^n$ and thereby acting on the fields as above, let us now suppose we have some matrix Lie group G acting on $\mathbb{R}^N$ and thereby acting on our set of fields. For example, if $N = 2$, then $SO(2)$ acts on $\mathbb{R}^2$ and therefore acts on $\varphi = (\varphi^1, \varphi^2)$ by rotating it.

$$\begin{pmatrix} \varphi^1 \\ \varphi^2 \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \varphi^1 \\ \varphi^2 \end{pmatrix}$$

Suppose now that this action of G on φ is a symmetry of the Lagrangian density, that is, that $\mathcal{L}$ is invariant under the action. Since G acts directly on the fields and not on the coordinates $x = (x^1, \ldots, x^n)$, the action is referred to as an *internal symmetry* of $\mathcal{L}$ to distinguish it from the *external (or spacetime) symmetries* of $\mathcal{L}$ arising from the action of the diffeomorphism group of $\mathbb{R}^n$ on $\mathbb{R}^n$. Each element A of the Lie algebra $\mathfrak{g}$ then gives rise to a 1-parameter group $\{e^{tA}\}_{t\in\mathbb{R}}$ of internal symmetries of $\mathcal{L}$. The infinitesimal generator of $\{e^{tA}\}_{t\in\mathbb{R}}$ is a vector field Y_A on $\mathbb{R}^N$. Now we will say that any vector field Y on $\mathbb{R}^N$ that is the infinitesimal generator for a 1-parameter group of internal symmetries of $\mathcal{L}$ is an *infinitesimal internal symmetry* of $\mathcal{L}$. Writing $y = (y^1, \ldots, y^N)$ for the standard coordinates on $\mathbb{R}^N$, any such Y can be written as

$$Y = Y^\beta \frac{\partial}{\partial y^\beta}.$$

Such an infinitesimal internal symmetry gives rise to a conserved current $J = (J^1, \ldots, J^n)$ given by

$$J^\alpha(Y) = -Y^\beta \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi^\beta)}, \ \ \alpha = 1, \ldots, n.$$

Specifically, the conservation law asserts that, if each J^α is evaluated on some solution $\varphi = (\varphi^1, \ldots, \varphi^N)$ to the Euler-Lagrange equations, then

$$div\, J(Y) = \partial_\alpha J^\alpha(Y) = 0.$$

This then is a version of Noether's Theorem for internal symmetries. We will have occasion to write out a concrete example in Section 2.3. We should mention also that Corollary 4.30 of [Olv] is a version of Noether's Theorem that allows for both internal and external symmetries.

Returning now to external symmetries, we would like to write out some concrete examples on $\mathbb{R}^{1,3}$. For these it is customary to suppress the dependence on X and also to write

$$(J^1, J^2, J^3, J^4) = (j^0, j^1, j^2, j^3) = (j^0, \mathbf{j}).$$

We will work in a fixed Minkowski coordinate system on $\mathbb{R}^{1,3}$ and will denote these coordinates $(t, x^1, x^2, x^3) = (t, \mathbf{x})$. Our conservation law $\partial_\alpha j^\alpha = 0$ then becomes

$$\frac{\partial j^0}{\partial t} = -div\, \mathbf{j} = -\sum_{i=1}^{3} \frac{\partial j^i}{\partial x^i}.$$

Assuming $j^0(t, \mathbf{x})$ is integrable on $\mathbb{R}^3$ for each fixed t we define the *charge*

$$Q(t) = \int_{\mathbb{R}^3} j^0(t, \mathbf{x})\, d^3\mathbf{x}.$$

If, for each t, $|\,\mathbf{j}(t, \mathbf{x})\,|$ approaches zero sufficiently fast as $|\mathbf{x}| \to \infty$, then the Divergence Theorem implies the *(global) conservation of charge*

$$\frac{dQ(t)}{dt} = 0.$$

Remark 2.1.7. Physicists have a tendency to refer to anything that is conserved in some sense as a "charge" or "current". The terminology originates in electromagnetic theory which, in addition to Poincaré invariance has a certain internal, gauge symmetry corresponding to the Lie group $U(1)$. This is not the type of symmetry we are discussing here (see Remark 2.1.6), but even so there are conservation laws associated with such symmetries and, for the electromagnetic case, Q turns out to be the familiar electric charge, while $\mathbf{j}$ is the electric current. In Section 2.3 we will describe an analogous situation for the complex Klein-Gordon field.

Notice that this version of the conservation of charge asserts that the total charge *in all of* $\mathbb{R}^3$ is constant in time. In particular, it allows for the rather inexplicable possibility that charge simply disappears from one region provided an equal amount appears at the same instant (in our fixed inertial frame) at the other end of the universe. There is, however, a much stronger local version that we will now describe. Let R be a bounded region in $\mathbb{R}^3$ with smooth, orientable boundary ∂R and $\mathbf{n}$ the outward unit normal vector field to ∂R. Define the *charge contained in R* at time t to be

$$Q_R(t) = \int_R j^0(t, \mathbf{x})\, d^3\mathbf{x}.$$

Then the Divergence Theorem gives

$$\frac{dQ_R(t)}{dt} = \int_R -div\, \mathbf{j}\, d^3\mathbf{x} = -\int_{\partial R} \mathbf{j} \cdot \mathbf{n}\, dS$$

so any change in the amount of charge contained in R must be accounted for by a flux of the current through the boundary of R. The charge in R cannot simply disappear, but must flow out.

For the examples to follow we want to consider the Klein-Gordon Lagrangian density $\mathcal{L}$ and the action of the Poincaré group $\mathcal{P}_+^\uparrow$ on $\mathbb{R}^{1,3}$. We know that, for every $g \in \mathcal{P}_+^\uparrow$, the diffeomorphism $x \mapsto g \cdot x = g(x)$ is a symmetry of $\mathcal{L}$ and, moreover, preserves the volume form on $\mathbb{R}^{1,3}$ since everything in $\mathcal{P}_+^\uparrow$ has Jacobian determinant one. We recall that one can produce infinitesimal symmetries from this action in the following way. Fix some A in the Lie algebra of $\mathcal{P}_+^\uparrow$ and consider the smooth curve $t \mapsto e^{tA}$ in $\mathcal{P}_+^\uparrow$. Each e^{tA} acts on $\mathbb{R}^{1,3}$ so we can define a smooth vector field X_A on $\mathbb{R}^{1,3}$ by

$$X_A(x) = \frac{d}{dt}(e^{tA} \cdot x)\Big|_{t=0} = X_A^\alpha(x)\,\partial_\alpha = x^\alpha(Ax)\,\partial_\alpha. \tag{2.11}$$

Then the unique integral curve of X_A through x at $t = 0$ is $t \mapsto e^{tA} \cdot x$ so X_A is complete and its 1-parameter group of diffeomorphisms $\{g_t : \mathbb{R}^{1,3} \to \mathbb{R}^{1,3}\}$ consists of the maps $g_t(x) = e^{tA} \cdot x$, each of which is a symmetry of $\mathcal{L}$. X_A is therefore an infinitesimal symmetry of $\mathcal{L}$. The information we will need regarding $\mathcal{P}_+^\uparrow$ and its Lie algebra will be drawn from Sections 2.4 and 2.5 of [Nab6].

Example 2.1.1. *Energy-Momentum*
We consider first the subgroup of $\mathcal{P}_+^\uparrow$ consisting of the translation group of $\mathbb{R}^{1,3}$. This is isomorphic to the additive group $\mathbb{R}^{1,3}$ and so its Lie algebra can also be identified with $\mathbb{R}^{1,3}$. More precisely, any A in the Lie algebra of the translation group is a real linear combination of the matrices $O_\alpha, \alpha = 0, 1, 2, 3$, described in Section 2.5.3 (page 71) of [Nab6], with coefficients $(a^0, a^1, a^2, a^3) \in \mathbb{R}^{1,3}$. The corresponding infinitesimal symmetry is therefore, by (2.11), just $a^\alpha \partial_\alpha$ and so the corresponding current satisfies

$$a^\beta \partial_\alpha \Big(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \partial_\beta \varphi - \delta^\alpha_\beta \mathcal{L} \Big) = 0.$$

But this must be satisfied for all (a^0, a^1, a^2, a^3) in $\mathbb{R}^{1,3}$ so we must have

$$\partial_\alpha \Big(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \partial_\beta \varphi - \delta^\alpha_\beta \mathcal{L} \Big) = 0, \;\; \beta = 0, 1, 2, 3.$$

Now define

$$T^\alpha{}_\beta = \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \partial_\beta \varphi - \delta^\alpha_\beta \mathcal{L}, \;\; \alpha, \beta = 0, 1, 2, 3 \tag{2.12}$$

so that

$$\partial_\alpha T^\alpha{}_\beta = 0, \;\; \beta = 0, 1, 2, 3.$$

It is often convenient to use instead the contravariant form

$$T^{\alpha\beta} = \eta^{\gamma\beta} T^{\alpha}{}_{\gamma} = \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \partial^\beta \varphi - \eta^{\alpha\beta} \mathcal{L}, \quad \alpha, \beta = 0, 1, 2, 3 \tag{2.13}$$

of $T^{\alpha}{}_{\beta}$. Since $T^{\alpha}{}_{\beta}$ and $T^{\alpha\beta}$ differ at most in a sign, $(T^{\alpha}{}_{\beta})$ and $(T^{\alpha\beta})$ are regarded as having the same physical content. Of course, one still has

$$\partial_\alpha T^{\alpha\beta} = 0, \quad \beta = 0, 1, 2, 3.$$

Consequently, for each fixed $\beta = 0, 1, 2, 3$, we obtain a conserved current

$$(T^{0\beta}, T^{1\beta}, T^{2\beta}, T^{3\beta}). \tag{2.14}$$

We will have something to say about the corresponding charges shortly.

In classical mechanics we found that time translation symmetry corresponded to conservation of energy and spatial translation symmetry corresponded to conservation of (linear) momentum (Appendix A.2 of [Nab6]), while in special relativity one is forced to conclude that energy and momentum are but different aspects of the same underlying physical quantity. Since translations in $\mathbb{R}^{1,3}$ involve both translations in time and translations in space we are motivated to identify the matrix $(T^{\alpha}{}_{\beta})$ with the energy-momentum content of the field φ. Indeed, $(T^{\alpha}{}_{\beta})$ is known in the physics literature as the *canonical energy-momentum tensor* of the field φ; more accurately, the $T^{\alpha}{}_{\beta}$ are the components of the energy-momentum tensor in the Lorentz frame in which we happen to be doing these calculations.

Remark 2.1.8. Special relativity not only merges *space* and *time* into *spacetime*, but also requires that various other classically distinct concepts, such as the energy and momentum of a particle, be regarded as merely different aspects of a single underlying physical concept. The reason is simply that different inertial observers agree, for example, on the energy-momentum 4-vector of a particle, but not on how much of it is energy and how much is momentum. For systems more complex than a single particle, such as a fluid or a field, there is more to this. One can think of a fluid, for example, as a huge swarm of particles that one might like to characterize by something like a mass-energy density. However, what is mass-energy density in one inertial frame will be some combination of mass-energy density and such things as energy flux density and momentum flux density (pressure) in another frame. A meaningful relativistic description must be in terms of an object that incorporates all of these. This object is called the energy-momentum tensor (or stress-energy tensor) and physicists have laid down a general scheme describing what they would like the components of this tensor to mean physically. However, defining the tensor in order to achieve these physical interpretations is a subtle business that necessarily involves very detailed knowledge of the physics of whatever situation is being modeled. The best way to get a sense of how this is done is to work through the definitions in some particularly simple case such as a so-called perfect fluid or an electromagnetic field. For those interested in pursuing this we mention that there is a very detailed discussion of the entire business in Minkowski spacetime in Chapter 5 of [MTW]. We will only point out that there are physical reasons for believing that such a tensor should be *symmetric* (see Section 5.7 of [MTW]) and that what we have called the canonical energy-momentum tensor *need not* have this property (although, as we will soon see, for the Klein-Gordon field it does). Consequently, this definition, although standard, is susceptible to serious physical objections and one should take care not to be entirely won over by the use of such words as energy and momentum in relation to it. This is a subject of much concern to physicists, but we will only refer those interested to the discussion in http://arxiv.org/abs/hep-th/0307199.

This *caveat* having been duly noted we proceed to introduce still more physical-sounding terminology. Notice that

$$T^{00} = T^0{}_0 = \frac{\partial \mathcal{L}}{\partial(\partial_0 \varphi)}\,\partial_0\varphi - \mathcal{L}.$$

This entry is called the *energy density* of φ in the inertial frame under consideration. Assuming its integral over $\mathbb{R}^3$ exists at any fixed time t,

$$P^0 = \int_{\mathbb{R}^3} T^{00}\, d^3\mathbf{x}$$

is the charge corresponding to the current $(T^{00}, T^{10}, T^{20}, T^{30})$. This is called the *total energy* of φ in the given inertial frame. This is just a definition, of course, and in the end definitions are justified only by their usefulness. For a bit of motivation, however, one can compare

$$T^{00} = \frac{\partial \mathcal{L}}{\partial(\partial_0 \varphi)}\,\partial_0\varphi - \mathcal{L}$$

with the expression

$$E_L = \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L$$

for the total energy in Lagrangian particle mechanics (see (A.7) of [Nab6]).

Assuming again that the integrals exist, the charges corresponding to the remaining currents $(T^{0k}, T^{1k}, T^{2k}, T^{3k}), k = 1, 2, 3$, are

$$P^k = \int_{\mathbb{R}^3} T^{0k}\, d^3\mathbf{x} = \int_{\mathbb{R}^3} \partial_0\varphi\, \partial^k\varphi\, d^3\mathbf{x}, \quad k = 1, 2, 3, \tag{2.15}$$

and are identified with the components of the *total momentum* of the field φ in the given frame. (P^0, P^1, P^2, P^3) are the components of the *total 4-momentum* of the field. At the risk of becoming tedious we emphasize once again that these are definitions, not theorems and one must look to the physics for motivation and/or justification. For more on the physical interpretation of the energy-momentum tensor we refer to Chapter 5 of [MTW] or Section 32 of [LaLi].

Notice that nothing we have done thus far depends on the specific form of the Lagrangian density $\mathcal{L}$, but only on its invariance with respect to translations in $\mathbb{R}^{1,3}$. Now we specialize to Klein-Gordon.

Exercise 2.1.6. Let $\mathcal{L} = \frac{1}{2}(\partial_\alpha\varphi\,\partial^\alpha\varphi - m^2\varphi^2)$.

1. Show that (with α labeling the row),

$$(T^\alpha{}_\beta) = \begin{pmatrix} (\partial_0\varphi)^2 - \mathcal{L} & \partial_0\varphi\,\partial_1\varphi & \partial_0\varphi\,\partial_2\varphi & \partial_0\varphi\,\partial_3\varphi \\ -\partial_0\varphi\,\partial_1\varphi & -(\partial_1\varphi)^2 - \mathcal{L} & -\partial_1\varphi\,\partial_2\varphi & -\partial_1\varphi\,\partial_3\varphi \\ -\partial_0\varphi\,\partial_2\varphi & -\partial_1\varphi\,\partial_2\varphi & -(\partial_2\varphi)^2 - \mathcal{L} & -\partial_2\varphi\,\partial_3\varphi \\ -\partial_0\varphi\,\partial_3\varphi & -\partial_1\varphi\,\partial_3\varphi & -\partial_2\varphi\,\partial_3\varphi & -(\partial_3\varphi)^2 - \mathcal{L} \end{pmatrix}$$

2. Show that $T^{\alpha\beta} = \partial^\alpha\varphi\,\partial^\beta\varphi - \eta^{\alpha\beta}\mathcal{L}$ so that $(T^{\alpha\beta})$ is symmetric. Write out the matrix $(T^{\alpha\beta})$ explicitly.
3. Show that the entries in each column of the matrix $(T^\alpha{}_\beta)$ are the components of a current 4-vector corresponding to some particular element of the Lie algebra of $\mathcal{P}_+^\uparrow$ and so each column has divergence zero.
4. Show that

$$T^{00} = \frac{\partial\mathcal{L}}{\partial(\partial_0\varphi)}\,\partial_0\varphi - \mathcal{L} = \frac{1}{2}\Big(\sum_{\alpha=0}^{3}(\partial_\alpha\varphi)^2 + m^2\varphi^2\Big)$$

and therefore

$$P^0 = \frac{1}{2}\int_{\mathbb{R}^3}\Big(\sum_{\alpha=0}^{3}(\partial_\alpha\varphi)^2 + m^2\varphi^2\Big)d^3\mathbf{x}.$$

provided the integral exists.

Exercise 2.1.7. Return now to the case of a general Lagrangian density $\mathcal{L}$. Assume that $T^{\alpha\beta}$ satisfies $T^{\alpha\beta} = T^{\beta\alpha}$ for all $\alpha, \beta = 0, 1, 2, 3$, and $\partial_\alpha T^{\alpha\beta} = 0$ for all $\beta = 0, 1, 2, 3$. Define $T^{\alpha\beta\gamma}$ by

$$T^{\alpha\beta\gamma} = x^\beta T^{\alpha\gamma} - x^\gamma T^{\alpha\beta}, \quad \alpha, \beta, \gamma = 0, 1, 2, 3.$$

Show that

$$\partial_\alpha T^{\alpha\beta\gamma} = 0, \quad \beta, \gamma = 0, 1, 2, 3.$$

$T^{\alpha\beta\gamma}$ are the components of the *canonical moment tensor* and will see their significance in the next example.

Example 2.1.2. *Angular Momentum*
Now we turn to the $\mathcal{L}_+^\uparrow$ subgroup of $\mathcal{P}_+^\uparrow$ and consider a Lagrangian density $\mathcal{L}$, such as

Klein-Gordon, that is invariant under the action of $\mathcal{L}_+^\uparrow$. The Lie algebra of $\mathcal{L}_+^\uparrow$ can be identified with the 4×4 matrices $(a^\alpha{}_\beta)$ for which $a_{\alpha\beta} = \eta_{\alpha\gamma} a^\gamma{}_\beta$ satisfies $a_{\beta\alpha} = -a_{\alpha\beta}$ (see Exercise 2.5.1 of [Nab6]). The corresponding infinitesimal symmetry is, by (2.11), $X = (a^\alpha{}_\beta x^\beta)\,\partial_\alpha$ so the associated current satisfies

$$\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}(a^\beta{}_\gamma x^\gamma)\partial_\beta\varphi - \delta^\alpha_\beta(a^\beta{}_\gamma x^\gamma)\mathcal{L}\right) = 0.$$

This is equivalent to each of the following.

$$a^\beta{}_\gamma\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}x^\gamma\partial_\beta\varphi - \delta^\alpha_\beta x^\gamma\mathcal{L}\right) = 0$$

$$\eta_{\mu\beta}a^\beta{}_\gamma\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}x^\gamma\eta^{\mu\beta}\partial_\beta\varphi - \eta^{\mu\beta}\delta^\alpha_\beta x^\gamma\mathcal{L}\right) = 0$$

$$a_{\mu\gamma}\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}x^\gamma\partial^\mu\varphi - \eta^{\mu\alpha}x^\gamma\mathcal{L}\right) = 0$$

Using the skew-symmetry of the $a_{\mu\gamma}$ we can write this as

$$\sum_{0\le\mu<\gamma\le3} a_{\mu\gamma}\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}(x^\gamma\partial^\mu\varphi - x^\mu\partial^\gamma\varphi) - (\eta^{\mu\alpha}x^\gamma - \eta^{\gamma\alpha}x^\mu)\mathcal{L}\right) = 0.$$

But the coefficients $a_{\mu\gamma}$ with $\mu < \gamma$ are arbitrary so we must have

$$\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}(x^\gamma\partial^\mu\varphi - x^\mu\partial^\gamma\varphi) - (\eta^{\mu\alpha}x^\gamma - \eta^{\gamma\alpha}x^\mu)\mathcal{L}\right) = 0 \tag{2.16}$$

for all $0 \le \mu < \gamma \le 3$.

Exercise 2.1.8. Let $T^{\alpha\beta} = \eta^{\gamma\beta}T^\alpha{}_\gamma = \frac{\partial\mathcal{L}}{\partial(\partial_\alpha\varphi)}\partial^\beta\varphi - \eta^{\alpha\beta}\mathcal{L}$, $\alpha,\beta = 0,1,2,3$, and consider the canonical moment tensor $T^{\alpha\gamma\mu}$ defined in Exercise 2.1.7. Show that (2.16) is equivalent to

$$\partial_\alpha T^{\alpha\gamma\mu} = 0, \quad \gamma,\mu = 0,1,2,3.$$

Consequently, for each fixed $\gamma,\mu = 0,1,2,3$, we have a conserved current

$$(T^{0\gamma\mu}, T^{1\gamma\mu}, T^{2\gamma\mu}, T^{3\gamma\mu})$$

and a corresponding charge

$$\int_{\mathbb{R}^3} T^{0\gamma\mu}\,d^3\mathbf{x} = \int_{\mathbb{R}^3}(x^\gamma T^{0\mu} - x^\mu T^{0\gamma})\,d^3\mathbf{x}$$

that will be conserved if the integrands decay sufficiently fast as $|\mathbf{x}| \to \infty$. We would like to make some specific choices for γ and μ and look for possible physical interpretations of the corresponding charges. First suppose $1 \le i < j \le 3$ and consider

$$J^{ij} = \int_{\mathbb{R}^3} (x^i T^{0j} - x^j T^{0i})\, d^3\mathbf{x}.$$

According to the scheme laid down by the physicists, each $T^{0k}, k = 1, 2, 3$, is a component of the momentum density of the field (see (2.15)). Consequently, $x^i T^{0j} - x^j T^{0i}$ bears a rather striking resemblance to a component of the familiar definition of angular momentum in classical mechanics (Example A.2.1 of [Nab6]). For this reason, J^{ij} is referred to as the k^{th}-component of the *total angular momentum* of the field φ, where $1 \le k \le 3, k \ne i, k \ne j$. In the following sense these three components are the conserved quantities associated with invariance under the rotation group.

Exercise 2.1.9. The calculations above leading to the conservation law $\partial_\alpha T^{\alpha\gamma\mu} = 0, \gamma, \mu = 0, 1, 2, 3$, were based on the invariance of the Lagrangian density under the entire group $\mathcal{L}_+^\uparrow$. Consequently, $(a^\alpha{}_\beta)$ represented an arbitrary element of the Lie algebra of $\mathcal{L}_+^\uparrow$. Repeat these computations, but assuming only invariance under the rotation subgroup of $\mathcal{L}_+^\uparrow$ so that $(a^\alpha{}_\beta)$ is some real linear combination of the generators M_1, M_2, and M_3 of rotations (see page 65 of [Nab6]). Show that one obtains in this way just three independent currents and that the resulting charges are precisely the components $J^{ij}, 1 \le i < j \le 3$, of the total angular momentum of φ.

Now take $\gamma = 0$ and $\mu = k = 1, 2, 3$. We obtain three charges

$$J^{0k} = \int_{\mathbb{R}^3} (x^0 T^{0k} - x^i T^{00})\, d^3\mathbf{x}, \quad k = 1, 2, 3.$$

Exercise 2.1.10. Carry out the procedure described in Exercise 2.1.9 with the generators N_1, N_2, and N_3 of the boosts in $\mathcal{L}_+^\uparrow$ (page 65 of [Nab6]) rather than M_1, M_2, and M_3 to show that there are just three independent currents and that the corresponding charges are $J^{0k}, k = 1, 2, 3$.

These conserved quantities $J^{0k}, k = 1, 2, 3$, associated with boosts do not appear to have been given any standard names and are rarely even mentioned in the physics literature. When they do put in an appearance it is generally in a somewhat different form that is described in the next exercise.

Exercise 2.1.11. Assuming the required integrability, the charges $J^{0k}, k = 1, 2, 3$, are conserved so their time derivatives are all zero. Write out

$$\frac{d}{dt} \int_{\mathbb{R}^3} (tT^{0k} - x^k T^{00})\, d^3\mathbf{x} = 0$$

explicitly to obtain, for each $k = 1, 2, 3$,

$$\frac{d}{dt}\int_{\mathbb{R}^3} x^k T^{00}\, d^3\mathbf{x} = P^k + t\,\frac{dP^k}{dt},$$

where P^k is given by 2.15. Now assume also that the Lagrangian density is invariant under translations and conclude that

$$\frac{d}{dt}\int_{\mathbb{R}^3} x^k T^{00}\, d^3\mathbf{x}$$

is constant for each $k = 1, 2, 3$. In physics this is taken to mean that *the center of energy of the field φ travels with constant velocity* and is thought of as an analogue of Newton's First Law.

2.2 Solutions to the Klein-Gordon Equation

We would now like to look at some explicit solutions to the Klein-Gordon equation $(\Box + m^2)\varphi = 0$. For these explicit formulas we will allow the solutions to be complex and just take real and imaginary parts if we want real solutions. We will be interested in solutions of various types, beginning with what we will call *strong*, or *classical solutions*; these are simply functions φ on $\mathbb{R}^{1,3}$ that are twice continuously differentiable with respect to each $x^\alpha, \alpha = 0, 1, 2, 3$, and satisfy the equation. A *distributional solution* to the Klein-Gordon equation is a tempered distribution φ that satisfies the Klein-Gordon equation when the derivatives are all interpreted as distributional derivatives. Equivalently, φ is a distributional solution to Klein-Gordon if it annihilates any Schwartz function in the image of the Klein-Gordon operator $\Box + m^2$, that is, if it satisfies

$$\langle\, \varphi, (\Box + m^2)\phi\, \rangle = 0 \;\; \forall \phi \in \mathcal{S}(\mathbb{R}^{1,3}).$$

Although we will tend to eschew the following notational convention, one often sees this condition written as an integral

$$\int_{\mathbb{R}^{1,3}} \varphi(x)\,(\Box + m^2)\,\phi(x)\,d^4x = 0 \;\; \forall \phi \in \mathcal{S}(\mathbb{R}^{1,3}),$$

whether or not the distribution φ happens to be regular.

We have seen that the natural action of $\mathcal{L}_+^\uparrow$ on functions carries classical solutions of Klein-Gordon onto other classical solutions (Section 2.1.1). There is also a natural action of $\mathcal{L}_+^\uparrow$ on the tempered distributions of $\mathbb{R}^{1,3}$. Specifically, we define $\Lambda \cdot \varphi$ for each $\Lambda \in \mathcal{L}_+^\uparrow$ by

$$\langle \Lambda \cdot \varphi, \phi \rangle = \langle \varphi, \Lambda \cdot \phi \rangle \ \ \forall \phi \in \mathcal{S}(\mathbb{R}^{1,3})$$

Notice, by the way, that $\Lambda \cdot \phi$ is a Schwartz function whenever $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$. A distribution φ is *invariant under* $\Lambda \in \mathcal{L}_+^\uparrow$ if $\Lambda \cdot \varphi = \varphi$, that is, if

$$\langle\, \varphi, \phi \circ \Lambda^{-1} \,\rangle = \langle\, \varphi, \phi \,\rangle$$

for every $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$ and *invariant under* $\mathcal{L}_+^\uparrow$ if this is true for every $\Lambda \in \mathcal{L}_+^\uparrow$. We will be looking for classical and distributional solutions to the Klein-Gordon equation and, in particular, those that are invariant under $\mathcal{L}_+^\uparrow$.

Now we begin our search for solutions as one does in the case of the electromagnetic field (Section 4.2 of [Nab5]) by looking for plane wave solutions in some fixed admissible basis for $\mathbb{R}^{1,3}$ of the form

$$\varphi(t, \mathbf{x}) = e^{i(\omega t - \mathbf{k}\cdot\mathbf{x})},$$

where $\omega \in \mathbb{R}$ and $\mathbf{k} \in \mathbb{R}^3$. These are all smooth, of course, so one can simply compute the derivatives and substitute into the Klein-Gordon equation. The result is

$$(-\omega^2 + \|\mathbf{k}\|^2 + m^2)e^{i(\omega t - \mathbf{k}\cdot\mathbf{x})} = 0$$

which can only occur if the *dispersion relation*

$$\omega^2 - \|\mathbf{k}\|^2 = m^2$$

is satisfied. Now recall that, for any plane wave, ω is interpreted physically as the wave's angular frequency and $\mathbf{k}$, the wave vector, points in the direction in which the wave is propagating and has magnitude equal to the wave's spatial frequency. Thus, $\omega/\|\mathbf{k}\|$ is the wave's speed, in the sense that this is the speed at which the crests of either $Re\,(\varphi(t, \mathbf{x}))$ or $Im\,(\varphi(t, \mathbf{x}))$ travel.

Now we interject some physics into the picture by applying the Planck-Einstein-de Broglie relations (see page 132 of [Nab5]) to the plane wave φ, that is, we define $p^0 = \hbar\omega = E$, $\mathbf{p} = \hbar\mathbf{k}$ and $p = (p^0, p^1, p^2, p^3) = (p^0, \mathbf{p})$. In natural units $\hbar = 1$ so we conclude that $p = (\omega, \mathbf{k})$. Thus, $\omega = E$ and $\mathbf{k} = \mathbf{p}$ and our plane wave solution is

$$\varphi(x) = \varphi(t, \mathbf{x}) = e^{i(Et - \mathbf{p}\cdot\mathbf{x})} = e^{ip_\alpha x^\alpha}, \tag{2.17}$$

where $p_\alpha = \eta_{\alpha\beta}p^\beta$. The dispersion relation now becomes

$$p_\alpha p^\alpha = m^2.$$

In order that φ be a scalar solution to the Klein-Gordon equation the dispersion relation must be satisfied in every frame of reference. To ensure this we define the components of p in any other admissible frame in such a way that p transforms as a 4-vector, that is, $\hat{p}^\alpha = \Lambda^\alpha{}_\beta p^\beta$ if the new frame is given by $\hat{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta$. Then

$\hat{p}_\alpha \hat{x}^\alpha = p_\alpha x^\alpha$ and $\hat{p}_\alpha \hat{p}^\alpha = p_\alpha p^\alpha = m^2$. We call p the *energy-momentum 4-vector* or simply the *4-momentum* of the plane wave φ. Thus, the 4-momentum of a plane wave solution to the Klein-Gordon equation is restricted to the mass hyperboloid $X_m = \{p \in \mathbb{P}^{1,3} : p_\alpha p^\alpha = m^2\}$; it can be either future directed $p^0 > 0$ and so lie in X_m^+ or past directed and so lie in X_m^-. When $p \in X_m^+$ we will call the plane wave φ a *positive frequency solution* or *positive energy solution*, whereas if $p \in X_m^-$ it is a *negative frequency solution* or *negative energy solution*.

Notice that these plane wave solutions are not in $L^2(\mathbb{R}^{1,3})$ since $|\, e^{i(Et-\mathbf{p}\cdot\mathbf{x})}\,|^2 = 1$ which is not integrable over $\mathbb{R}^{1,3}$. Nevertheless, if ϕ is any Schwartz function on $\mathbb{R}^{1,3}$, $e^{i(Et-\mathbf{p}\cdot\mathbf{x})}\phi(x)$ *is* integrable on $\mathbb{R}^{1,3}$ so $e^{i(Et-\mathbf{p}\cdot\mathbf{x})}$ can be identified with the tempered distribution whose value at any $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$ is

$$\langle\, e^{i(Et-\mathbf{p}\cdot\mathbf{x})}, \phi \,\rangle = \int_{\mathbb{R}^{1,3}} e^{i(Et-\mathbf{p}\cdot\mathbf{x})}\phi(x)\, d^4x = \int_{\mathbb{R}^{1,3}} e^{ip_\alpha x^\alpha}\phi(x)\, d^4x.$$

Exercise 2.2.1. Show that $e^{i(Et-\mathbf{p}\cdot\mathbf{x})}$, regarded as an element of $\mathcal{S}'(\mathbb{R}^{1,3})$, is a distributional solution to the Klein-Gordon equation.

Remark 2.2.1. Although we have already pointed out that there are a number reasons to doubt that the Klein-Gordon equation is the appropriate relativistic analogue of the Schrödinger equation, let us put those aside for a moment and try to regard the plane wave solution φ of Klein-Gordon as a wave function for a single, relativistic particle of mass m. Not being in $L^2(\mathbb{R}^3)$ for fixed t, the plane wave solutions are not normalizable, but they are instructive and, in a sense to be described shortly, every solution is a superposition of these plane waves. We will assume that the particle associated with φ has the same energy-momentum 4-vector p as the plane wave. Now notice that the dispersion relation implies

$$E = \pm\sqrt{m^2 + \|\mathbf{p}\|^2}.$$

In particular, this allows for the possibility that the total energy E of the particle can be negative. These *negative energy states* present rather substantial difficulties for the physical interpretation and one would like to simply ignore them on physical grounds. Indeed, this is precisely what is done in classical (pre-quantum) special relativity where $E^2 = m^2 + \|\mathbf{p}\|^2$ is simply taken to mean $E = \sqrt{m^2 + \|\mathbf{p}\|^2}$. In quantum mechanics, however, the situation is not so simple and one can argue that there is no particularly compelling reason to assume that negative energy solutions can simply be ignored. This is a subtle business and still not resolved to everyone's satisfaction so we will be content with a few simple remarks. We mentioned earlier that a general real Klein-Gordon field can be written as a superposition of plane waves. Specifically, one finds in the physics literature explicit formulas such as the following for such a field.

$$\varphi(t,\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} \frac{1}{2\omega_{\mathbf{p}}} \left(e^{i(\omega_{\mathbf{p}} t - \mathbf{p}\cdot\mathbf{x})}\, a(\mathbf{p}) + e^{-i(\omega_{\mathbf{p}} t - \mathbf{p}\cdot\mathbf{x})}\, \overline{a}(\mathbf{p}) \right) d^3\mathbf{p} \tag{2.18}$$

Here $\omega_{\mathbf{p}} = \sqrt{m^2 + \|\mathbf{p}\|^2}$ is positive so the expansion of $\varphi(t,\mathbf{x})$ requires both positive and negative energy plane waves (we will have more to say about where such a formula might come from as we proceed). Without negative energy solutions one therefore cannot describe all of the real Klein-Gordon fields and so will generally not be able to describe the evolution of the field from initial data. It turns out, in fact, that when one includes terms in the equation that represent an interaction with an electromagnetic field an initially positive energy solution will necessarily evolve into a negative energy state (through the emission of photons, for example). These issues still exist and are more acute and more physically relevant in the case of the Dirac equation where rather elaborate schemes have been devised to resolve them. For more on the physics behind all of this one should consult a book on relativistic quantum mechanics, for example, [BD1] or [Gri].

Now we set about solving the Klein-Gordon equation

$$(\square + m^2)\,\varphi = 0$$

by applying the Fourier transform to both sides. Since the functions and distributions of interest are defined on Minkowski spacetime and we are looking for solutions invariant under $\mathcal{L}_+^\uparrow$ we will use the Minkowski-Fourier transform $\mathcal{F}_M$ introduced in Appendix A. For clarity we will also revert to the notation of Section 2.6.1 of [Nab6] and write $\mathbb{P}^{1,3}$ for the copy of $\mathbb{R}^{1,3}$ on which $\mathcal{F}_M\phi = \tilde{\phi}$ is defined for any $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$. Recall that, for $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$, $\mathcal{F}_M$ is given by

$$(\mathcal{F}_M\phi)(p) = \tilde{\phi}(p) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^{1,3}} e^{-i\langle p,x\rangle} \phi(x)\, d^4x,$$

where $\langle p, x\rangle = p_\alpha x^\alpha$ is the Lorentz inner product, and then the definition is extended to $L^2(\mathbb{R}^{1,3})$ and to the tempered distributions on $\mathbb{R}^{1,3}$ in the same way as the usual Fourier transform. On $\mathcal{S}(\mathbb{P}^{1,3})$ the inverse Minkowski-Fourier transform is given by

$$(\mathcal{F}_M^{-1}\psi)(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^{1,3}} e^{i\langle x,p\rangle} \psi(p)\, d^4p.$$

To apply the Fourier transform to both sides of the Klein-Gordon equation we must assume that $(\square + m^2)\,\varphi$ *has* a Fourier transform and therefore is, at least, a tempered distribution. Consequently, we will be working in the context of $\mathcal{S}'(\mathbb{R}^{1,3})$ and will sort out at the end what degree of smoothness the solutions we construct actually have. The following two exercises describe the advantages of $\mathcal{F}_M$ over the usual Fourier transform.

Exercise 2.2.2. Show that, for every $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$,

$$\mathcal{F}_M(\Lambda \cdot \phi) = \Lambda \cdot (\mathcal{F}_M\phi)$$

so, in particular, ϕ is invarint under $\mathcal{L}_+^\uparrow$ if and only if its Minkowski-Fourier transform $\mathcal{F}_M\phi$ is invariant under $\mathcal{L}_+^\uparrow$, that is,

$$\Lambda \cdot \phi = \phi \Leftrightarrow \Lambda \cdot (\mathcal{F}_M\phi) = \mathcal{F}_M\phi.$$

Also prove the same results for the inverse Fourier transform $\mathcal{F}_M^{-1}$. *Hint*: Apply the Change of Variables Formula (Theorem 2.6.1 of [Nab6]). Prove and use the fact that, with respect to the Lorentz inner product, the adjoint of $\Lambda \in \mathcal{L}_+^\uparrow$ is Λ^{-1}.

Exercise 2.2.3. Show that a tempered distribution $\varphi \in \mathcal{S}'(\mathbb{R}^{1,3})$ is invariant under $\Lambda \in \mathcal{L}_+^\uparrow$ if and only if its Minkowski-Fourier transform is invariant under Λ, that is,

$$\Lambda \cdot \varphi = \varphi \Leftrightarrow \Lambda \cdot (\mathcal{F}_M\varphi) = \mathcal{F}_M\varphi.$$

For $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$ the basic properties of the Fourier transform (Appendix A) imply that

$$\mathcal{F}_M\,[\,(\Box + m^2)\phi\,] = (-p^2 + m^2)\,\mathcal{F}_M\phi = (-p^2 + m^2)\tilde{\phi} \tag{2.19}$$

where $p^2 = p_\alpha p^\alpha$.

Exercise 2.2.4. Show that that (2.19) remains valid for distributions, that is,

$$\mathcal{F}_M\,[\,(\Box + m^2)\varphi\,] = (-p^2 + m^2)\,\mathcal{F}_M\varphi = (-p^2 + m^2)\tilde{\varphi} \tag{2.20}$$

for every $\varphi \in \mathcal{S}'(\mathbb{R}^{1,3})$. In particular,

$$(\Box + m^2)\varphi = 0 \Leftrightarrow (p^2 - m^2)\,\tilde{\varphi} = 0$$

so, in momentum space $\mathbb{P}^{1,3}$, the Klein-Gordon equation takes the form

$$(p^2 - m^2)\,\tilde{\varphi} = 0. \tag{2.21}$$

If $\tilde{\varphi}$ is a smooth function, then (2.21) implies that it must vanish on the complement of the mass hyperboloid $X_m = \{p \in \mathbb{P}^{1,3} : p_\alpha p^\alpha = m^2\}$. A distribution is said to *vanish on an open set U* if it takes the value zero on any Schwartz function ϕ whose support is contained in U and the *support* of the distribution is the complement of largest open set on which it vanishes.

Exercise 2.2.5. Show that any distribution $\tilde{\varphi}$ on $\mathbb{P}^{1,3}$ satisfying (2.21) vanishes on the complement $\mathbb{P}^{1,3} - X_m$ of the mass hyperboloid. Conclude that the value of $\tilde{\varphi}$ on any Schwartz function is uniquely determined by the restriction of the Schwartz function to X_m.

What we would like to do then is to find distributions on momentum space satisfying (2.21) and then apply the inverse Fourier transform $\mathcal{F}_M^{-1}$ to obtain solutions to the Klein-Gordon equation $(\Box + m^2)\varphi = 0$ on $\mathbb{R}^{1,3}$. It should be clear from the previous exercise that the mass hyperboloid X_m will figure heavily in this so we will pause to make a few observations. X_m is clearly a smooth submanifold of $\mathbb{R}^{1,3}$ with two connected components $X_m = X_m^+ \sqcup X_m^-$. To ease the exposition a bit we will restrict most of our remarks to the upper branch X_m^+ of the hyperboloid. By definition, the action of $\mathcal{L}_+^\uparrow$ on $\mathbb{P}^{1,3}$ carries X_m^+ to itself. Indeed, X_m^+ is the complete orbit of any one of its points under this action (see page 95 of [Nab6]). Being the graph in $\mathbb{P}^{1,3}$ of the function $h : \mathbb{R}^3 \to \mathbb{R}$ defined by $h(\mathbf{p}) = h(p_1, p_2, p_3) = \omega_{\mathbf{p}} = \sqrt{m^2 + \|\mathbf{p}\|^2}$, X_m^+ has a global chart given by the projection $\pi_+(\omega_{\mathbf{p}}, p_1, p_2, p_3) = (p_1, p_2, p_3)$. In particular, X_m^+ is diffeomorphic to $\mathbb{R}^3$. We will denote by $\iota_+ : X_m^+ \hookrightarrow \mathbb{P}^{1,3}$ the inclusion map of X_m^+ into $\mathbb{P}^{1,3}$ and by $s_+ : \mathbb{R}^3 \to X_m^+$ the diffeomorphism $s_+(\mathbf{p}) = (\omega_{\mathbf{p}}, \mathbf{p})$.

Exercise 2.2.6. Show that the restriction $\iota_+^*\eta$ of the Minkowski metric on $\mathbb{P}^{1,3}$ to X_m^+ is a Riemannian metric because every tangent vector to X_m^+ in $\mathbb{P}^{1,3}$ is spacelike.

With this metric X_m^+ has constant sectional curvature $-\frac{1}{m^2}$ (see the Proposition on page 113, Chapter 4, of [O'N]) and we will refer to it as *m-hyperbolic space*. When $m = 1$ it is known simply as *hyperbolic space* and is one of the three famous models of hyperbolic geometry. We have already seen (Section 2.6.2 of [Nab6]) that X_m^+ admits a Lorentz invariant measure μ_m defined as follows. For any Borel set $B \subseteq X_m^+$, $\pi_+(B)$ is a Borel subset of $\mathbb{R}^3$ so we can define

$$\mu_m(B) = \int_{\pi_+(B)} \frac{d^3\mathbf{p}}{2\omega_{\mathbf{p}}} = \int_{\pi_+(B)} \frac{d^3\mathbf{p}}{2\sqrt{m^2 + \|\mathbf{p}\|^2}}\,,$$

where $d^3\mathbf{p} = dp_1\, dp_2\, dp_3$ denotes integration with respect to Lebesgue measure on $\mathbb{R}^3$.

Now we return to the problem of solving the Klein-Gordon equation. We have seen that solutions arise as inverse Fourier transforms of distributions on $\mathbb{P}^{1,3}$ that vanish on the complement of the mass hyperboloid X_m^+. The Lorentz invariant measure μ_m on X_m^+ that we have just introduced provides a means of constructing one such distribution. It is called the *Dirac delta supported on* X_m^+, usually denoted $\delta_+(p^2 - m^2)$, and defined at any Schwartz function $\phi \in \mathcal{S}(\mathbb{P}^{1,3})$ by restricting ϕ to X_m^+ and integrating the restriction with respect to μ_m, that is,

$$\langle\, \delta_+(p^2 - m^2),\, \phi \,\rangle = \int_{X_m^+} \iota_+^*\phi\, d\mu_m = \int_{\mathbb{R}^3} \frac{\phi(\sqrt{m^2 + \|\mathbf{p}\|^2}\,, p_1, p_2, p_3)}{2\sqrt{m^2 + \|\mathbf{p}\|^2}}\, dp_1 dp_2 dp_3. \tag{2.22}$$

The distribution $\delta_+(p^2 - m^2)$ is invariant under $\mathcal{L}_+^\uparrow$ because the measure μ_m is Lorentz invariant.

Remark 2.2.2. We point out once again the custom of writing the value of $\delta_+(p^2 - m^2)$ at ϕ as an integral

$$\int_{\mathbb{P}^{1,3}} \phi(p)\delta_+(p^2 - m^2)\, d^4p \tag{2.23}$$

despite the fact $\delta_+(p^2 - m^2)$ is not a regular distribution.

Notice that the integrand in (2.22) is actually a Schwartz function on $\mathbb{R}^3$. In fact, we can define a continuous map $\Pi_+ : \mathcal{S}(\mathbb{P}^{1,3}) \to \mathcal{S}(\mathbb{R}^3)$ from the test functions on $\mathbb{P}^{1,3}$ to the test functions on $\mathbb{R}^3$ by

$$\Pi_+(\phi(p_0, p_1, p_2, p_3)) = \frac{\phi(\sqrt{m^2 + \|\mathbf{p}\|^2}\,, p_1, p_2, p_3)}{2\sqrt{m^2 + \|\mathbf{p}\|^2}}.$$

Since $\delta_+(p^2 - m^2)$ is certainly supported on X_m^+ its inverse Fourier transform $\mathcal{F}_M^{-1}[\delta_+(p^2 - m^2)]$ is a solution to the Klein-Gordon equation on $\mathbb{R}^{1,3}$ that is invariant under $\mathcal{L}_+^\uparrow$. Computing this inverse transform explicitly requires quite a lot of work and the result is a rather complicated formula for a distribution on $\mathbb{R}^{1,3}$ involving Hankel and modified Bessel functions. Since we will not require the result we will simply refer those interested in seeing this done to [deJag], where the calculations are carried out in great detail.

Remark 2.2.3. Everything we have just done for the upper branch X_m^+ of the mass hyperboloid X_m could equally well be done for the lower branch X_m^-. In particular, we have a Dirac delta $\delta_-(p^2 - m^2)$ supported on X_m^- given by

$$\langle\, \delta_-(p^2 - m^2),\, \phi\, \rangle = \int_{X_m^-} \iota_-^* \phi\, d\mu_m = \int_{\mathbb{R}^3} \frac{\phi(-\sqrt{m^2 + \|\mathbf{p}\|^2}\,, p_1, p_2, p_3)}{2\sqrt{m^2 + \|\mathbf{p}\|^2}}\, dp_1 dp_2 dp_3$$

and a corresponding invariant solution $\mathcal{F}_M^{-1}[\delta_-(p^2 - m^2)]$ to the Klein-Gordon equation as well as a projection $\Pi_- : \mathcal{S}(\mathbb{P}^{1,3}) \to \mathcal{S}(\mathbb{R}^3)$ given by

$$\Pi_-(\phi(p_0, p_1, p_2, p_3)) = \frac{\phi(-\sqrt{m^2 + \|\mathbf{p}\|^2}\,, p_1, p_2, p_3)}{2\sqrt{m^2 + \|\mathbf{p}\|^2}}.$$

We can also combine $\delta_+(p^2 - m^2)$ and $\delta_-(p^2 - m^2)$ into a Dirac delta $\delta(p^2 - m^2)$ supported on X_m

$$\delta(p^2 - m^2) = \delta_+(p^2 - m^2) + \delta_-(p^2 - m^2),$$

where the sum is defined pointwise on $\mathcal{S}(\mathbb{P}^{1,3})$.

Exercise 2.2.7. 1. Show that any Schwartz function of the form $(p^2 - m^2)\phi$ with $\phi \in \mathcal{S}(\mathbb{P}^{1,3})$ satisfies

$$\Pi_\pm((p^2 - m^2)\phi) = 0.$$

2. Let η_+ and η_- be two arbitrary tempered distributions on $\mathbb{R}^3$. Define a tempered distribution $\tilde{\varphi}$ on $\mathbb{P}^{1,3}$ by

$$\langle\, \tilde{\varphi}, \phi \,\rangle = \langle\, \eta_+, \Pi_+(\phi) \,\rangle + \langle\, \eta_-, \Pi_-(\phi) \,\rangle$$

for every $\phi \in \mathcal{S}(\mathbb{P}^{1,3})$. Show that $(p^2 - m^2)\tilde{\varphi} = 0$.

Remark 2.2.4. Conversely, *every* distribution $\tilde{\varphi}$ on $\mathbb{P}^{1,3}$ satisfying $(p^2 - m^2)\tilde{\varphi} = 0$ is of the form $\langle\, \eta_+, \Pi_+(\phi) \,\rangle + \langle\, \eta_-, \Pi_-(\phi) \,\rangle$ for some η_+, η_- and ϕ. This is proved in Section 3.3 of [deJag].

The previous Exercise and Remark describe all of the distributional solutions to $(p^2 - m^2)\tilde{\varphi} = 0$ on $\mathbb{P}^{1,3}$. One of the major results of [deJag], also proved in Section 3.3 of that paper, is that every distribution on $\mathbb{P}^{1,3}$ satisfying $(p^2 - m^2)\tilde{\varphi} = 0$ and invariant under $\mathcal{L}_+^\uparrow$ is a constant linear combination of $\delta_+(p^2 - m^2)$ and $\delta_-(p^2 - m^2)$, that is,

$$\tilde{\varphi} = c_+\delta_+(p^2 - m^2) + c_-\delta_-(p^2 - m^2). \tag{2.24}$$

This effectively describes all of the Lorentz invariant solutions to the momentum space form of the Klein-Gordon equation and therefore, by taking inverse Fourier transforms $\mathcal{F}_M^{-1}$, all of the Lorentz invariant distributional solutions to the Klein-Gordon equation. Thus, from (2.24) we conclude that every Lorentz invariant solution to the Klein-Gordon equation on $\mathbb{R}^{1,3}$ is of the form

$$c_+\mathcal{F}_M^{-1}[\delta_+(p^2 - m^2)] + c_-\mathcal{F}_M^{-1}[\delta_-(p^2 - m^2)],$$

where c_+ and c_- are constants. We will not give the rather involved proof of this here (see [deJag] for this as well as explicit computations of the inverse transforms). Instead we will conclude by writing out some (not necessarily Lorentz invariant) solutions to the Klein-Gordon equation in the hope of illustrating how formulas such as (2.18) arise.

Begin by selecting a smooth, complex-valued function $c_+ : X_m^+ \to \mathbb{C}$ with compact support on X_m^+. Extend c_+ arbitrarily to a smooth, complex-valued function $C_+ : \mathbb{P}^{1,3} \to \mathbb{C}$ on $\mathbb{P}^{1,3}$ with compact support.

Exercise 2.2.8. Describe a procedure for constructing such an extension C_+. *Hint*: First extend c_+ to $\mathbb{P}^{1,3} \cong X_m^+ \times \mathbb{R}$ and then multiply by an appropriate cut-off function (see Exercise 2-26 of [Sp1]).

Since C_+ has compact support the distribution $C_+\delta_+(p^2-m^2)$ is defined, is given by

$$\langle\, C_+\delta_+(p^2-m^2),\, \phi\,\rangle = \langle\, \delta_+(p^2-m^2),\, C_+\phi\,\rangle = \int_{X_m^+} \iota_+^*(\,C_+\phi\,)\,d\mu_m = \int_{X_m^+} c_+\iota_+^*\phi\,d\mu_m$$

and has compact support in X_m^+.

Exercise 2.2.9. Let $\tilde{\varphi}$ be a tempered distribution on $\mathbb{P}^{1,3}$ with compact support $K\subseteq X_m$. By definition, $\tilde{\varphi}$ is defined on every element of $\mathcal{S}(\mathbb{P}^{1,3})$. Show that $\tilde{\varphi}$ extends to a map $\tilde{\varphi}'$ defined on all of $C^\infty(\mathbb{P}^{1,3})$. *Hint*: Choose a bounded open set U in $\mathbb{P}^{1,3}$ with $K\subseteq U$ and a smooth real-valued ("bump") function f on $\mathbb{P}^{1,3}$ that is 1 on K and 0 on $\mathbb{P}^{1,3}-U$ (see Exercise 2-26 of [Sp1]). Define $\tilde{\varphi}'(g)=\tilde{\varphi}(fg)$ for any smooth function g on $\mathbb{P}^{1,3}$ and show that the definition is independent of the choice of f.

According to Exercise 2.2.9, $C_+\delta_+(p^2-m^2)$ extends to a map on the smooth functions on $\mathbb{P}^{1,3}$ (we will omit the prime and use the same symbol for this extension). Fix an $x=(x^0,x^1,x^2,x^3)=(t,\mathbf{x})$ in $\mathbb{R}^{1,3}$ and consider the smooth function of p defined by $(2\pi)^{-3/2}e^{i\langle x,p\rangle}$, where the $\langle x,p\rangle = x^\alpha p_\alpha$. Now we evaluate

$$\begin{aligned}\langle\, C_+\delta_+(p^2-m^2),\,(2\pi)^{-3/2}e^{i\langle x,p\rangle}\,\rangle &= (2\pi)^{-3/2}\langle\,\delta_+(p^2-m^2),\,C_+e^{i\langle x,p\rangle}\,\rangle\\ &= (2\pi)^{-3/2}\int_{X_m^+}\iota_+^*(C_+e^{i\langle x,p\rangle})\,d\mu_m\\ &= (2\pi)^{-3/2}\int_{X_m^+} c_+(\omega_{\mathbf{p}},\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d\mu_m\\ &= (2\pi)^{-3/2}\int_{\mathbb{R}^3}\frac{1}{2\omega_{\mathbf{p}}}\,(c_+\circ s_+)(\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d^3\mathbf{p}.\end{aligned}$$

Notice that this is a smooth function of x. We claim that it is, in fact, the inverse Fourier transform of the distribution $\sqrt{2\pi}\,C_+\delta_+(p^2-m^2)$ and therefore a solution to the Klein-Gordon equation. It is instructive to check directly that this integral defines a solution to the Klein-Gordon equation so we will leave this to you.

Exercise 2.2.10. Verify by direct differentiation that

$$\varphi_+(t,\mathbf{x}) = (2\pi)^{-3/2}\int_{\mathbb{R}^3}\frac{1}{2\omega_{\mathbf{p}}}\,c_+(\omega_{\mathbf{p}},\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d^3\mathbf{p}$$

satisfies the Klein-Gordon equation. Being a superposition of positive frequency plane wave solutions, we will refer to φ_+ as a *positive frequency solution* or *positive energy solution* to the Klein-Gordon equation.

To show that $\mathcal{F}_M^{-1}[\,\sqrt{2\pi}\,C_+\delta_+(p^2-m^2)] = (2\pi)^{-3/2}\int_{\mathbb{R}^3}\frac{1}{2\omega_{\mathbf{p}}}\,(c_+\circ s_+)(\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d^3\mathbf{p}$ we proceed as follows. *A priori* this inverse transform is a distribution so we must treat the function defined by the integral as a (regular) distribution as well. Now, the value of the distribution

$$(2\pi)^{-3/2}\int_{\mathbb{R}^3}\frac{1}{2\omega_{\mathbf{p}}}\,(c_+\circ s_+)(\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d^3\mathbf{p} = (2\pi)^{-3/2}\int_{X_m^+} c_+(\omega_{\mathbf{p}},\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d\mu_m$$

on a Schwartz function $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$ is

$$\begin{aligned}\int_{\mathbb{R}^{1,3}}\Big[(2\pi)^{-3/2}\int_{X_m^+} c_+(\omega_{\mathbf{p}},\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d\mu_m\Big]\phi(x)\,d^4x \\ = (2\pi)^{-3/2}\int_{X_m^+}\int_{\mathbb{R}^{1,3}} c_+(\omega_{\mathbf{p}},\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,\phi(x)\,d^4x\,d\mu_m.\end{aligned}$$

On the other hand, the definition of the inverse Fourier transform of the distribution $\sqrt{2\pi}\,C_+\delta_+(p^2-m^2)$ gives, for each $\phi \in \mathcal{S}(\mathbb{R}^{1,3})$,

$$\begin{aligned}\langle\,\mathcal{F}_M^{-1}[\,\sqrt{2\pi}\,C_+\delta_+(p^2-m^2)],\,\phi\,\rangle &= \langle\,\sqrt{2\pi}\,C_+\delta_+(p^2-m^2),\,\mathcal{F}_M^{-1}[\phi]\,\rangle \\ &= \sqrt{2\pi}\,\langle\,\delta_+(p^2-m^2),\,C_+\mathcal{F}_M^{-1}[\phi]\,\rangle \\ &= \sqrt{2\pi}\int_{X_m^+}\iota_+^*(C_+\mathcal{F}_M^{-1}[\phi])\,d\mu_m \\ &= \sqrt{2\pi}\int_{X_m^+} c_+(\omega_{\mathbf{p}},\mathbf{p})\iota_+^*(\mathcal{F}_M^{-1}[\phi])\,d\mu_m.\end{aligned}$$

But

$$\begin{aligned}\iota_+^*(\mathcal{F}_M^{-1}[\phi]) &= \iota_+^*\Big(\frac{1}{(2\pi)^2}\int_{\mathbb{R}^{1,3}} e^{i\langle x,p\rangle}\phi(x)\,d^4x\Big) \\ &= \frac{1}{(2\pi)^2}\int_{\mathbb{R}^{1,3}} e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\phi(x)\,d^4x\end{aligned}$$

so

$$\begin{aligned}\langle\,\mathcal{F}_M^{-1}[\,\sqrt{2\pi}\,C_+\delta_+(p^2-m^2)],\,\phi\,\rangle &= \langle\,\sqrt{2\pi}\,C_+\delta_+(p^2-m^2),\,\mathcal{F}_M^{-1}[\phi]\,\rangle \\ &= \sqrt{2\pi}\int_{X_m^+} c_+(\omega_{\mathbf{p}},\mathbf{p})\Big[\frac{1}{(2\pi)^2}\int_{\mathbb{R}^{1,3}} e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\phi(x)\,d^4x\Big]d\mu_m \\ &= (2\pi)^{-3/2}\int_{X_m^+}\int_{\mathbb{R}^{1,3}} c_+(\omega_{\mathbf{p}},\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,\phi(x)\,d^4x\,d\mu_m\end{aligned}$$

and therefore, as distributions,

$$\mathcal{F}_M^{-1}[\,\sqrt{2\pi}\,C_+\delta_+(p^2-m^2)] = (2\pi)^{-3/2}\int_{\mathbb{R}^3}\frac{1}{2\omega_{\mathbf{p}}}\,(c_+\circ s_+)(\mathbf{p})\,e^{i(\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x})}\,d^3\mathbf{p}.$$

Remark 2.2.5. Once again, everything we have just done for the upper branch X_m^+ of the mass hyperboloid X_m can equally well be done for the lower branch X_m^- by simply beginning with a smooth function c_- on X_m^- with compact support. This will give rise to an analogous solution $\varphi_-(t, \mathbf{x})$ to the Klein-Gordon equation.

Exercise 2.2.11. Let $c_- : X_m^- \to \mathbb{C}$ be a smooth function with compact support and show that

$$\varphi_-(t, \mathbf{x}) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \frac{1}{2\omega_{\mathbf{p}}}\, c_-(-\omega_{\mathbf{p}}, -\mathbf{p})\, e^{-i(\omega_{\mathbf{p}} t - \mathbf{p}\cdot\mathbf{x})}\, d^3\mathbf{p}$$

satisfies the Klein-Gordon equation. Being a superposition of negative frequency plane wave solutions, we will refer to φ_- as a *negative frequency solution* or *negative energy solution* to the Klein-Gordon equation.

More generally, one can add a positive frequency solution $\varphi_+(t, \mathbf{x})$ and a negative frequency solution $\varphi_-(t, \mathbf{x})$ to obtain a solution $\varphi(t, \mathbf{x}) = \varphi_+(t, \mathbf{x}) + \varphi_-(t, \mathbf{x})$ that is a superposition of both positive and negative frequencies. Then φ_+ is referred to as the *positive frequency part* of φ and φ_- is its *negative frequency part*.

Now notice that, since s_+ is a diffeomorphism of $\mathbb{R}^3$ onto X_m^+,

$$a = c_+ \circ s_+$$

can be regarded as an arbitrary smooth function with compact support of $\mathbb{R}^3$. We can therefore write

$$\varphi_+(t, \mathbf{x}) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \frac{1}{2\omega_{\mathbf{p}}}\, a(\mathbf{p})\, e^{i(\omega_{\mathbf{p}} t - \mathbf{p}\cdot\mathbf{x})}\, d^3\mathbf{p},$$

where $a(\mathbf{p})$ is an arbitrary smooth function with compact support on $\mathbb{R}^3$. To obtain a real solution we add the conjugate and arrive at

$$\varphi(t, \mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} \frac{1}{2\omega_{\mathbf{p}}} \left(e^{i(\omega_{\mathbf{p}} t - \mathbf{p}\cdot\mathbf{x})}\, a(\mathbf{p}) + e^{-i(\omega_{\mathbf{p}} t - \mathbf{p}\cdot\mathbf{x})}\, \overline{a}(\mathbf{p}) \right) d^3\mathbf{p} \tag{2.25}$$

which is just (2.18).

Remark 2.2.6. The function $a(\mathbf{p})$ on $\mathbb{R}^3$ need not have compact support in order for (2.25) to be a solution to the Klein-Gordon equation. All that is required is enough smoothness and a sufficiently rapid decay as $\|\mathbf{p}\| \to \infty$ to justify differentiating under the integral sign. For example, it is enough for $a(\mathbf{p})$ to be a Schwartz function (a notion that is not defined on general manifolds such as $X_m^\pm$).

We should also point out that one often finds in the literature integral representations for real solutions that differ from (2.25) in various powers of 2π and $\omega_{\mathbf{p}}$. All of these can be obtained from (2.25) by renormalizing $a(\mathbf{p})$ and we will have occasion to make one such cosmetic change a bit later.

Notice that we have certainly not shown that every solution to the Klein-Gordon equation is of the form (2.18). Indeed, all of the solutions we have found of this form are smooth. The essential point to be taken from our discussion is that solutions to Klein-Gordon on $\mathbb{R}^{1,3}$ are inverse Fourier transforms of distributions supported on the mass hyperboloid and that, for the particularly nice distributions we have been discussing, there is an explicit integral formula for these inverse transforms.

2.3 Complex Klein-Gordon Fields

Real and complex Klein-Gordon fields have different physical interpretations and the purpose of this brief section is to try to get to the root of this difference. We let φ denote a complex-valued solution to $(\Box + m^2)\varphi = 0$. Write φ in terms of real and imaginary parts as

$$\varphi = \frac{1}{\sqrt{2}}\,(\varphi^1 + i\varphi^2)$$

(the $2^{-1/2}$ is conventional). Then, since the Klein-Gordon equation is linear, φ^1 and φ^2 are clearly real solutions to $(\Box + m^2)\varphi = 0$ and so

$$\overline{\varphi} = \frac{1}{\sqrt{2}}\,(\varphi^1 - i\varphi^2)$$

is another complex solution. Furthermore,

$$\varphi^1 = \frac{1}{\sqrt{2}}\,(\varphi + \overline{\varphi})$$

and

$$\varphi^2 = \frac{-i}{\sqrt{2}}\,(\varphi - \overline{\varphi}).$$

Conversely, if φ^1 and φ^2 are two arbitrary real solutions to $(\Box + m^2)\varphi = 0$, then $\varphi = \frac{1}{\sqrt{2}}\,(\varphi^1 + i\varphi^2)$ and $\overline{\varphi} = \frac{1}{\sqrt{2}}\,(\varphi^1 - i\varphi^2)$ are two complex solutions. Complex solutions are therefore nothing more or less than pairs of real solutions and occasionally it is convenient to think of them explicitly in this way by employing vector notation.

$$\varphi = \frac{1}{\sqrt{2}}\begin{pmatrix}\varphi^1\\ \varphi^2\end{pmatrix}$$

We have defined the Klein-Gordon Lagrangian density $\mathcal{L}$ for *real-valued* functions φ by

$$\mathcal{L}(\varphi) = \frac{1}{2}\,(\partial_\alpha\varphi\,\partial^\alpha\varphi - m^2\varphi^2).$$

Since $\mathcal{L}$ must be real-valued, this clearly will not do if φ is complex. To handle the complex case we write $\varphi = \frac{1}{\sqrt{2}}\,(\varphi^1 + i\varphi^2)$ and define

$$\mathcal{L}(\varphi) = \mathcal{L}(\varphi^1) + \mathcal{L}(\varphi^2) = \frac{1}{2}\Big(\partial_\alpha\varphi^1\,\partial^\alpha\varphi^1 + \partial_\alpha\varphi^2\,\partial^\alpha\varphi^2 - m^2[\,(\varphi^1)^2 + (\varphi^2)^2\,]\Big).$$

Exercise 2.3.1. Show that

$$\mathcal{L}(\varphi) = \partial_\alpha\varphi\,\partial^\alpha\,\overline{\varphi} - m^2\varphi\overline{\varphi}. \tag{2.26}$$

Notice that the Klein-Gordon Lagrangian density (2.26) clearly has a $U(1)$-symmetry that is not present in the real case. Indeed, if $g = e^{i\theta}$, $\theta \in \mathbb{R}$, is any complex number of modulus one, then $\varphi \to g\cdot\varphi = e^{i\theta}\varphi$ leaves $\mathcal{L}$ invariant because

$$\begin{aligned}\mathcal{L}(g\cdot\varphi) &= \partial_\alpha(g\cdot\varphi)\,\partial^\alpha\,\overline{(g\cdot\varphi)} - m^2(g\cdot\varphi)(\overline{g\cdot\varphi})\\ &= (e^{i\theta}\partial_\alpha\varphi)\,(e^{-i\theta}\partial^\alpha\,\overline{\varphi}) - m^2(e^{i\theta}\varphi)(e^{-i\theta}\,\overline{\varphi})\\ &= \partial_\alpha\varphi\,\partial^\alpha\,\overline{\varphi} - m^2\varphi\overline{\varphi}\\ &= \mathcal{L}(\varphi).\end{aligned}$$

If we identify $U(1)$ with SO(2) via

$$\begin{pmatrix} e^{i\theta} \end{pmatrix} \to \begin{pmatrix} cos\,\theta & -sin\,\theta \\ sin\,\theta & cos\,\theta \end{pmatrix}$$

and identify the complex function $\varphi = \frac{1}{\sqrt{2}}(\varphi^1 + i\varphi^2)$ with the pair

$$\frac{1}{\sqrt{2}}\begin{pmatrix}\varphi^1\\ \varphi^2\end{pmatrix}$$

we have an SO(2)-action under which

$$\mathcal{L}(\varphi) = \frac{1}{2}\Big(\partial_\alpha\varphi^1\,\partial^\alpha\varphi^1 + \partial_\alpha\varphi^2\,\partial^\alpha\varphi^2 - m^2[\,(\varphi^1)^2 + (\varphi^2)^2\,]\Big)$$

is invariant.

We would now like to apply the version of Noether's Theorem appropriate to such internal symmetries (see Remark 2.1.6). SO(2) is 1-dimensional and therefore so is its Lie algebra $\mathfrak{so}(2)$. A generator of $\mathfrak{so}(2)$ is

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Consequently, we will have one infinitesimal internal symmetry

$$Y = Y^1\partial_1 + Y^2\partial_2$$

and one corresponding conserved current

$$j^\alpha = -Y^1 \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi^1)} - Y^2 \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi^2)}, \quad \alpha = 0, 1, 2, 3.$$

Exercise 2.3.2. Show that

$$j^\alpha = \varphi^1 \partial^\alpha \varphi^2 - \varphi^2 \partial^\alpha \varphi^1 = i\,(\varphi\,\partial^\alpha \overline{\varphi} - \overline{\varphi}\,\partial^\alpha \varphi), \quad \alpha = 0, 1, 2, 3.$$

As we have seen (in the discussion following Remark 2.1.6) one can now associate with the complex Klein-Gordon field $\varphi = \frac{1}{\sqrt{2}}(\varphi^1 + i\varphi^2)$ a conserved charge

$$\int_{\mathbb{R}^3} j^0(t, \mathbf{x})\, d^3\mathbf{x} = \int_{\mathbb{R}^3} (\varphi^1 \partial^0 \varphi^2 - \varphi^2 \partial^0 \varphi^1)\, d^3\mathbf{x} = i \int_{\mathbb{R}^3} (\overline{\varphi}\,\partial^0 \varphi - \varphi\,\partial^0 \overline{\varphi})\, d^3\mathbf{x}$$

(assuming appropriate integrability and decay conditions). In this sense, *complex Klein-Gordon fields are charged*. This charge is the result of a $U(1)$-symmetry possessed by complex fields, but not by real fields so real Klein-Gordon fields are *uncharged*, or *neutral*. Notice that a complex field φ and its conjugate $\overline{\varphi}$ have opposite charge so they represent distinct, independent physical fields; after quantization these correspond to a particle and its antiparticle. Since all of these particles have one (real or complex) component they have spin 0 and are therefore bosons (see Remark 9.5 of [Nab5]).

The charge associated with a complex Klein-Gordon field is a $U(1)$-charge and, to this extent at least, is analogous to ordinary electric charge. Whether or not it should be identified with electric charge is a question that we simply cannot answer at this stage. Electric charge is a coupling constant; it determines the strength of an interaction with the electromagnetic field. Quite aside from the fact that we have not discussed interactions at all, one should keep in mind that there is no known physical particle corresponding to a classical Klein-Gordon field and therefore nothing to watch interact with an electromagnetic field.

Remark 2.3.1. There is much more to be said about Klein-Gordon fields from the point of view of physics. For those interested in pursuing this we suggest Chapter 9 of [BD1], Sections 2.2, 3.2, and 3.3 of [Ryd], or Chapter 1 of [Gri].

2.4 Classical Klein-Gordon as a Hamiltonian System

2.4.1 Introduction

We would now like to turn from the Lagrangian to the Hamiltonian view of classical field theory. Needless to say, this involves a generalization of the Hamiltonian picture of classical mechanics which is discussed in some detail in Section 2.3 of [Nab5] and is summarized in Appendix A.3 of [Nab6]. Field theory requires an infinite-dimensional version of this and here, as one might expect, there are technical issues to be addressed so we will proceed more slowly. We will consider only the linear case (specifically, the Klein-Gordon field) where these technical issues are less severe. As usual, we will provide specific references to the arguments we do not include, but some general sources are [AM], [AMR], [Arn2], [ChM1], [Mar2], and [GS1].

2.4.2 Motivation: Heat Flow and Bilinear Forms

The underlying mathematical structure of finite-dimensional Hamiltonian mechanics is easy to describe. Each possible state of a physical system is represented by a point in a symplectic manifold (P, ω) called its phase space. The total energy of the system in any state is given by a smooth, real-valued function $H : P \to \mathbb{R}$ on phase space which, together with ω, determines a Hamiltonian vector field X_H whose flow completely describes the time evolution of the state. This flow is determined by Hamilton's equations, which are *ordinary* differential equations on phase space. It so happens that, with a few technical adjustments, the same basic philosophy can be employed to describe the evolution of certain physical systems governed by *partial* differential equations. The technical adjustments amount to moving to an infinite-dimensional phase space and finding appropriate infinite-dimensional analogues of symplectic forms, Hamiltonian vector fields and flows. This is what we would like to do here for the classical real Klein-Gordon field. Our basic references for this section are [ChM1], [Evans], [Mar2], and [Mar3], but for some material on semigroups of operators and the Hille-Yosida Theorem (reviewed in Appendix B) we will also refer to Sections X.8 and X.9 of [RS1], Chapter 7 of [Pazy], Chapter 34 of [Lax], and Chapters IX and XIV of [Yos].

Before proceeding with the formal definitions we will try to motivate some of what is coming by revisiting an old friend from [Nab5], where we had a brief encounter with the heat flow on $\mathbb{R}$ (Example 5.2.12 and Example 8.4.8 of [Nab5]). Although this system is not "Hamiltonian" in the sense we just described, it sheds much light on many of the issues we will need to address. We consider the Cauchy problem

$$\begin{aligned} &\frac{\partial\psi}{\partial t} - \Delta\psi = 0, \quad (x,t) \in \mathbb{R} \times (0,\infty) \\ &\lim_{t\to 0^+} \psi(x,t) = \psi_0(x), \quad x \in \mathbb{R}, \end{aligned} \tag{2.27}$$

for the heat equation, where we write Δ for the 1-dimensional spatial Laplacian (second derivative with respect to x). There exists a strongly continuous, contractive semigroup $\{T_t\}_{t\geq 0}$ of bounded operators on $L^2(\mathbb{R})$ which has the property that, if the initial state ψ_0 is in $L^2(\mathbb{R})$, then the state at time t is $\psi_t(x) = \psi(x,t) = (T_t\psi_0)(x)$; see Exercise 5.2.18 of [Nab5] and also Example 8.4.8 of [Nab5]. This is called the *heat semigroup* and it describes the "flow" of the solution from its initial state to its state at time t in precisely the same way that the flow of a Hamiltonian vector field describes the evolution of the state in classical mechanics.

Remark 2.4.1. Working in $L^2(\mathbb{R})$ is just a sly way of introducing boundary conditions at infinity, that is, growth conditions on the solutions as $|x| \to \infty$. Other such growth conditions are certainly possible and when we formulate the general definitions we will allow for this by assuming only that the flow takes place in a Banach space (such as some other L^p-space). L^2 is particularly natural for our purposes since the energy of a field configuration is generally defined to be the sum of various squared L^2-norms.

The question we would now like to address is whether or not the "flow" on $L^2(\mathbb{R})$ described by the semigroup $\{T_t\}_{t\geq 0}$ can be regarded as, in some sense, the flow of a "vector field" on $L^2(\mathbb{R})$. Here matters become a bit more ticklish and we would like to explain why. For any $u \in L^2(\mathbb{R})$, $u(t) = T_t u$ describes a curve in $L^2(\mathbb{R})$. Ideally, one would like to regard this as an integral curve of the "vector field" we are hoping to define. In the best of all possible worlds $u(t)$ would be differentiable and the heat equation $\frac{\partial u}{\partial t} = \Delta u$ would be equivalent to

$$\frac{du}{dt} = \Delta u,$$

where $\frac{du}{dt} = \lim_{h\to 0} \frac{T_{t+h}u - T_t u}{h}$ with the limit computed in the Hilbert space $L^2(\mathbb{R})$.

Remark 2.4.2. Take careful note of the fact that $\frac{du}{dt}$ and $\frac{\partial u}{\partial t}$ are very different t-derivatives (see Exercise 6.2.1 of [Nab5] and the Remark preceding it).

In this form the equation is analogous to a first order linear equation $\frac{d\mathbf{x}}{dt} = A\mathbf{x}$ on $\mathbb{R}^n$, where A is some $n \times n$ matrix (that is, linear operator on $\mathbb{R}^n$). In this finite-dimensional context we can think of A as a vector field on $\mathbb{R}^n$ whose value at any point $\mathbf{x}$ is $A\mathbf{x}$ and then $\frac{d\mathbf{x}}{dt} = A\mathbf{x}$ is the equation for its integral curves, which always exist. But now the issues we must deal with come into focus. The Laplacian Δ is an *unbounded* linear operator on $L^2(\mathbb{R})$, defined only on the dense, linear subspace

$H^2(\mathbb{R})$ of $L^2(\mathbb{R})$ so our equation makes sense only on this subspace. Moreover, even if $u \in H^2(\mathbb{R})$, Δu is generally not in $H^2(\mathbb{R})$, but only in $L^2(\mathbb{R})$ and is therefore not "tangent to $H^2(\mathbb{R})$" in the usual sense (the tangent space to a linear space is that same linear space). This suggests that a "vector field" on $L^2(\mathbb{R})$ will generally be defined only on a dense linear subspace and will need to take values outside this subspace. In this case, it may not be so clear what one means by the "flow" of the "vector field" and it is certainly not clear that, whatever it means, it must exist. These difficulties are resolved (to the extent that they can be resolved) by the general theory of semigroups of operators and the famous Hille-Yosida Theorem and we have reviewed what we need of this in Appendix B.

There is one more item that deserves a bit of preliminary discussion. On a finite-dimensional manifold P a symplectic form ω is, by definition, nondegenerate. This means that, at each point $p \in P$, the mapping $v_p \to \iota_{v_p}\omega = \omega(v_p, \cdot)$ is an isomorphism of $T_p(P)$ onto $T_p^*(P)$. Because both of these vector spaces are of the same finite-dimension this is equivalent to the map being injective. However, we will not be operating in finite-dimensions here so "injective" is no longer the same thing as "isomorphism" and it so happens that, for many of the interesting examples, "isomorphism" is too much to ask. To describe some of these examples, however, we first need to set the stage .

We consider a *real* Banach space $\mathcal{E}$. Suppose we are given a continuous, skew-symmetric, bilinear form $\Omega : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ on $\mathcal{E}$. Then Ω determines a 2-form ω on $\mathcal{E}$ in the following way. For any $e \in \mathcal{E}$ the tangent space $T_e(\mathcal{E})$ to $\mathcal{E}$ at e is canonically identified with $\mathcal{E}$ and we will not distinguish between their elements notationally. Now define

$$\omega_e : T_e(\mathcal{E}) \times T_e(\mathcal{E}) \to \mathbb{R}$$

by

$$\omega_e(v, w) = \Omega(v, w), \quad \forall v, w \in \mathcal{E}.$$

Since ω is constant in e, it is both smooth and closed and it is nondegenerate exactly when the bilinear form Ω has the property that the map

$$\Omega^\flat : \mathcal{E} \to \mathcal{E}^*$$

defined by

$$\Omega^\flat(v) = \iota_v\Omega = \Omega(v, \cdot)$$

is injective. We will see in the next Lemma that, if the Banach space $\mathcal{E}$ is not reflexive (Section 4.10 of [Fried]), then $\Omega^\flat$ *cannot* be an isomorphism of Banach spaces.

Lemma 2.4.1. *Let $\mathcal{E}$ be a real Banach space and $\Omega : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ a continuous, skew-symmetric, bilinear form on $\mathcal{E}$. If the map $\Omega^\flat : \mathcal{E} \to \mathcal{E}^*$ is an isomorphism, then $\mathcal{E}$ is reflexive.*

Proof. We assume that $\Omega^\flat$ is an isomorphism and show that $\mathcal{E}$ is reflexive, that is, that the natural inclusion $\kappa : \mathcal{E} \to \mathcal{E}^{**}$, of $\mathcal{E}$ into its second dual $\mathcal{E}^{**}$ is a Banach space isomorphism of $\mathcal{E}$ onto $\mathcal{E}^{**}$.

Remark 2.4.3. Recall that $\kappa(e) = e^{**}$, where $e^{**}(\beta) = \beta(e)$ for every $\beta \in \mathcal{E}^*$. Recall also that, if $\mathcal{E}$ is reflexive, then κ is actually an isometric isomorphism of $\mathcal{E}$ onto $\mathcal{E}^{**}$ (Theorem 4.10.2 of [Fried]).

Since $\Omega^\flat : \mathcal{E} \to \mathcal{E}^*$ is an isomorphism, so is $(\Omega^\flat)^{-1} : \mathcal{E}^* \to \mathcal{E}$.

Exercise 2.4.1. Show that the map $((\Omega^\flat)^{-1})^* : \mathcal{E}^* \to \mathcal{E}^{**}$, which takes $\alpha \in \mathcal{E}^*$ to $((\Omega^\flat)^{-1})^*\alpha \in \mathcal{E}^{**}$ defined by

$$[\,((\Omega^\flat)^{-1})^*\alpha\,](\beta) = \alpha(\,(\Omega^\flat)^{-1}\beta\,) \ \ \forall \beta \in \mathcal{E}^*$$

is also an isomorphism and therefore

$$-((\Omega^\flat)^{-1})^* \circ \Omega^\flat : \mathcal{E} \to \mathcal{E}^{**}$$

is an isomorphism. Complete the proof by showing that $-((\Omega^\flat)^{-1})^* \circ \Omega^\flat = \kappa$.

□

Next we will describe a general procedure for manufacturing such examples that we will put to use in the next section.

Example 2.4.1. Let $\mathcal{B}$ be an arbitrary real Banach space, $\mathcal{B}^*$ its dual and $\mathcal{E} = \mathcal{B} \oplus \mathcal{B}^*$. Define $\Omega : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ by

$$\Omega(\,(a,\alpha),(b,\beta)\,) = \beta(a) - \alpha(b), \ \ \forall a,b \in \mathcal{B} \ \text{ and } \ \forall \alpha,\beta \in \mathcal{B}^*.$$

Exercise 2.4.2. Show that Ω is bilinear, skew-symmetric, and continuous and that $\Omega^\flat$ is injective. *Hint*: For the injectivity of $\Omega^\flat$ use the fact that the continuous linear functionals on a Banach space separate points; this is a consequence of the Hahn-Banach Theorem (Corollary 4.8.5 of [Fried]).

We will show now that $\Omega^\flat$ is surjective and therefore an isomorphism by the Open Mapping Theorem (Theorem 4.6.1 of [Fried]) if and only if $\mathcal{B}$ is reflexive. For this we note that, for any $(a,\alpha) \in \mathcal{E} = \mathcal{B} \oplus \mathcal{B}^*$,

$$\Omega^\flat(a,\alpha) = \iota_{(a,\alpha)}\Omega = \Omega(\,(a,\alpha),(\,\cdot\,,\,\cdot\,)\,).$$

For any such (a,α) in $\mathcal{E}$, $(-\alpha, \kappa(a)) = (-\alpha, a^{**})$, acting on $\mathcal{B} \oplus \mathcal{B}^*$ coordinatewise, is in $\mathcal{E}^*$. Specifically,

$$(-\alpha,\kappa(a))(b,\beta) = -\alpha(b) + a^{**}(\beta) = \beta(a) - \alpha(b) = [\,\iota_{(a,\alpha)}\Omega\,](b,\beta) = [\,\Omega^\flat(a,\alpha)\,](b,\beta).$$

Consequently,

$$\Omega^\flat(a,\alpha) = (-\alpha,\kappa(a)).$$

In particular, $\Omega^\flat$ is surjective if and only if κ is surjective, that is, if and only if $\mathcal{B}$ is reflexive.

It is worth recording a special case of this example.

Example 2.4.2. Let $\mathcal{H}$ be a real Hilbert space. Then $\mathcal{H}$ is reflexive and, moreover, we can canonically identify $\mathcal{H}^*$ with $\mathcal{H}$ by the Riesz Representation Theorem (Theorem 6.2.4 of [Fried]). Let $\mathcal{E} = \mathcal{H} \oplus \mathcal{H}^* = \mathcal{H} \oplus \mathcal{H}$ and define $\Omega : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ by

$$\Omega(\,(e_1,e_2),(f_1,f_2)\,) = \langle e_1, f_2\rangle_{\mathcal{H}} - \langle e_2, f_1\rangle_{\mathcal{H}}.$$

Then Ω is a continuous, skew-symmetric bilinear form on $\mathcal{H}\oplus\mathcal{H}$ and $\Omega^\flat : \mathcal{H}\oplus\mathcal{H} \to \mathcal{H}^* \oplus \mathcal{H}^* = \mathcal{H} \oplus \mathcal{H}$ is a linear homeomorphism.

2.4.3 Infinite-Dimensional Hamiltonian Systems

With the examples from the preceding section as motivation we can now proceed with the general definitions. Let $\mathcal{E}$ be a real Banach space. Any continuous, bilinear form $B : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ induces a continuous linear map $B^\flat : \mathcal{E} \to \mathcal{E}^*$ defined by $B^\flat(e) = \iota_e B = B(e,\,\cdot\,)$. B is said to be *weakly nondegenerate* if $B^\flat$ is injective and B is *strongly nondegenerate* if $B^\flat$ is an isomorphism of Banach spaces. By the Open Mapping Theorem (Theorem 4.6.1 of [Fried]), B is strongly nondegenerate if and only if it is weakly nondegenerate and $B^\flat$ is surjective. A *weak symplectic form* on $\mathcal{E}$ is a 2-form ω on $\mathcal{E}$ such that

1. ω is closed ($d\omega = 0$), and
2. for each $e \in \mathcal{E}$, $\omega_e : T_e(\mathcal{E}) \times T_e(\mathcal{E}) \to \mathbb{R}$ is weakly nondegenerate.

If, in (2), each ω_e is strongly nondegenerate, then ω is a *strong symplectic form* (often called simply a *symplectic form*).

Example 2.4.3. The information in Examples 2.4.1 and 2.4.2 provides a wide class of examples, both weak and strong. If $\mathcal{B}$ is an arbitrary real Banach space, then one defines a (constant) 2-form ω on $\mathcal{B} \oplus \mathcal{B}^*$ by $\omega(\,(a,\alpha),(b,\beta)\,) = \beta(a) - \alpha(b)$ at each point of $\mathcal{B} \oplus \mathcal{B}^*$ and for all $(a,\alpha),(b,\beta) \in \mathcal{B} \oplus \mathcal{B}^*$. If $\mathcal{B}$ is reflexive, then ω is a strong symplectic form on $\mathcal{B} \oplus \mathcal{B}^*$. If $\mathcal{B}$ is not reflexive, then ω is a weak (but not strong) symplectic form on $\mathcal{B} \oplus \mathcal{B}^*$. When $\mathcal{B}$ is a real Hilbert space $\mathcal{H}$ one identifies $\mathcal{H}^*$ with $\mathcal{H}$ by the Riesz Representation Theorem (Theorem 6.2.4 of

[Fried]) and the strong symplectic form ω on $\mathcal{H}$ is given, at each point of $\mathcal{H} \oplus \mathcal{H}$, by $\omega(\,(e_1, e_2), (f_1, f_2)\,) = \langle e_1, f_2 \rangle_{\mathcal{H}} - \langle e_2, f_1 \rangle_{\mathcal{H}}$ for all $(e_1, e_2), (f_1, f_2) \in \mathcal{H} \oplus \mathcal{H}$. The special case in which $\mathcal{H} = L^2(\mathbb{R}^3; \mathbb{R})$ will put in an appearance soon when we return to the Klein-Gordon field.

Remark 2.4.4. $L^2(\mathbb{R}^3; \mathbb{R})$ is the real Hilbert space of real-valued square integrable functions on $\mathbb{R}^3$.

The first thing one would probably like to know about an infinite-dimensional strong/weak symplectic form is whether or not there is an infinite-dimensional version of the Darboux Theorem (Theorem A.3.2 of [Nab6]). The answer is yes/no. In 1969, Weinstein [Weins] used an idea of Moser [Mos] to give a remarkably simple and elegant proof that any strong symplectic form on a Banach space is locally constant.

Theorem 2.4.2. *Let ω be a strong symplectic form on the Banach space $\mathcal{E}$. Then, for each $e \in \mathcal{E}$, there exists an open neighborhood U_e of e in $\mathcal{E}$, a constant symplectic form ω_0 on $\mathcal{E}$, an open neighborhood V_0 of $0 \in \mathcal{E}$, and a diffeomorphism F of U_e onto V_0 such that $F^*\omega_0 = \omega$.*

Remark 2.4.5. This immediately implies the same result for Banach manifolds and so, in particular, for finite-dimensional manifolds (of even dimension, of course). A bit of linear algebra then gives the usual coordinate form of the finite-dimensional Darboux Theorem. Proofs of Theorem 2.4.2 are available in [Weins], on pages 29-30 of [Mar2] and on page 535 of [AMR]; one can also consult Theorem 8.1 of [Lang3]. On the other hand, in 1972, Marsden [Mar1] constructed examples of weak symplectic forms on Hilbert spaces that are not locally constant. One can prove the local constancy of certain types of weak sympletic forms, but additional hypotheses are required (see, for example, Theorem 1.2 of [Mar3]).

The examples of most interest to us here are only weak symplectic forms and this is the source of some of the technical issues we will need to confront in this section. We begin by trying to isolate the appropriate phase space and weak symplectic form for the Hamiltonian formulation of Klein-Gordon theory. These are, of course, choices we must make and it would be naive to think that the choices are uniquely determined and can be "derived" in any meaningful sense. Nevertheless, it is the physics that must point us in the right direction and that is where we will start.

The phase space of a physical system is, in a very real sense, only the mathematical arena in which the discussion takes place; physicists very often do not mention it at all. The *physics* is contained in the Hamiltonian (total energy). We will therefore adopt the point of view that the phase space should be regarded as just the natural domain of the Hamiltonian and so what we should do is write down this Hamiltonian and let it tell us where it wants to be defined. Fine, but how do you write down Hamiltonians? This is actually a nontrivial question and it is a question for physicists, not mathematicians. However, one can find a hint in Section 2.3 of [Nab5] .

In classical mechanics one often arrives at the Hamiltonian $H(q, p)$ from the energy function $E_L(q, \dot{q})$ associated with the Lagrangian $L(q, \dot{q})$ by introducing conjugate momenta $p = \partial L/\partial \dot{q}$ and rewriting E_L in terms of q and p, that is, by moving E_L from the tangent bundle to the cotangent bundle via the Legendre transformation. The result is that $E_L(q, \dot{q}) = \frac{\partial L}{\partial \dot{q}}\dot{q} - L(q, \dot{q})$ becomes $H(q, p) = p\dot{q} - L(q, \dot{q})$, where it is understood that the $\dot{q}$ are now written in terms of q and p by solving $p = \partial L/\partial \dot{q}$ for $\dot{q}$.

The usual intuition behind moving from the formalism of mechanics to the formalism of field theory goes something like this. The discrete index i labeling the "degrees of freedom" (the coordinates in the configuration space M) is replaced by the continuous position variable x (one degree of freedom for each position); the coordinates q themselves (functions of t along the trajectory) are replaced, at each fixed value of t, by the value $\varphi(t, x)$ of the field φ at time t and location x; the $\dot{q}$ (also functions of t along the trajectory) are replaced, at each fixed value of t, by the time derivatives $\partial_0 \varphi$ of the coordinates (often denoted $\dot{\varphi}$ in this context). Let's see what happens if we just formally apply this admittedly rather heuristic algorithm in the case of the Klein-Gordon field (which is precisely what physicists do in practice).

We already know that the Lagrangian density for the Klein-Gordon field is

$$\mathcal{L}(\varphi, \partial_0\varphi, \partial_1\varphi, \partial_2\varphi, \partial_3\varphi) = \frac{1}{2}\big((\partial_0\varphi)^2 - (\partial_1\varphi)^2 - (\partial_2\varphi)^2 - (\partial_3\varphi)^2 - m^2\varphi^2 \big).$$

Now define the *conjugate momentum density* associated with $\mathcal{L}$ by

$$\frac{\partial \mathcal{L}}{\partial(\partial_0\varphi)} = \partial_0\varphi \stackrel{def}{=} \dot{\varphi}.$$

Writing $\nabla\varphi$ for the spatial gradient $(\partial_1\varphi, \partial_2\varphi, \partial_3\varphi)$ one obtains, as the analogue of the Hamiltonian in mechanics, the *Hamiltonian density*

$$\mathcal{H}(\varphi, \dot{\varphi}) = \frac{1}{2}\big(\dot{\varphi}^2 + \nabla\varphi \cdot \nabla\varphi + m^2\varphi^2 \big). \tag{2.28}$$

The spatial integral

$$H(\varphi, \dot{\varphi}) = \int_{\mathbb{R}^3} \mathcal{H}(\varphi, \dot{\varphi})\, d^3\mathbf{x} = \frac{1}{2}\int_{\mathbb{R}^3} \big(\dot{\varphi}^2 + \nabla\varphi \cdot \nabla\varphi + m^2\varphi^2 \big)\, d^3\mathbf{x} \tag{2.29}$$

of the Hamiltonian density is then called simply the *Hamiltonian*.

Needless to say, there is nothing that pretends to be a proof here. We have arrived, in the manner of the physicists, at a potential candidate for the Hamiltonian of Klein-Gordon theory. We must now see if we can fit this candidate neatly into a rigorous view of an infinite-dimensional Hamiltonian system. The first step involves a somewhat subtle change in perspective, but one that we have seen before in mechanics. A solution $\varphi(t, \mathbf{x})$ to the Klein-Gordon equation is a function of both t and $\mathbf{x}$ and, with this interpretation, H, as defined by (2.29) is a function of t. However, our goal is to regard the time evolution of the system as a flow along the integral curves of a vector field on phase space and in this interpretation t enters only as a parameter along these integral curves. For this we must look again at (2.29) and forget where φ and $\dot{\varphi}$ came from so that we can view them as simply independent variables in some appropriate spaces of functions on $\mathbb{R}^3$ (just as we forgot, in Section 2.3 of [Nab5], where $q, \dot{q}$ and q, p came from and regarded them as coordinates on the tangent and cotangent bundles, respectively). But what are these "appropriate spaces of functions on $\mathbb{R}^3$" ? Notice that (2.29) is just a sum of squared $L^2(\mathbb{R}^3;\mathbb{R})$ norms. The first term indicates that $\dot{\varphi}$ must therefore be in $L^2(\mathbb{R}^3;\mathbb{R})$ for each t. The second and third terms require that both φ and its gradient $\nabla\varphi$ must also be in $L^2(\mathbb{R}^3;\mathbb{R})$ for each t and so we find that φ must be in $H^1(\mathbb{R}^3;\mathbb{R})$ for each t. Consequently,

$$(\varphi, \dot{\varphi}) \in H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$$

and H can be regarded as simply a smooth real-valued function on $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$, provided $\nabla\varphi$ is now taken to be the distributional gradient (see Appendix A). We would like to adopt $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ as our phase space, but this will only make sense if $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ admits at least a weak symplectic form. It does, as we will now show.

Example 2.4.4. Notice that $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ is a dense linear subspace of $L^2(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$, although $H^1(\mathbb{R}^3;\mathbb{R})$ and $L^2(\mathbb{R}^3;\mathbb{R})$ do not have the same inner products. We already know that $L^2(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ admits a natural strong symplectic form $\omega : L^2(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R}) \to \mathbb{R}$ given, at every point, by $\omega(\,(e_1, e_2), (f_1, f_2)\,) = \langle e_1, f_2\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} - \langle e_2, f_1\rangle_{L^2(\mathbb{R}^3;\mathbb{R})}$ for all $(e_1, e_2), (f_1, f_2) \in L^2(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$. We will use the same formula to define ω at every point of $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ by

$$\begin{aligned}\omega(\,(\varphi, \dot{\varphi}), (\psi, \dot{\psi})\,) &= \langle \varphi, \dot{\psi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} - \langle \dot{\varphi}, \psi\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} \\ &= \int_{\mathbb{R}^3} (\varphi\dot{\psi} - \dot{\varphi}\psi)\, d^3\mathbf{x}\end{aligned}$$

for all $(\varphi, \dot{\varphi}), (\psi, \dot{\psi}) \in H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$. This is clearly (at least) a weak symplectic form on $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$. As you will now show, that's all it is.

Exercise 2.4.3. Argue as in Example 2.4.1 to show that $\omega^\flat(\varphi, \dot{\varphi}) = (-\dot{\varphi}, \varphi)$ and conclude that $\omega^\flat$ is not surjective so ω is not a strong symplectic form on $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$.

Although we have not yet fully justified the terminology, we are sufficiently encouraged by what we have seen thus far to refer to $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ as the *Klein-Gordon phase space*. The rest of the program consists of finding a Hamiltonian vector field on this phase space whose flow describes the time evolution of the Klein-Gordon field (as determined by the Klein-Gordon equation). Our discussion of the heat equation in Section 2.4.2, however, suggests that we will require a new notion of "vector field", defined on only a dense linear subspace and not necessarily taking values in this subspace, and that the notion of a "flow" for such a vector field may become problematic.

To see where such a "vector field" might come from let's write the Klein-Gordon equation as $\frac{\partial^2\varphi}{\partial t^2} = (\varDelta - m^2)\varphi$ and employ the usual device of translating this into a first order system of equations. Thus, we let $u = \varphi$ and $v = \partial\varphi/\partial t = \dot{\varphi}$ to obtain

$$(u_t, v_t) = (v, (\varDelta - m^2)u),$$

which we write more suggestively as

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0 & id \\ \varDelta - m^2 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Now take $\varDelta$ to be the distributional Laplacian on $\mathbb{R}^3$ and consider the linear operator

$$\begin{pmatrix} 0 & id \\ \varDelta - m^2 & 0 \end{pmatrix}$$

that carries (u, v) onto $(v, (\varDelta - m^2)u)$. Since there is no time dependence we can regard it as a linear operator on $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$. It is, however, an unbounded operator since it is defined only when $v \in H^1(\mathbb{R}^3;\mathbb{R})$ and $\varDelta u \in L^2(\mathbb{R}^3;\mathbb{R})$. The domain of the operator is therefore

$$H^2(\mathbb{R}^3;\mathbb{R}) \oplus H^1(\mathbb{R}^3;\mathbb{R}),$$

which is a dense linear subspace of $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$. It is this operator that we would like to identify with the "vector field" we are after, but first we should introduce a few general definitions.

Let $\mathcal{E}$ be a real Banach space and $\mathcal{D}$ a dense, linear subspace of $\mathcal{E}$ that is complete in some norm stronger than that of $\mathcal{E}$ (such as $H^2(\mathbb{R}^3;\mathbb{R}) \oplus H^1(\mathbb{R}^3;\mathbb{R})$ in $H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$). In particular, $\mathcal{D}$ is also a Banach space. Let $X : \mathcal{D} \to \mathcal{E}$ be a linear operator on $\mathcal{E}$ with domain $\mathcal{D}$. We canonically identify each tangent space $T_e(\mathcal{E})$ with $\mathcal{E}$ and thereby regard the value of X at $x \in \mathcal{D}$ as a tangent vector $X(x) \in \mathcal{E} = T_x(\mathcal{E})$ to $\mathcal{E}$ at x. Thought of in this way we will refer to X as a *(linear) vector field on* $\mathcal{D}$ *with values in* $\mathcal{E}$. Now suppose $\mathcal{E}$ has defined on it a weak symplectic form ω. Then the vector field X is said to be *Hamiltonian* if there exists a smooth, real-valued function $H : \mathcal{D} \to \mathbb{R}$ on $\mathcal{D}$ such that $dH_x(e) = \omega_x(X(x), e)$ for every $x \in \mathcal{D}$ and every $e \in T_x(\mathcal{D}) = \mathcal{D} \subseteq \mathcal{E} = T_x(\mathcal{E})$. From this it follows that dH_x extends to a bounded linear operator on $\mathcal{E} = T_x(\mathcal{E})$ so we will write this condition simply as

$$dH = \iota_X \omega.$$

When such an H exists and has been fixed we will generally write X as X_H.

Remark 2.4.6. A few remarks are in order here. Since ω is assumed to be only a *weak* symplectic form, it need not be the case that such an X_H exists for any given H and soon this will complicate the problem of defining Poisson brackets. We point out also that most of what we will have to say here extends in a fairly straightforward way to Banach manifolds, but we have restricted our attention to *linear* problems and for these it is sufficient to consider Banach spaces; for a more complete story one can consult [ChM1] and [Mar3].

Example 2.4.5. Let $\mathcal{E} = H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ be the phase space for the Klein-Gordon field with its weak symplectic form ω given, at each point of $\mathcal{E}$, by

$$\omega(\,(\varphi,\dot\varphi),(\psi,\dot\psi)\,) = \langle\varphi,\dot\psi\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} - \langle\dot\varphi,\psi\rangle_{L^2(\mathbb{R}^3;\mathbb{R})}$$

for all $(\varphi,\dot\varphi),(\psi,\dot\psi) \in H^1(\mathbb{R}^3;\mathbb{R})\oplus L^2(\mathbb{R}^3;\mathbb{R})$. Also let $\mathcal{D} = H^2(\mathbb{R}^3;\mathbb{R})\oplus H^1(\mathbb{R}^3;\mathbb{R})$. We consider the vector field X defined on $\mathcal{D}$ with values in $\mathcal{E}$ given by

$$X(\varphi,\dot\varphi) = (\dot\varphi,(\varDelta - m^2)\varphi)\ \ \forall(\varphi,\dot\varphi) \in \mathcal{D}.$$

We claim that $X = X_H$, where

$$H(\varphi,\dot\varphi) = \frac{1}{2}\int_{\mathbb{R}^3}(\,\dot\varphi^2 + \nabla\varphi\cdot\nabla\varphi + m^2\varphi^2\,)\,d^3\mathbf{x} = \frac{1}{2}(\,\langle\dot\varphi,\dot\varphi\rangle + \langle\nabla\varphi,\nabla\varphi\rangle + m^2\langle\varphi,\varphi\rangle\,),$$

and the inner products are all in $L^2(\mathbb{R}^3;\mathbb{R})$. To see this we must prove that $dH = \iota_X\omega$. In more detail, this is

$$dH_{(\varphi,\dot\varphi)}(\psi,\dot\psi) = \omega(\,(\dot\varphi,\ \varDelta\varphi - m^2\varphi),\ (\psi,\dot\psi)\,)$$

for all $(\varphi,\dot\varphi) \in \mathcal{D}$ and all $(\psi,\dot\psi) \in T_{(\varphi,\dot\varphi)}(\mathcal{D}) = \mathcal{D}$. For the right-hand side we get

$$\omega(\,(\dot\varphi,\ \varDelta\varphi - m^2\varphi),\ (\psi,\dot\psi)\,) = \langle\dot\varphi,\dot\psi\rangle - \langle\varDelta\varphi,\psi\rangle + m^2\langle\varphi,\psi\rangle.$$

For the left-hand side,

$$dH_{(\varphi,\dot\varphi)}(\psi,\dot\psi) = \frac{d}{d\varepsilon}H(\,(\varphi,\dot\varphi) + \varepsilon(\psi,\dot\psi)\,)|_{\varepsilon=0}.$$

Exercise 2.4.4. Compute this derivative to show that

$$dH_{(\varphi,\dot\varphi)}(\psi,\dot\psi) = \langle\dot\varphi,\dot\psi\rangle + \langle\nabla\varphi,\nabla\psi\rangle + m^2\langle\varphi,\psi\rangle.$$

Thus, to complete the proof we need only show that $\langle\nabla\varphi,\nabla\psi\rangle = -\langle\varDelta\varphi,\psi\rangle$. In other words, we must show that

$$\int_{\mathbb{R}^3}\Big(\frac{\partial\varphi}{\partial x^1}\frac{\partial\psi}{\partial x^1} + \frac{\partial\varphi}{\partial x^2}\frac{\partial\psi}{\partial x^2} + \frac{\partial\varphi}{\partial x^3}\frac{\partial\psi}{\partial x^3}\Big)d^3\mathbf{x}$$
$$= -\int_{\mathbb{R}^3}\Big(\frac{\partial^2\varphi}{(\partial x^1)^2}\psi + \frac{\partial^2\varphi}{(\partial x^2)^2}\psi + \frac{\partial^2\varphi}{(\partial x^3)^2}\psi\Big)d^3\mathbf{x}.$$

The integrals make sense because φ and ψ are in $H^2(\mathbb{R}^3;\mathbb{R})$. If φ and ψ are Schwartz functions, then each term on the left-hand side is equal to the corresponding term on the right-hand side by the Integration by Parts formula . The equality follows in general from the density of $\mathcal{S}(\mathbb{R}^3;\mathbb{R})$ in $H^2(\mathbb{R}^3;\mathbb{R})$.

Where do we stand at this point? Our goal is to represent the time evolution of the Klein-Gordon field as the flow of a Hamiltonian vector field on a weak symplectic Banach space. Thus far we have isolated a Banach (actually, Hilbert) space with a weak symplectic form and a vector field on it that is the Hamiltonian vector field corresponding to our candidate for the Klein-Gordon Hamiltonian. The essential item, however, is the flow for it is this that describes the time evolution of the system. For a finite-dimensional manifold and a tangent vector field defined everywhere on an open subset of the manifold the existence of a flow (or local flow) is a consequence of basic existence and uniqueness theorems for ordinary differential equations. Our situation is quite different, however. The Banach space is infinite-dimensional, the vector field/operator is defined only on a dense linear subspace and its values are not tangent to this subspace in the usual sense. One needs to define precisely what is meant by a "flow" in this context and have theorems available that establish their existence. All of this is to be found in the theory of semigroups of operators due to Hille and Yosida, which is discussed in some detail in Remark 8.55 of [Nab5] and summarized in Appendix B..

Example 2.4.6. (*Klein-Gordon as a Hamiltonian Flow*) We return now to the Klein-Gordon phase space $\mathcal{E} = H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$, with its weak symplectic form $\omega(\,(\varphi,\dot{\varphi}),(\psi,\dot{\psi})\,) = \langle \varphi,\dot{\psi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} - \langle \dot{\varphi},\psi\rangle_{L^2(\mathbb{R}^3;\mathbb{R})}$, Hamiltonian

$$\begin{aligned} H(\varphi,\dot{\varphi}) &= \frac{1}{2}\int_{\mathbb{R}^3} (\,\dot{\varphi}^2 + \nabla\varphi\cdot\nabla\varphi + m^2\varphi^2\,)\, d^3\mathbf{x} \\ &= \frac{1}{2}(\,\langle\dot{\varphi},\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + \langle\nabla\varphi,\nabla\varphi\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + m^2\langle\varphi,\varphi\rangle_{L^2(\mathbb{R}^3;\mathbb{R})}\,), \end{aligned}$$

and corresponding Hamiltonian vector field X_H, defined on $\mathcal{D}(X_H) = H^2(\mathbb{R}^3;\mathbb{R}) \oplus H^1(\mathbb{R}^3;\mathbb{R})$ by $X_H(\varphi,\dot{\varphi}) = (\dot{\varphi},(\Delta - m^2)\varphi)$. We have two objectives. First we will indicate how the Hille-Yosida Theorem can be applied to show that X_H is the infinitesimal generator for a C^0-semigroup $\{T_t\}_{t\geq 0}$ of operators on $\mathcal{E}$. Given this, one can "flow" any given initial state in $\mathcal{E}$ to its state at time $t > 0$ by T_t and we will indicate why this flow actually tracks the evolution of the Klein-Gordon field, that is, satisfies the Klein-Gordon equation.

Remark 2.4.7. Twice we have used the word "indicate" rather than "prove". Careful proofs would require a number of results on the existence and uniqueness of solutions to certain partial differential equations. We will try to give the flavor of the arguments without a lengthy digression into details that would take us too far afield. For those who need more, Theorem 6 in Section 7.4 of [Evans] contains a very readable account of the application of Hille-Yosida, as does Section 7.4 of [Pazy]. Section XIV.3 of [Yos] is a bit more condensed, but discusses both the application of Hille-Yosida and the fact that the resulting flow gives solutions to the Klein-Gordon equation. Finally, we must, in good conscience, point out that in glossing over the PDEs we have, in fact, glossed over the essential ingredient, that is, the fact that the

Klein-Gordon operator $K = -\Delta + m^2$ is *elliptic*. Elliptic regularity results are crucial to the success of the procedure we will describe. For reference we will record the version of the Hille-Yosida Theorem that we require (see Appendix B).

Theorem 2.4.3. *(Hille-Yosida Theorem) Let $\mathcal{E}$ be a Banach space and $A : \mathcal{D}(A) \to \mathcal{E}$ a densely defined, closed operator on $\mathcal{E}$. Then A is the infinitesimal generator of a contractive semigroup of operators on $\mathcal{E}$ if and only if*

1. $(0, \infty) \subseteq \rho(A)$, *and*
2. $\|R_\lambda(A)\| = \|(\lambda - A)^{-1}\| \leq \frac{1}{\lambda} \;\; \forall \lambda > 0.$

Before getting started it will be convenient to introduce a bit of notation and an equivalent norm on $H^1(\mathbb{R}^3; \mathbb{R}) \oplus L^2(\mathbb{R}^3; \mathbb{R})$. For $u, v \in H^1(\mathbb{R}^3; \mathbb{R})$ define

$$B[u, v] = \langle \nabla u, \nabla v \rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + m^2 \langle u, v \rangle_{L^2(\mathbb{R}^3;\mathbb{R})}$$

so that

$$B[u, u] = \|\nabla u\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + m^2 \|u\|^2_{L^2(\mathbb{R}^3;\mathbb{R})}.$$

Except for the factor of m^2 these are just the usual $H^1(\mathbb{R}^3; \mathbb{R})$-inner product and squared norm.

Exercise 2.4.5. Show that $B[u, u]^{1/2}$ is a norm on $H^1(\mathbb{R}^3; \mathbb{R})$ and is equivalent to $\| \; \|_{H^1(\mathbb{R}^3;\mathbb{R})}$, that is,

$$c\|u\|_{H^1(\mathbb{R}^3;\mathbb{R})} \leq B[u, u]^{1/2} \leq C\|u\|_{H^1(\mathbb{R}^3;\mathbb{R})}$$

for some constants $0 < c \leq C$.

Now, for $(\phi_1, \psi_1), (\phi_2, \psi_2) \in H^1(\mathbb{R}^3; \mathbb{R}) \oplus L^2(\mathbb{R}^3; \mathbb{R})$, define

$$\mathcal{B}(\,(\phi_1, \psi_1), (\phi_2, \psi_2)\,) = B[\phi_1, \phi_2] + \langle \psi_1, \psi_2 \rangle_{L^2(\mathbb{R}^3;\mathbb{R})}$$

and write

$$|||\,(\phi, \psi)\,|||^2 = \mathcal{B}(\,(\phi, \psi), (\phi, \psi)\,).$$

Then $||| \; |||$ is a norm on $H^1(\mathbb{R}^3; \mathbb{R}) \oplus L^2(\mathbb{R}^3; \mathbb{R})$ and is equivalent to $\| \; \|_{H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})}$.

Exercise 2.4.6. Show that $\mathcal{B}$ is symmetric, bilinear and positive semi-definite so that there is a Cauchy-Schwarz Inequality that takes the form

$$\Big| \mathcal{B}\,(\,(\phi_1, \psi_1), (\phi_2, \psi_2)\,) \Big| \leq |||\,(\phi_1, \psi_1)\,|||\;|||\,(\phi_2, \psi_2)\,|||.$$

Now we'll return to the Klein-Gordon Hamiltonian flow. The objective is to verify the hypotheses of the Hille-Yosida Theorem. Since we know that $\mathcal{D}(X_H) =$

$H^2(\mathbb{R}^3;\mathbb{R}) \oplus H^1(\mathbb{R}^3;\mathbb{R})$ is dense in $\mathcal{E} = H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$, the first thing to check is that the operator $X_H(\varphi,\dot\varphi) = (\dot\varphi,(\varDelta - m^2)\varphi)$ is closed. For this we let $\{(\varphi_k,\dot\varphi_k)\}_{k=1}^\infty$ be a sequence in $\mathcal{D}(X_H)$ satisfying $(\varphi_k,\dot\varphi_k) \to (\varphi,\dot\varphi)$ in $\mathcal{E}$ and $X_H(\varphi_k,\dot\varphi_k) \to (f,g)$ in $\mathcal{E}$. We must show that $(\varphi,\dot\varphi)$ is in $\mathcal{D}(X_H)$ and $X_H(\varphi,\dot\varphi) = (f,g)$. Part of this is easy. Since $\dot\varphi_k \to \dot\varphi$ in $L^2(\mathbb{R}^3;\mathbb{R})$ and $X_H(\varphi_k,\dot\varphi_k) = (\dot\varphi_k,(\varDelta - m^2)\varphi_k) \to (f,g)$ in $\mathcal{E}$ we must have $\dot\varphi_k \to f$ in $H^1(\mathbb{R}^3;\mathbb{R})$ and therefore in $L^2(\mathbb{R}^3;\mathbb{R})$ so $f = \dot\varphi$. Moreover, $(-\varDelta + m^2)\varphi_k \to -g$ in $L^2(\mathbb{R}^3;\mathbb{R})$. For the rest we will need an estimate.

Lemma 2.4.4. *Let* $K : H^2(\mathbb{R}^3;\mathbb{R}) \to L^2(\mathbb{R}^3;\mathbb{R})$ *be the operator* $K = -\varDelta + m^2$. *Then* K *is self-adjoint on* $H^2(\mathbb{R}^3;\mathbb{R})$ *and, moreover, there exists a positive constant* C *such that, for every* $\varphi \in H^2(\mathbb{R}^3;\mathbb{R})$*,*

$$\|\varphi\|_{H^2(\mathbb{R}^3;\mathbb{R})} \le C\big(\,\|K\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} + \|\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})}\,\big).$$

Proof. The self-adjointness of K follows from Theorem 8.5 of [Nab5]. Now notice that it will suffice to prove the inequality for the dense subset $\mathcal{S}(\mathbb{R}^3;\mathbb{R})$ of $H^2(\mathbb{R}^3;\mathbb{R})$ so we let $\varphi \in \mathcal{S}(\mathbb{R}^3;\mathbb{R})$. Write

$$\|\varphi\|^2_{H^2(\mathbb{R}^3;\mathbb{R})} = \int_{\mathbb{R}^3} \varphi^2\, d^3\mathbf{x} + \sum_{i=1}^3 \int_{\mathbb{R}^3} \partial_i\varphi\, \partial_i\varphi\, d^3\mathbf{x} + \sum_{i,j=1}^3 \int_{\mathbb{R}^3} \partial_i\partial_j\varphi\, \partial_i\partial_j\varphi\, d^3\mathbf{x}.$$

Exercise 2.4.7. Integrate by parts, reverse the order of the derivatives and integrate by parts again to show that

$$\sum_{i,j=1}^3 \int_{\mathbb{R}^3} \partial_i\partial_j\varphi\, \partial_i\partial_j\varphi\, d^3\mathbf{x} = \|\,\varDelta\varphi\,\|^2_{L^2(\mathbb{R}^3;\mathbb{R})}.$$

Exercise 2.4.8. Integrate by parts and use the inequality $|ab| \le \frac{1}{2}(a^2 + b^2), a, b \in \mathbb{R}$, to show that

$$\sum_{i=1}^3 \int_{\mathbb{R}^3} \partial_i\varphi\, \partial_i\varphi\, d^3\mathbf{x} \le \frac{1}{2}\big(\,\|\varphi\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + \|\varDelta\varphi\|^2_{L^2(\mathbb{R}^3;\mathbb{R})}\,\big).$$

Consequently,

$$\|\varphi\|^2_{H^2(\mathbb{R}^3;\mathbb{R})} \le \frac{3}{2}\Big(\,\|\varphi\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + \|\,\varDelta\varphi\,\|^2_{L^2(\mathbb{R}^3;\mathbb{R})}\,\Big).$$

Since $-\varDelta\varphi = K\varphi - m^2\varphi$ we have $\|\,\varDelta\varphi\,\|_{L^2(\mathbb{R}^3;\mathbb{R})} \le \|K\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} + m^2\|\varphi\|_{L^2(dR^3;\mathbb{R})}$ and so

$$
\begin{aligned}
\|\varphi\|_{H^2(\mathbb{R}^3;\mathbb{R})} &\leq \sqrt{\frac{3}{2}}\Big(\|\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} + \|\Delta\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} \Big) \\
&\leq \sqrt{\frac{3}{2}}\Big(\|\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} + \|K\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} + m^2\|\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} \Big) \\
&\leq \sqrt{\frac{3}{2}}\Big((1+m^2)\|\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} + (1+m^2)\|K\varphi\|_{L^2(\mathbb{R}^3;\mathbb{R})} \Big)
\end{aligned}
$$

so taking $C = \sqrt{\frac{3}{2}}\,(1+m^2)$ completes the proof. □

Now, since $\varphi_k \to \varphi$ and $K\varphi_k \to -g$ in $L^2(\mathbb{R}^3;\mathbb{R})$, these sequences are Cauchy in $L^2(\mathbb{R}^3;\mathbb{R})$. Lemma 2.4.4 implies that

$$\|\varphi_k - \varphi_l\|_{H^2(\mathbb{R}^3;\mathbb{R})} \leq C\big(\|K\varphi_k - K\varphi_l\|_{L^2(\mathbb{R}^3;\mathbb{R})} + \|\varphi_k - \varphi_l\|_{L^2(\mathbb{R}^3;\mathbb{R})} \big)$$

and therefore $\{\varphi_k\}_{k=1}^{\infty}$ is Cauchy in $H^2(\mathbb{R}^3;\mathbb{R})$. Consequently, $\{\varphi_k\}_{k=1}^{\infty}$ converges in $H^2(\mathbb{R}^3;\mathbb{R})$. It must converge to φ since H^2-convergence implies L^2-convergence. Thus, $\varphi \in H^2(\mathbb{R}^3;\mathbb{R}) = \mathcal{D}(K)$. Since K is self-adjoint on $\mathcal{D}(K)$ it is closed and so $\varphi_k \to \varphi$, $K\varphi_k \to -g$ and $\varphi \in \mathcal{D}(K)$ imply $K\varphi = -g$. Thus, $(\varphi,\dot{\varphi}) \in \mathcal{D}(X_H)$ and $X_H(\varphi,\dot{\varphi}) = (\dot{\varphi}, -K\varphi) = (f,g)$ as required. We conclude that $X_H : \mathcal{D}(X_H) \to \mathcal{E}$ is a densely defined, closed operator.

Next we must check the resolvent conditions in the Hille-Yosida Theorem. Let λ be a nonzero real number and consider the operator $\lambda - X_H : \mathcal{D}(X_H) \to \mathcal{E}$. Then, for any $(\varphi,\dot{\varphi}) \in \mathcal{D}(X_H)$,

$$(\lambda - X_H)(\varphi,\dot{\varphi}) = (\lambda\varphi - \dot{\varphi}, \lambda\dot{\varphi} + K\varphi).$$

Thus, for any $(f,g) \in \mathcal{E}$, the equation $(\lambda - X_H)(\varphi,\dot{\varphi}) = (f,g)$ is equivalent to the system

$$\lambda\varphi - \dot{\varphi} = f \tag{2.27}$$

$$\lambda\dot{\varphi} + K\varphi = g \tag{2.28}$$

and these give

$$\lambda^2\varphi + K\varphi = \lambda f + g.$$

With $K = -\Delta + m^2$ we can write this as

$$\varphi - \nu\Delta\varphi = F,$$

where $\nu = (\lambda^2 + m^2)^{-1} > 0$ and $F = \nu(\lambda f + g) \in L^2(\mathbb{R}^3;\mathbb{R})$. Now we need a theorem from partial differential equations. The following is a special case of Lemma 4.2, Section 7.4, of [Pazy].

Lemma 2.4.5. *If* $\nu > 0$ *and* $F \in L^2(\mathbb{R}^3;\mathbb{R})$*, then there is a unique* $\varphi \in H^2(\mathbb{R}^3;\mathbb{R})$ *satisfying*

$$\varphi - \nu\Delta\varphi = F.$$

Thus, we have a unique $H^2(\mathbb{R}^3;\mathbb{R})$-solution φ to $\lambda^2\varphi + K\varphi = \lambda f + g$ and then $\dot{\varphi} = \lambda\varphi - f \in H^1(\mathbb{R}^3;\mathbb{R})$ gives the unique solution $(\varphi,\dot{\varphi}) \in \mathcal{D}(X_H)$ to $(\lambda - X_H)(\varphi,\dot{\varphi}) = (f,g)$ for any $(f,g) \in \mathcal{E}$. We conclude that the operator $\lambda - X_H : \mathcal{D}(X_H) \to \mathcal{E}$ is one-to-one and onto for any nonzero $\lambda \in \mathbb{R}$. Consequently, every nonzero real number (and, in particular, every positive real number) is in the resolvent set $\rho(X_H)$ and the resolvent operator $R_\lambda(X_H) = (\lambda - X_H)^{-1} : \mathcal{E} \to \mathcal{D}(X_H)$ is bounded for any such λ.

Next we must estimate the norm of the resolvent operator. Whenever (2.27) and (2.28) are satisfied and $\lambda > 0$ we write $(\varphi,\dot{\varphi}) = R_\lambda(X_H)(f,g)$ and our objective is to show that $|||\,(\varphi,\dot{\varphi})\,||| \le \frac{1}{\lambda}\,|||\,(f,g)\,|||$. From (2.28) and (2.27)

$$\begin{aligned}
\lambda\dot{\varphi} + K\varphi = g \Rightarrow\ & \lambda\dot{\varphi} - \Delta\varphi + m^2\varphi = g \\
&\Rightarrow \lambda\langle\dot{\varphi},\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + \langle -\Delta\varphi,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + m^2\langle\varphi,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} = \langle g,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} \\
&\Rightarrow \lambda\|\dot{\varphi}\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[\varphi,\dot{\varphi}] = \langle g,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} \\
&\Rightarrow \lambda\|\dot{\varphi}\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[\varphi,\lambda\varphi - f] = \langle g,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} \\
&\Rightarrow \lambda\big[\,\|\dot{\varphi}\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[\varphi,\varphi]\,\big] = \langle g,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + B[f,\varphi]
\end{aligned}$$

The Cauchy-Schwarz Inequality applied to $\langle g,\dot{\varphi}\rangle_{L^2(\mathbb{R}^3;\mathbb{R})} + B[f,\varphi] = \mathcal{B}((f,g),(\varphi,\dot{\varphi}))$ gives

$$\lambda\Big[\,\|\dot{\varphi}\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[\varphi,\varphi]\,\Big] \le \Big[\,\|g\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[f,f]\,\Big]^{1/2}\Big[\,\|\dot{\varphi}\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[\varphi,\varphi]\,\Big]^{1/2}$$

and then

$$\Big[\,\|\dot{\varphi}\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[\varphi,\varphi]\,\Big]^{1/2} \le \frac{1}{\lambda}\Big[\,\|g\|^2_{L^2(\mathbb{R}^3;\mathbb{R})} + B[f,f]\,\Big]^{1/2}$$

which is just $|||\,(\varphi,\dot{\varphi})\,||| \le \frac{1}{\lambda}\,|||\,(f,g)\,|||$.

We have verified all of the hypotheses of the Hille-Yosida Theorem B.0.3 and therefore conclude that $X_H : \mathcal{D}(X_H) \to \mathcal{E}$ is the infinitesimal generator for a contractive C^0-semigroup $\{T_t\}_{t\ge 0}$ of operators on $\mathcal{E}$. Stated otherwise, the Klein-Gordon Hamiltonian vector field $X_H : H^2(\mathbb{R}^3;\mathbb{R})\oplus H^1(\mathbb{R}^3;\mathbb{R}) \to H^1(\mathbb{R}^3;\mathbb{R})\oplus L^2(\mathbb{R}^3;\mathbb{R})$ has a flow on $H^1(\mathbb{R}^3;\mathbb{R})\oplus L^2(\mathbb{R}^3;\mathbb{R})$. The only question remaining is whether or not this flow actually represents the time evolution of the Klein-Gordon field as it is specified by the Klein-Gordon equation. More precisely, let $(f,g) \in H^1(\mathbb{R}^3;\mathbb{R}) \oplus L^2(\mathbb{R}^3;\mathbb{R})$ and flow (f,g) by T_t for any $t \ge 0$, that is, define

$$(\,\varphi(t,\mathbf{x}),\dot{\varphi}(t,\mathbf{x})\,) = T_t(f,g).$$

One would like to know then that $\varphi(t, \mathbf{x})$, defined in this way, is a solution to the initial value problem

$$\frac{\partial^2 \varphi}{\partial t^2} - \Delta\varphi + m^2\varphi = 0$$

$$\varphi(0, \mathbf{x}) = f(x) \qquad \frac{\partial \varphi}{\partial t}(0, \mathbf{x}) = g(x).$$

This is, indeed, the case, but not obviously so. If f and g are Schwartz functions, then elliptic regularity results show that $\varphi(t, \mathbf{x})$ is smooth on all of $\mathbb{R} \times \mathbb{R}^3$ and satisfies the initial value problem in the classical sense. Otherwise, $\varphi(t, \mathbf{x})$ is only a distributional solution and there is more work to do. We will leave the matter at this and simply refer to Chapter XIV, Section 3, of [Yos] for those who would like more information.

We have seen how at least one classical field theory can be reformulated as a Hamiltonian system. Technical difficulties arise due to the infinite-dimensionality of the phase space, the "weakness" of the symplectic form, and the fact that the Hamiltonian vector field is only densely defined. These difficulties persist when the analogy with Hamiltonian mechanics is pushed further. Consider, for example, the problem of defining Poisson brackets. We let $\mathcal{E}$ denote a real Banach space with a weak symplectic form ω. We have already seen that a smooth, real-valued function f on $\mathcal{E}$ need not have a symplectic gradient, that is, there need not exist a vector field X_f on $\mathcal{E}$ with $df = \iota_{X_f}\omega$ and, even if it does exist, it will generally be defined only on a dense linear subspace $\mathcal{D}(X_f)$ of $\mathcal{E}$. Let us suppose, however, that we have two smooth, real-valued functions f and g on $\mathcal{E}$ for which X_f and X_g both exist, on $\mathcal{D}(X_f)$ and $\mathcal{D}(X_g)$, respectively. Then we can define the *Poisson bracket*

$$\{f, g\} : \mathcal{D}(X_f) \cap \mathcal{D}(X_g) \to \mathbb{R}$$

of f and g by

$$\{f, g\}(x) = \omega_x(X_f(x), X_g(x))$$

for every $x \in \mathcal{D}(X_f) \cap \mathcal{D}(X_g)$. Even in this best case scenario, however, one must still contend with the possibility that $\mathcal{D}(X_f) \cap \mathcal{D}(X_g)$ might consist of the zero vector alone so that nothing much has been defined. This, in turn, complicates the issue of determining conserved quantities. Recall that, in Hamiltonian mechanics, a smooth real-valued function on the phase space is conserved (constant on the trajectories of the Hamiltonian vector field) precisely when it Poisson commutes with the Hamiltonian ($\{f, H\} = 0$). The analogous notion in field theory would be a real-valued function on phase space that is constant on the trajectories of the *flow* of the Hamiltonian vector field (which, of course, presupposes the existence of a flow). Since $\{f, H\} = 0$ can now occur simply because $\mathcal{D}(X_f)$ and $\mathcal{D}(X_H)$ intersect trivially one cannot expect the situation to be quite so simple in field theory. An additional complication arises from the fact that physically interesting quantities are themselves often not globally defined on phase space; even the Hamiltonian can be only densely defined (see Example (b), page 316, of [ChM1]). All-in-all the issue of conservation laws in field theory is considerably more delicate. We will conclude with just one simple result that implies, in particular, the conservation of energy for the Klein-Gordon field. The following is Theorem 1, Section 2, of [ChM1] and we will record the proof here as well.

Theorem 2.4.6. *Let $\mathcal{E}$ be a real Banach space and ω a weak symplectic form on $\mathcal{E}$. Let $X : \mathcal{D}(X) \to \mathcal{E}$ be a vector field on $\mathcal{D}(X) \subseteq \mathcal{E}$ with values in $\mathcal{E}$. Assume that X is the infinitesimal generator of a C^0-semigroup $\{T_t\}_{t\geq 0}$ of operators on $\mathcal{E}$. Then the following are equivalent.*

1. *$\iota_X\omega$ is a closed 1-form on $\mathcal{D}(X)$ ($d(\iota_X\omega) = 0$).*
2. *X is skew-symmetric with respect to ω, that is,*

$$\omega(X\varphi, \psi) = -\omega(\varphi, X\psi) \ \ \forall \varphi, \psi \in \mathcal{D}(X).$$

3. *$X = X_H$, where*

$$H(\varphi) = \frac{1}{2}\omega(X\varphi, \varphi) \ \ \forall \varphi \in \mathcal{D}(X).$$

4. *Each T_t preserves ω in the sense that $T_t^*\omega = \omega \ \forall t \geq 0$.*

Moreover, if any one (and therefore every one) of these conditions is satisfied, then energy is conserved in the sense that

$$H \circ T_t = H$$

on $\mathcal{D}(X)$ for every $t \geq 0$.

Proof. We will prove (1) $\Leftrightarrow$ (2), (2) $\Rightarrow$ (3), (3) $\Rightarrow$ (1), (2) $\Leftrightarrow$ (4) and then the conservation of energy. Write α for the 1-form $\iota_X\omega$. Then, at any point $x \in \mathcal{D}(X)$, $\alpha(x)$ is a real-valued linear functional on $T_x(\mathcal{D}(X)) = \mathcal{D}(X)$ and we will write its value at any $\varphi \in \mathcal{D}(X)$ as

$$\alpha(x) \cdot \varphi = \omega(X(x), \varphi).$$

By definition, the exterior derivative of α is given by

$$d\alpha(x) \cdot (\varphi, \psi) = (D\alpha(x) \cdot \varphi) \cdot \psi - (D\alpha(x) \cdot \psi) \cdot \varphi,$$

where $D\alpha(x)$ is the Fréchet derivative of α at x (Section 1.1). Thus,

$$\begin{aligned}
d\alpha(x)\cdot(\varphi,\psi) &= \frac{d}{d\varepsilon}\alpha(x+\varepsilon\varphi)\Big|_{\varepsilon=0}\cdot\psi - \frac{d}{d\varepsilon}\alpha(x+\varepsilon\psi)\Big|_{\varepsilon=0}\cdot\varphi \\
&= \frac{d}{d\varepsilon}\omega(X(x+\varepsilon\varphi),\psi)\Big|_{\varepsilon=0} - \frac{d}{d\varepsilon}\omega(X(x+\varepsilon\psi),\varphi)\Big|_{\varepsilon=0} \\
&= \frac{d}{d\varepsilon}[\omega(Xx,\psi)+\varepsilon\omega(X\varphi,\psi)]\Big|_{\varepsilon=0} - \frac{d}{d\varepsilon}[\omega(Xx,\varphi)+\varepsilon\omega(X\psi,\varphi)]\Big|_{\varepsilon=0} \\
&= \omega(X\varphi,\psi) - \omega(X\psi,\varphi) \\
&= \omega(X\varphi,\psi) + \omega(\varphi,X\psi)
\end{aligned}$$

so $d\alpha = 0$ if and only if $\omega(X\varphi,\psi) = -\omega(\varphi,X\psi)$ and this proves (1) $\Leftrightarrow$ (2). Next we assume X is skew-symmetric with respect to ω and let $H : \mathcal{D}(H) \to \mathbb{R}$ be defined by $H(\varphi) = \frac{1}{2}\omega(X\varphi,\varphi)$.

Exercise 2.4.9. Show that, for any $x,\varphi \in \mathcal{D}(X)$,

$$dH(x)\cdot\varphi = \frac{d}{d\varepsilon}H(x+\varepsilon\varphi)\Big|_{\varphi=0} = (\iota_X\omega)_x\varphi.$$

This proves (2) $\Rightarrow$ (3). (3) $\Rightarrow$ (1) is clear since $\iota_X\omega = dH$ implies $d(\iota_X\omega) = d(dH) = 0$.

Exercise 2.4.10. Let $\varphi,\psi \in \mathcal{D}(X)$ and consider the real-valued function of t given by $t \to \omega(T_t\varphi, T_t\psi)$. Regard this as the composition

$$t \to (T_t\varphi, T_t\psi) \to \omega(T_t\varphi, T_t\psi).$$

Use the Chain Rule, the fact that ω is linear in each argument, and $\frac{d}{dt}T_t(x) = XT_tx$ to show that

$$\frac{d}{dt}\omega(T_t\varphi, T_t\psi) = \omega(XT_t\varphi, T_t\psi) + \omega(T_t\varphi, XT_t\psi).$$

Consequently, if (2) is satisfied, then $\omega(T_t\varphi, T_t\psi)$ is constant and equal to $\omega(\varphi,\psi)$. Since this is true for any $\varphi,\psi \in \mathcal{D}(X)$ and $\mathcal{D}(X)$ is dense in $\mathcal{E}$, it is also true for any $\varphi,\psi \in \mathcal{E}$ and this proves (2) $\Rightarrow$ (4). Conversely, if (4) is satisfied, then, for any $\varphi,\psi \in \mathcal{D}(X)$, we have

$$0 = \frac{d}{dt}\omega(T_t\varphi, T_t\psi)\Big|_{t=0} = \omega(X\varphi,\psi) + \omega(\varphi,X\psi)$$

so (2) is also satisfied. Thus, we have shown (2) $\Leftrightarrow$ (4).

To prove the conservation of energy we will use (4) and the fact that $XT_t = T_tX$ on $\mathcal{D}(X)$. We therefore have, for every $\varphi \in \mathcal{D}(X)$,

$$H(T_t\varphi) = \frac{1}{2}\omega(XT_t\varphi, T_t\varphi) = \frac{1}{2}\omega(T_tX\varphi, T_t\varphi) = \frac{1}{2}\omega(X\varphi,\varphi) = H(\varphi)$$

as required. □

Exercise 2.4.11. Apply Theorem 2.4.6 to show that the energy of the Klein-Gordon field is conserved.

Appendix A
Tempered Distributions, Sobolev Spaces and Fourier Transforms

Our primary references for this material are [RS1], Section V.3, [RS2], Chapter IX, and [Yos], Chapter VI. There is also a more detailed synopsis with additional motivation and many examples in Sections 5.2 and 8.4.1 of [Nab5]. Let $\mathbb{N}$ denote the set of non-negative integers and $\mathbb{N}^N = \mathbb{N} \times \overset{N}{\cdots} \times \mathbb{N}$ the set of N-tuples of non-negative integers. An element $\alpha = (\alpha_1, \ldots, \alpha_N)$ of $\mathbb{N}^N$ will be called a *multi-index*. For each such multi-index α we write $|\alpha|$ for the sum $\alpha_1 + \cdots + \alpha_N$. If $(q^1, \ldots, q^N)$ are the coordinates of $q \in \mathbb{R}^N$ with respect to some orthonormal basis and ϕ is a smooth real- or complex-valued function on $\mathbb{R}^N$, we will denote by $\partial_\alpha \phi$ the partial derivative

$$(\partial_\alpha \phi)(q) = \left(\frac{\partial}{\partial q^1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial q^N}\right)^{\alpha_N} \phi(q^1, \ldots, q^N).$$

If $\alpha = (0, \ldots, 0)$, then $\partial_\alpha \phi = \phi$. If $\alpha = (1, 0, 0, \ldots, 0, 0), (0, 1, 0, \ldots, 0, 0), \ldots,$ $(0, 0, 0, \ldots, 0, 1)$, we will write $\partial_\alpha \phi$ as $\partial_1 \phi, \partial_2 \phi, \ldots, \partial_N \phi$ so that $\partial_k \phi = \partial \phi / \partial q^k$ for $k = 1, 2, \ldots, N$. We will write q^α for the monomial

$$q^\alpha = (q^1)^{\alpha_1} \cdots (q^N)^{\alpha_N}.$$

The *Schwartz space* $\mathcal{S}(\mathbb{R}^N)$ consists of all smooth, complex-valued functions ϕ on $\mathbb{R}^N$ for which

$$\sup_{q \in \mathbb{R}^N} \left| q^\alpha (\partial_\beta \phi)(q) \right| < \infty$$

for all multi-indices α and β. These are the functions which, together with all of their partial derivatives, decay more rapidly than the reciprocal of any polynomial in $q^1, \ldots, q^N$ as $\|q\| \to \infty$; they are called *Schwartz functions* on $\mathbb{R}^N$. Certainly, $\mathcal{S}(\mathbb{R}^N)$ contains the space $C_0^\infty(\mathbb{R}^N)$ of smooth, complex-valued functions on $\mathbb{R}^N$ with compact support. The set of real-valued elements of $\mathcal{S}(\mathbb{R}^N)$ will be denoted $\mathcal{S}(\mathbb{R}^N; \mathbb{R})$. $\mathcal{S}(\mathbb{R}^N)$ is a vector space over $\mathbb{C}$ and $\mathcal{S}(\mathbb{R}^N; \mathbb{R})$ is a vector space over $\mathbb{R}$.

Recall that a *semi-norm* on a vector space $\mathcal{V}$ is a map $\rho : \mathcal{V} \to [0, \infty)$ satisfying

1. $\rho(v_1 + v_2) \le \rho(v_1) + \rho(v_2) \quad \forall v_1, v_2 \in \mathcal{V}$
2. $\rho(\alpha v) = |\alpha| \rho(v) \quad \forall \alpha \in \mathbb{R}$ or $\mathbb{C}$ and $\forall v \in \mathcal{V}$.

On $\mathcal{S}(\mathbb{R}^N)$ we can define a countable family of semi-norms $\| \; \|_{\alpha,\beta}$

$$\|\phi\|_{\alpha,\beta} = \sup_{q \in \mathbb{R}^N} \left| q^\alpha (\partial_\beta \phi)(q) \right|,$$

parametrized by pairs of multi-indices $\alpha, \beta \in \mathbb{N}^N$. Although each $\| \; \|_{\alpha,\beta}$ is only a semi-norm, the family of all such has the property that $\|\phi\|_{\alpha,\beta} = 0 \, \forall \alpha, \beta \in \mathbb{N}^N \Rightarrow \phi = 0$ so these combine to give a metric

$$\rho(\phi_1, \phi_2) = \sum_{\alpha,\beta \in \mathbb{N}^N} \frac{1}{2^{|\alpha|+|\beta|}} \frac{\|\phi_1 - \phi_2\|_{\alpha,\beta}}{1 + \|\phi_1 - \phi_2\|_{\alpha,\beta}}$$

that is, moreover, complete, that is, Cauchy sequences converge (Theorem V.9 of [RS1]). We supply $\mathcal{S}(\mathbb{R}^N)$ with the topology determined by this metric. A sequence $\{\phi_n\}_{n=1}^\infty$ in $\mathcal{S}(\mathbb{R}^N)$ converges to ϕ in $\mathcal{S}(\mathbb{R}^N)$ if and only if $\{ \|\phi - \phi_n\|_{\alpha,\beta} \}_{n=1}^\infty$ converges to zero in $\mathbb{R}$ for all $\alpha, \beta \in \mathbb{N}^N$. This is clearly a very restrictive notion of convergence. Nevertheless, using cutoff functions one can show that $C_0^\infty(\mathbb{R}^N)$ is dense in $\mathcal{S}(\mathbb{R}^N)$. $\mathcal{S}(\mathbb{R}^N; \mathbb{R})$ is a closed subset of $\mathcal{S}(\mathbb{R}^N)$ and is therefore also complete.

Remark A.0.1. The topology we have just introduced provides the vector space $\mathcal{S}(\mathbb{R}^N)$ with the structure of a *Fréchet space*. . We will not need any general results on Fréchet spaces, but for those who would like to know more we recommend the very thorough treatment in [Ham].

The complex-valued, linear functionals on $\mathcal{S}(\mathbb{R}^N)$ that are continuous with respect to this Fréchet topology are called *tempered distributions* on $\mathbb{R}^N$ and the linear space of all such is denoted $\mathcal{S}'(\mathbb{R}^N)$. If $T \in \mathcal{S}'(\mathbb{R}^N)$, then we will write the value of T at $\phi \in \mathcal{S}(\mathbb{R}^N)$ either as $T[\phi]$ or as $\langle T, \phi \rangle$. The elements of $\mathcal{S}(\mathbb{R}^N)$ are called *test functions*. Notice that, because of their rapid decay at infinity, the elements of $\mathcal{S}(\mathbb{R}^N)$ are all square integrable on $\mathbb{R}^N$ so $\mathcal{S}(\mathbb{R}^N)$ is a *linear subspace* (but *not* a topological subspace) of $L^2(\mathbb{R}^N)$.

$$\mathcal{S}(\mathbb{R}^N) \subseteq L^2(\mathbb{R}^N)$$

Indeed, since $C_0^\infty(\mathbb{R}^N)$ is dense in both $\mathcal{S}(\mathbb{R}^N)$ and $L^2(\mathbb{R}^N)$, the Schwartz functions are dense in $L^2(\mathbb{R}^N)$. Next observe that $L^2(\mathbb{R}^N)$, being a Hilbert space, is isometrically isomorphic to its dual via the Riesz Representation Theorem. Thus, every ψ in $L^2(\mathbb{R}^N)$ gives rise to (and can be identified with) a linear functional T_ψ on $L^2(\mathbb{R}^N)$. The restriction of T_ψ to $\mathcal{S}(\mathbb{R}^N)$ is a linear functional on $\mathcal{S}(\mathbb{R}^N)$ and is, in fact, an element of $\mathcal{S}'(\mathbb{R}^N)$. We will often simply identify T_ψ with ψ so we can then identify $L^2(\mathbb{R}^N)$ with a subset of $\mathcal{S}'(\mathbb{R}^N)$. Thus, we have

$$\mathcal{S}(\mathbb{R}^N) \subseteq L^2(\mathbb{R}^N) \subseteq \mathcal{S}'(\mathbb{R}^N)$$

which is an example of what is called a *Gelfand triple* or *rigged Hilbert space*. With this in mind one often allows oneself such abuses of terminology as "T is a distribution in $L^2(\mathbb{R}^N)$". Distributions of the form T_ψ for some $\psi \in L^2(\mathbb{R}^N)$ are called *regular distributions*, while all of the others are called *singular distributions*. An example of a singular distribution is the *Dirac delta* at any $a \in \mathbb{R}^N$, denoted δ_a and defined by

$$\delta_a[\phi] = \phi(a) \quad \forall \phi \in \mathcal{S}(\mathbb{R}^N).$$

We should point out that the dual of a Fréchet space is generally not a Fréchet space so $\mathcal{S}'(\mathbb{R}^N)$ does not come equipped with a ready-made topology. However, one defines sequential convergence in $\mathcal{S}'(\mathbb{R}^N)$ pointwise on $\mathcal{S}(\mathbb{R}^N)$, that is, a sequence $\{T_n\}$ in $\mathcal{S}'(\mathbb{R}^N)$ converges to T in $\mathcal{S}'(\mathbb{R}^N)$ if and only if $\{T_n[\phi]\}$ converges in $\mathbb{C}$ to $T[\phi]$ for every $\phi \in \mathcal{S}(\mathbb{R}^N)$. Then every element of $\mathcal{S}'(\mathbb{R}^N)$ is the limit of a sequence in $L^2(\mathbb{R}^N)$ (see Example 5.2.10 of [Nab5]).

For any multi-index α, the *α^{th}-distributional derivative* $\partial_\alpha T$ of a distribution T is defined by

$$\partial_\alpha T[\phi] = (-1)^{|\alpha|} T[\partial_\alpha \phi]$$

for every $\phi \in \mathcal{S}(\mathbb{R}^N)$. If $T = T_\psi$ for some $\psi \in L^2(\mathbb{R}^N)$, then $\partial_\alpha T_\psi$ may or may not be regular, that is, may or may not be in $L^2(\mathbb{R}^N)$. For $\alpha = (1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)$ the distributional derivatives $\partial_\alpha T$ are written $\partial_1 T, \partial_2 T, \ldots, \partial_N T$, respectively. Similarly we will write $\partial_{k_1}\partial_{k_2} T$ for $\partial_\alpha T$ when α has 1 in the k_1 and k_2 slots and 0 elsewhere. The *distributional gradient of T* is the N-tuple

$$\nabla T = (\partial_1 T, \partial_2 T, \ldots, \partial_N T).$$

The particular use we would like to make of distributional derivatives at the moment is the description of a certain class of Hilbert spaces. We should be clear on the notational conventions we will employ in the following definitions. If $\psi \in L^2(\mathbb{R}^N)$ we can regard ψ as a tempered distribution, that is, we can identify ψ with $T_\psi \in \mathcal{S}'(\mathbb{R}^N)$. We can then write the distributional derivatives of this distribution as $\partial_\alpha \psi$. One says that the generally non-differentiable function $\psi \in L^2(\mathbb{R}^N)$ has *derivatives in the sense of distributions*. As we mentioned, these may or may not be in $L^2(\mathbb{R}^N)$. The Sobolev spaces are defined by selecting those $\psi \in L^2(\mathbb{R}^N)$ for which various distributional derivatives *are* in $L^2(\mathbb{R}^N)$.

Remark A.0.2. The Sobolev spaces are defined in Section IX.6 of [RS2] in terms of Fourier transforms and we will get to this shortly. The equivalence of this definition with ours is Proposition 1 of that section.

We define the *Sobolev space* $H^1(\mathbb{R}^N)$ to be the subset of $L^2(\mathbb{R}^N)$ consisting of those elements for which the first order distributional derivatives are also in $L^2(\mathbb{R}^N)$, that is,

$$H^1(\mathbb{R}^N) = \Big\{\psi \in L^2(\mathbb{R}^N) : \partial_k\psi \in L^2(\mathbb{R}^N),\ k = 1, 2, \ldots, N\Big\}.$$

This is a linear subspace of $L^2(\mathbb{R}^N)$, but on $H^1(\mathbb{R}^N)$ we will define a new inner product by summing the L^2-inner products of the functions and all of their corresponding first order distributional derivatives. More precisely, we define

$$\langle\psi_1, \psi_2\rangle_{H^1} = \langle\psi_1, \psi_2\rangle_{L^2} + \langle\partial_1\psi_1, \partial_1\psi_2\rangle_{L^2} + \cdots + \langle\partial_N\psi_1, \partial_N\psi_2\rangle_{L^2} \tag{A.1}$$

so that the corresponding norm is given by

$$\|\psi\|^2_{H^1} = \|\psi\|^2_{L^2} + \sum_{k=1}^{N} \|\partial_k\psi\|^2_{L^2}. \tag{A.2}$$

With this inner product, $H^1(\mathbb{R}^N)$ is complete and therefore a Hilbert space. Relative to the norm topology determined by (A.2), the smooth functions on $\mathbb{R}^N$ are dense. Indeed, one can show that the set $C_0^\infty(\mathbb{R}^N)$ of smooth functions with compact support is dense in $H^1(\mathbb{R}^N)$ relative to the $H^1(\mathbb{R}^N)$-norm topology.

Next define the *Sobolev space* $H^2(\mathbb{R}^N)$ to be the subset of $L^2(\mathbb{R}^N)$ consisting of those elements for which the first and second order distributional derivatives are in $L^2(\mathbb{R}^N)$, that is,

$$H^2(\mathbb{R}^N) = \Big\{\psi \in L^2(\mathbb{R}^N) : \partial_k\psi, \partial_{k_1}\partial_{k_2}\psi \in L^2(\mathbb{R}^N), k, k_1, k_2 = 1, 2, \ldots, N\Big\}.$$

$H^2(\mathbb{R}^N)$ is also a Hilbert space with inner product

$$\langle\psi_1, \psi_2\rangle_{H^2} = \langle\psi_1, \psi_2\rangle_{L^2} + \sum_{k=1}^{N}\langle\partial_k\psi_1, \partial_k\psi_2\rangle_{L^2} + \sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\langle\partial_{k_1}\partial_{k_2}\psi_1, \partial_{k_1}\partial_{k_2}\psi_2\rangle_{L^2} \tag{A.3}$$

and corresponding norm

$$\|\psi\|^2_{H^2} = \|\psi\|^2_{L^2} + \sum_{k=1}^{N}\|\partial_k\psi\|^2_{L^2} + \sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\|\partial_{k_1}\partial_{k_2}\psi\|^2_{L^2}. \tag{A.4}$$

Remark A.0.3. We will need only $H^1(\mathbb{R}^N)$ and $H^2(\mathbb{R}^N)$, but, for integers $K \geq 3$, the Sobolev spaces $H^K(\mathbb{R}^N)$ are defined in an entirely analogous manner. *As sets*,

$$\cdots \subseteq H^K(\mathbb{R}^N) \subseteq \cdots \subseteq H^2(\mathbb{R}^N) \subseteq H^1(\mathbb{R}^N) \subseteq L^2(\mathbb{R}^N), \tag{A.5}$$

although each of these has a different inner product. Much more refined information about these inclusions and about the degree of regularity one can expect of the elements of a given Sobolev space can be obtained from the so-called *Sobolev Inequalities*. We mention also that, for $\mathbb{C}^k$-valued functions, the Sobolev norms are defined to be the sum of the Sobolev norms of the coordinate functions and one thereby obtains Sobolev spaces of $\mathbb{C}^k$-valued functions. On occasion we will need to restrict our attention to the real-valued elements of $H^K(\mathbb{R}^N)$ and we will denote the set of all such by $H^K(\mathbb{R}^N;\mathbb{R})$. This is a closed subset of $H^K(\mathbb{R}^N)$ so it is complete with respect to the H^K-norm and therefore $H^K(\mathbb{R}^N;\mathbb{R})$ is a real Hilbert space with respect to the inner product on $H^K(\mathbb{R}^N)$. We will have another way of looking at the Sobolev spaces after reviewing the final topic of this section.

The *Fourier transform* of $\phi \in \mathcal{S}(\mathbb{R}^N)$ is defined by

$$(\mathcal{F}\phi)(p) = \hat{\phi}(p) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} e^{-ip\cdot q}\phi(q)\, d^N q, \tag{A.6}$$

where $p \cdot q = \sum_{i=1}^N p^i q^i$ if $q = (q^1, \ldots, q^N) \in \mathbb{R}^N$ and $p = (p^1, \ldots, p^N) \in \mathbb{R}^N$.

Remark A.0.4. One should really think of the $\mathbb{R}^N$ in which q lives and the $\mathbb{R}^N$ in which p lives as distinct. One can do this by identifying the p copy of $\mathbb{R}^N$ with the vector space dual $(\mathbb{R}^N)^*$ of the q copy of $\mathbb{R}^N$. Then $(q^1, \ldots, q^N)$ are components with respect to some orthonormal basis for $\mathbb{R}^N$ and $(p^1, \ldots, p^N)$ are components with respect to the dual basis for $(\mathbb{R}^N)^*$. It is common to refer to $(\mathbb{R}^N)^*$ as *Fourier space* or *momentum space*. The Schwartz space $\mathcal{S}((\mathbb{R}^N)^*)$ is defined with respect to the dual coordinates $(p^1, \ldots, p^N)$. However, since a Schwartz function $\phi(q) = \phi(q^1, \ldots, q^N)$ of the orthonormal coordinates $(q^1, \ldots, q^N)$ uniquely gives rise to the same Schwartz function $\phi(p) = \phi(p^1, \ldots, p^N)$ of the dual coordinates $(p^1, \ldots, p^N)$ and conversely one generally need not bother to distinguish the two. Unless it is likely to lead to some ambiguity we will follow this practice and write $\mathcal{S}(\mathbb{R}^N)$ for both.

Example A.0.1. The Fourier transform of $\phi(q)$ is defined as a function $\hat{\phi}(p)$ of p, but it can equally well be thought of as a function of q as follows.

$$(\mathcal{F}\phi)(q) = \hat{\phi}(q) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} e^{-iq\cdot x}\phi(x) d^N x$$

Notice that one can think of $p\cdot q$ as the natural pairing on $(\mathbb{R}^N)^* \times \mathbb{R}^N$. In fact, there is nothing sacred about the bilinear form $\langle p, q\rangle = p\cdot q$. All of the essential properties of the Fourier transform that we will describe remain true if it is replaced in the definition by any nondegenerate, symmetric bilinear form $\langle\ ,\ \rangle$ and we will on occasion (Section 2.2) make use of this flexibility by taking $\langle\ ,\ \rangle$ to be the Minkowski inner product on $\mathbb{R}^{1,3}$. We will discuss this a bit more at the end of this section.

Example A.0.2. Let A be an $N \times N$, symmetric, positive definite matrix. Then the Fourier transform of the Gaussian function

$$\phi(q) = e^{-\frac{1}{2}q\cdot Aq}$$

is given by

$$\hat{\phi}(p) = \frac{1}{\sqrt{\det A}}\, e^{-\frac{1}{2}p\cdot A^{-1}p}.$$

The $N = 1$ case is Example 5.2.8 of [Nab5] and the general case is Exercise 8.4.3 of [Nab5].

The Fourier transform of a Schwartz function of q is a Schwartz function of p. Indeed, the mapping $\mathcal{F} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ that sends ϕ to $\mathcal{F}\phi = \hat{\phi}$ is a (Fréchet) continuous, linear, bijection with a continuous inverse $\mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ given by

$$(\mathcal{F}^{-1}\psi)(q) = \check{\psi}(q) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} e^{iq\cdot p}\psi(p)\, d^N p. \tag{A.7}$$

This is called the *Fourier Inversion Theorem* and is Theorem IX.1 of [RS2] and Theorem 1, Section 1, Chapter VI, of [Yos]. Notice that

$$(\mathcal{F}^{-1}\psi)(q) = (\mathcal{F}\psi)(-q)$$

for any $\psi \in \mathcal{S}(\mathbb{R}^N)$.

Next we will record some of the most commonly used properties of the Fourier transform and its inverse (see Section IX.1 of [RS1] or Section 4.3.1 of [Evans]). For any $\phi, \phi_1, \phi_2, \psi \in \mathcal{S}(\mathbb{R}^N)$, any multi-index α, any $r \neq 0$ in $\mathbb{R}$, and any $a \in \mathbb{R}^N$,

1. $\mathcal{F}(\partial_\alpha\phi)(p) = (ip)^\alpha(\mathcal{F}\phi)(p)$
2. $\mathcal{F}((-iq)^\alpha\phi)(p) = \partial_\alpha(\mathcal{F}\phi)(p)$
3. $\mathcal{F}^{-1}(\partial_\alpha\psi)(q) = (-iq)^\alpha(\mathcal{F}^{-1}\psi)(q)$
4. $\mathcal{F}^{-1}((ip)^\alpha\psi)(q) = \partial_\alpha(\mathcal{F}^{-1}\psi)(q)$
5. $\mathcal{F}(\phi(q-a)) = e^{-ia\cdot p}\hat{\phi}(p)$
6. $\mathcal{F}(e^{ia\cdot q}\phi(q)) = \hat{\phi}(p-a)$
7. $\mathcal{F}(\phi(rq)) = \frac{1}{|r|}\,\hat{\phi}\left(\frac{1}{r}\,p\right)$
8. $\mathcal{F}(\phi_1 * \phi_2)(p) = (2\pi)^{N/2}\hat{\phi}_1(p)\hat{\phi}_2(p)$, where the *convolution product* $\phi_1 * \phi_2$ is defined by

$$(\phi_1 * \phi_2)(q) = \int_{\mathbb{R}^N} \phi_1(q-y)\phi_2(y)\, d^N y.$$

Furthermore, $\mathcal{F} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ preserves the L^2-norm, that is,

$$\int_{\mathbb{R}^N} |\phi(q)|^2 d^N q = \int_{\mathbb{R}^N} |\hat{\phi}(p)|^2 d^N p$$

for every $\phi \in \mathcal{S}(\mathbb{R}^N)$ (Corollary to Theorem IX.1 of [RS2]). Since $\mathcal{S}(\mathbb{R}^N)$ is dense in $L^2(\mathbb{R}^N)$ and $\mathcal{F}$ carries $\mathcal{S}(\mathbb{R}^N)$ onto $\mathcal{S}(\mathbb{R}^N)$, this implies that $\mathcal{F}$ extends by continuity to a unitary operator of $L^2(\mathbb{R}^N)$ onto itself, which we will continue to denote

$$\mathcal{F} : L^2(\mathbb{R}^N) \to L^2(\mathbb{R}^N).$$

This is called the *Plancherel Theorem*.

We will continue to refer to $\mathcal{F} : L^2(\mathbb{R}^N) \to L^2(\mathbb{R}^N)$ as the Fourier transform, although it is often called the *Fourier-Plancherel transform*. $\mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N)$ extends to the L^2-adjoint $\mathcal{F}^* : L^2(\mathbb{R}^N) \to L^2(\mathbb{R}^N)$ of $\mathcal{F}$, that is, to the inverse of $\mathcal{F} : L^2(\mathbb{R}^N) \to L^2(\mathbb{R}^N)$ since $\mathcal{F}$ is unitary. For ϕ in $L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$, $\mathcal{F}\phi$ is computed from the integral (A.6), but for an element of $L^2(\mathbb{R}^N)$ that is not Lebesgue integrable on $\mathbb{R}^N$ this integral will not converge. One can compute $\mathcal{F}\phi$ either as a limit in $L^2(\mathbb{R}^N)$ of the Fourier transforms of a sequence of functions in $L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$ converging to ϕ or as

$$(\mathcal{F}\phi)(p) = \hat{\phi}(p) = \lim_{M\to\infty} \frac{1}{(2\pi)^{N/2}} \int_{\|q\|\leq M} e^{-ip\cdot q} \phi(q)\, d^N q,$$

where the limit is in $L^2(\mathbb{R}^N)$; see Corollary 1, Section 2, Chapter VI, of [Yos] or page 11 of [RS2]. Similarly,

$$(\mathcal{F}^{-1}\psi)(q) = \check{\psi}(q) = \lim_{M\to\infty} \frac{1}{(2\pi)^{N/2}} \int_{\|p\|\leq M} e^{iq\cdot p} \phi(p)\, d^N p.$$

Remark A.0.5. The Fourier transform actually extends to all $\phi \in L^1(\mathbb{R}^N)$ by (A.6), but $\hat{\phi}$ will, in general, only be in the space $C^0_\infty(\mathbb{R}^N)$ of continuous functions that vanish at infinity ($|\hat{\phi}(p)| \to 0$ as $\|p\| \to \infty$). This is the so-called *Riemann-Lebesgue Lemma* (see Theorem IX.7 of [RS2]). Moreover, $\mathcal{F}$ maps $L^1(\mathbb{R}^N)$ into, but not onto $C^0_\infty(\mathbb{R}^N)$. Identifying the range of the Fourier transform is a delicate issue and is discussed in Sections IX.2 and IX.3 of [RS2].

The Fourier transform and its inverse extend beyond $L^2(\mathbb{R}^N)$ to the tempered distributions. On $\mathcal{S}(\mathbb{R}^N)$ the Fourier transform is a linear bijection

$$\mathcal{F} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{S}(\mathbb{R}^N).$$

We extend the map to

$$\mathcal{F} : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)$$

as follows. Let T be an element of $\mathcal{S}'(\mathbb{R}^N)$. Then $\mathcal{F}T = \hat{T} \in \mathcal{S}'(\mathbb{R}^N)$ is defined for each $\phi \in \mathcal{S}(\mathbb{R}^N)$ by

$$\langle \mathcal{F}T, \phi \rangle = \langle T, \mathcal{F}\phi \rangle \tag{A.8}$$

or

$$\langle \hat{T}, \phi \rangle = \langle T, \hat{\phi} \rangle. \tag{A.9}$$

As motivation for the definition we consider the case in which $T = T_\psi$ is a regular distribution with $\psi \in L^2(\mathbb{R}^N)$ and show that $\langle T_{\hat{\psi}}, \phi \rangle = \langle T_\psi, \hat{\phi} \rangle$ for every $\phi \in \mathcal{S}(\mathbb{R}^N)$.

$$\begin{aligned}
\langle T_{\hat{\psi}}, \phi \rangle &= \int_{\mathbb{R}^N} \hat{\psi}(q)\phi(q)d^N q \\
&= \int_{\mathbb{R}^N} \left(\frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} e^{-iq\cdot x}\psi(x)d^N x \right) \phi(q)d^N q \\
&= \int_{\mathbb{R}^N} \left(\frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} e^{-ix\cdot q}\phi(q)d^N q \right) \psi(x)d^N x \qquad \text{(Fubini's Theorem)} \\
&= \int_{\mathbb{R}^N} \hat{\phi}(x)\psi(x)d^N x \\
&= \langle T_\psi, \hat{\phi} \rangle.
\end{aligned}$$

Thus, for $\psi \in L^2(\mathbb{R}^N)$,

$$\hat{T}_\psi = T_{\hat{\psi}}.$$

It is common to suppress the distinction between $\psi \in L^2(\mathbb{R}^N)$ and $T_\psi \in \mathcal{S}'(\mathbb{R}^N)$ in which case one says that the distributional Fourier transform of ψ is the same as its L^2-Fourier transform.

Exercise A.0.1. Show that, for $\phi \in \mathcal{S}(\mathbb{R}^N)$,

$$\hat{\hat{\phi}}(q) = \phi(-q).$$

Similarly, we define $\mathcal{F}^{-1} : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)$ by

$$\langle \mathcal{F}^{-1}T, \phi \rangle = \langle T, \mathcal{F}^{-1}\phi \rangle \tag{A.10}$$

or

$$\langle \check{T}, \phi \rangle = \langle T, \check{\phi} \rangle. \tag{A.11}$$

Both $\mathcal{F} : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)$ and $\mathcal{F}^{-1} : \mathcal{S}'(\mathbb{R}^N) \to \mathcal{S}'(\mathbb{R}^N)$ are linear bijections.

If T is any tempered distribution and f is an integrable function on $\mathbb{R}^N$ with the property that it and all of its derivatives are bounded by polynomials, then, in particular, $f\phi$ is a Schwartz function whenever ϕ is a Schwartz function and then

$$\langle fT, \phi \rangle = \langle T, f\phi \rangle$$

defines a tempered distribution fT. This is certainly the case if f is a polynomial on $\mathbb{R}^N$. With this definition one can show that properties (1)-(4) above are still valid when ϕ is taken to be a tempered distribution in $L^2(\mathbb{R}^N)$. For example,

$$\mathcal{F}(\partial_\alpha T) = (ip)^\alpha \mathcal{F}T$$

(see Example 3, Section 2, Chapter VI, of [Yos]). Thus, even for distributions, the Fourier transform takes derivatives to products, which is essentially its *raison d'être*. For example, if ψ is in $L^2(\mathbb{R}^N)$ and if $\partial_k\psi$ is also in $L^2(\mathbb{R}^N)$ for each $k = 1, \ldots, N$, then, for each $p = (p^1, \ldots, p^k, \ldots, p^N)$,

$$\mathcal{F}(\partial_k\psi)(p) = ip^k\,\hat{\psi}(p).$$

But $\mathcal{F}$ is an isometry on $L^2(\mathbb{R}^N)$ so

$$\|\,\partial_k\psi\,\|_{L^2}^2 = \int_{\mathbb{R}^N} (p^k)^2\,|\hat{\psi}(p)|^2\,d^N p.$$

Consequently,

$$\|\,\psi\,\|_{L^2}^2 + \sum_{k=1}^{N} \|\,\partial_k\psi\,\|_{L^2}^2 = \int_{\mathbb{R}^N} (1 + \|p\|^2)\,|\hat{\psi}(p)|^2\,d^N p$$

and we conclude that if ψ is in $H^1(\mathbb{R}^N)$, then $(1 + \|p\|^2)^{\frac{1}{2}}\hat{\psi}(p)$ is in $L^2(\mathbb{R}^N)$ and then

$$\|\,\psi\,\|_{H^1} = \left\|\,(1 + \|p\|^2)^{\frac{1}{2}}\,\hat{\psi}(p)\,\right\|_{L^2}.$$

For functions in $L^2(\mathbb{R}^N)$ it is also true that, conversely, if $(1 + \|p\|^2)^{\frac{1}{2}}\hat{\psi}(p)$ is in $L^2(\mathbb{R}^N)$, then ψ is in $H^1(\mathbb{R}^N)$. One often sees $H^1(\mathbb{R}^N)$ defined as

$$H^1(\mathbb{R}^N) = \{\psi \in L^2(\mathbb{R}^N) : (1 + \|p\|^2)^{\frac{1}{2}}\hat{\psi}(p) \in L^2(\mathbb{R}^N)\}.$$

There are analogous Fourier transform characterizations of all of the Sobolev spaces (Proposition 1, Section IX.6, of [RS2]).

The final issue we would like to address in this section is a modest modification of the usual definition of the Fourier transform that on occasion will be more appropriate to our needs. In abstract harmonic analysis (see [Fol2] and [Rud1]) there is a much more general notion of Fourier transform that we will briefly describe in order to see how our generalization might arise. For this we let G denote an arbitrary locally compact, Hausdorff, Abelian group (such as the additive group $\mathbb{R}^n$). The character group of G is denoted $\hat{G}$ and consists of all continuous homomorphisms $\xi : G \to S^1$ from G to the group S^1 of all complex numbers of modulus one (see Remark 1.5.2 of [Nab6]). $\hat{G}$ is also a locally compact, Hausdorff, Abelian group. G admits a nontrivial regular Borel measure μ_G that is translation invariant ($\mu_G(g + B) = \mu_G(B)$ for every $g \in G$ and every Borel set B in G) and is unique up to a positive multiplicative constant; this is called a *Haar measure* on G. For $G = \mathbb{R}^n$ the usual Lebesgue measure (or any positive multiple of it) is a Haar measure and $\widehat{\mathbb{R}^n} \cong \mathbb{R}^n$ (Remark 1.5.2 of [Nab6]). Now for any $\phi \in L^1(G, \mu_G)$ we define the *Fourier transform* $\mathcal{F}_G\phi : \hat{G} \to \mathbb{C}$ to be the complex-valued function on the character group $\hat{G}$ defined by

$$[\mathcal{F}_G\phi](\xi) = \int_G \xi(g)^{-1}\phi(g)\, d\mu_G(g) = \int_G \overline{\xi(g)}\phi(g)\, d\mu_G(g) \tag{A.12}$$

for every $\xi \in \hat{G}$. One can show that there exists a unique Haar measure $\mu_{\hat{G}}$ on $\hat{G}$ such that $\mathcal{F}_G\phi$ is in $L^1(\hat{G}, \mu_{\hat{G}})$ whenever ϕ is in $L^1(G, \mu_G)$ and

$$\phi(g) = \int_{\hat{G}} \xi(g)(\mathcal{F}_G\phi)(\xi)\, d\mu_{\hat{G}}(\xi)$$

for almost every $g \in G$. The *inverse Fourier transform* is given by

$$[\mathcal{F}_G^{-1}\phi](g) = \int_{\hat{G}} \xi(g)\phi(\xi)\, d\mu_{\hat{G}}(\xi)$$

for every $\phi \in L^1(\hat{G}, \mu_{\hat{G}})$ and every $g \in G$.

Now let's specialize to the additive group $G = \mathbb{R}^{1,3}$. We will denote the points in $\mathbb{R}^{1,3}$ by $q = (q^0, q^1, q^2, q^3)$. In Remark 1.5.2 of [Nab6] it was shown that any element of the character group $\hat{\mathbb{R}}^{1,3}$ can be written in the form

$$\xi(q) = \xi(q^0, q^1, q^2, q^3) = e^{i(p_0q^0 + p_1q^1 + p_2q^2 + p_3q^3)} = e^{ip\cdot q}$$

for some unique $p = (p_0, p_1, p_2, p_3)$ in $\mathbb{R}^4$ and that this gives a group isomorphism between $\hat{\mathbb{R}}^{1,3}$ and $\mathbb{R}^4$. With this identification of $\hat{\mathbb{R}}^{1,3}$ and $\mathbb{R}^4$ and with

$$\frac{1}{(2\pi)^2}\, d^4q = \frac{1}{(2\pi)^2}\, dq^0dq^1dq^2dq^3$$

as Haar measure, (A.12) reduces to the usual Fourier transform

$$(\mathcal{F}\phi)(p) = \hat{\phi}(p) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} e^{-ip\cdot q}\phi(q)\, d^4q.$$

However, it was shown in Section 2.6.3 of [Nab6] that $\hat{\mathbb{R}}^{1,3}$ is naturally identified with the vector space dual $\mathbb{P}^{1,3}$ of $\mathbb{R}^{1,3}$ so it would seem desirable to incorporate the inner product structure of $\mathbb{R}^{1,3}$ rather than that of $\mathbb{R}^4$ into the definition of the Fourier transform. This is easily done by noting that any $\xi \in \hat{\mathbb{R}}^{1,3}$ can equally well be written as

$$\xi(q) = \xi(q^0, q^1, q^2, q^3) = e^{i(p_0q^0 - p_1q^1 - p_2q^2 - p_3q^3)}$$

for some unique $p = (p_0, p_1, p_2, p_3)$. With this (A.12) becomes what we will refer to as the *Minkowski-Fourier transform* and which we will write as $\mathcal{F}_{\mathcal{M}}\phi$ or simply

$\tilde{\phi}$ if the context is clear.

$$(\mathcal{F}_{\mathcal{M}}\phi)(p) = \tilde{\phi}(p) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^{1,3}} e^{-i\langle p,q\rangle}\phi(q)\,d^4q \tag{A.13}$$

Here $\langle p, q\rangle = p_0q^0 - p_1q^1 - p_2q^2 - p_3q^3 = p_\alpha q^\alpha$ and we write $\mathbb{R}^{1,3}$ rather than $\mathbb{R}^4$ simply to emphasize the appearance of the Minkowski inner product (the Haar measures are the same for both, of course). The *inverse Minkowski-Fourier transform* is then given by

$$(\mathcal{F}_{\mathcal{M}}^{-1}\phi)(q) = \frac{1}{(2\pi)^2}\int_{\mathbb{P}^{1,3}} e^{i\langle q,p\rangle}\phi(p)\,d^4p.$$

These have all of the desirable properties of the usual Fourier transform and, in particular, extend to $L^2(\mathbb{R}^{1,3})$ and to the tempered distributions on $\mathbb{R}^{1,3}$.

Remark A.0.6. Note that the Minkowski inner product plays no role in the definitions of $L^2(\mathbb{R}^{1,3})$ or the tempered distributions on $\mathbb{R}^{1,3}$ so these are the same as the corresponding objects on $\mathbb{R}^4$.

The advantages of the Minkowski-Fourier transform will emerge in Section 2.2 and, most particularly, in Exercises 2.2.2 and 2.2.3 where you will show that elements of $\mathcal{S}(\mathbb{R}^{1,3})$ and $\mathcal{S}'(\mathbb{R}^{1,3})$ are Lorentz invariant if and only if their Minkowski-Fourier transforms are Lorentz invariant.

Appendix B
Semigroups of Operators

Let $\mathcal{E}$ denote a Banach space and, for each $t \geq 0$, let $T_t : \mathcal{E} \to \mathcal{E}$ be a bounded linear operator on $\mathcal{E}$. If $\{T_t\}_{t\geq 0}$ satisfies

1. $T_0 = id_{\mathcal{E}}$,
2. $T_{t+s} = T_t T_s, \quad \forall t, s \geq 0$, and
3. for each $x \in \mathcal{E}$,

$$t \to T_t x : [0, \infty) \to \mathcal{E}$$

   is continuous,

then $\{T_t\}_{t\geq 0}$ is called a *strongly continuous semigroup of operators*, or a C^0*-semigroup of operators* on $\mathcal{E}$. $\{T_t\}_{t\geq 0}$ is *contractive* if each T_t has operator norm $\|T_t\| \leq 1$.

Example B.0.1. Let $\mathcal{E}$ be a Banach space and $A : \mathcal{E} \to \mathcal{E}$ a *bounded* linear operator on $\mathcal{E}$. For each $t \geq 0$ define $T_t : \mathcal{E} \to \mathcal{E}$ by

$$T_t = e^{tA} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}.$$

Since $\|A\| < \infty$ the series converges, for each fixed t, in the Banach space $\mathcal{B}(\mathcal{E})$ of bounded operators on $\mathcal{E}$ to a bounded operator. Clearly, $T_0 = id_{\mathcal{E}}$ and, since tA and sA commute, $e^{(t+s)A} = e^{tA} e^{sA}$ so $T_{t+s} = T_t T_s$. Because A is bounded we actually have a much stronger continuity condition than the definition requires. Indeed, since

$$\|T_t - id_{\mathcal{E}}\| = \left\| \sum_{n=1}^{\infty} \frac{(tA)^n}{n!} \right\| \leq \sum_{n=1}^{\infty} \frac{t^n \|A\|^n}{n!} = e^{t\|A\|} - 1$$

and $e^{t\|A\|} - 1 \to 0$ as $t \to 0^+$, $t \to e^{tA}$ is actually continuous as a map into $\mathcal{B}(\mathcal{E})$. $\{T_t\}_{t\geq 0}$ is therefore, in particular, a strongly continuous semigroup of operators on $\mathcal{E}$, but it is generally not contractive.

Exercise B.0.1. Show that, for any $x \in \mathcal{E}$,

$$\lim_{h \to 0} \frac{e^{(t+h)A}(x) - e^{tA}(x)}{h} = Ae^{tA}(x).$$

Write this as

$$\frac{d}{dt} T_t(x) = AT_t(x)$$

and recall that

$$T_0(x) = x.$$

Since every tangent space to a vector space can be identified with that same vector space we can think of the operator A as defining a *vector field* on $\mathcal{E}$ whose value at $x \in \mathcal{E}$ is $Ax \in T_x(\mathcal{E})$. This suggests regarding $T_t(x)$ as the *integral curve* of the vector field on $\mathcal{E}$ represented by A that starts at the identity operator.

Example B.0.2. There is an important semigroup of operators associated with the heat flow on $\mathbb{R}$ that was discussed in Example 5.2.13 and Example 8.4.8 of [Nab5]. We summarize the results. Consider the initial value problem

$$\begin{aligned} \frac{\partial \psi(t,x)}{\partial t} - D\,\frac{\partial^2 \psi(t,x)}{\partial x^2} &= 0, \quad (t,x) \in (0,\infty) \times \mathbb{R} \\ \lim_{t \to 0^+} \psi(t,x) &= \psi_0(x), \quad x \in \mathbb{R} \end{aligned} \tag{B.1}$$

for the *1-dimensional heat (or diffusion) equation*, where D is a positive real number (the *diffusion constant*). Now define the *1-dimensional heat kernel* $H_D : (0,\infty) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$H_D(t,x,y) = \frac{1}{\sqrt{4\pi Dt}} e^{-(x-y)^2/4Dt}.$$

Then a solution to the initial value problem (B.1) is given by

$$\psi(t,x) = \int_{\mathbb{R}} H_D(t,x,y)\psi_0(y)\,dy$$

so $H_D(t,x,y)$ propagates the initial state $\psi(0,x) = \psi_0(x)$ to the state at time t. For each $t > 0$ define a map T_t on $L^2(\mathbb{R})$ by

$$(T_t u)(x) = \int_{\mathbb{R}} H_D(t,x,y)\,u(y)\,dy$$

for every $u \in L^2(\mathbb{R})$ and take T_0 to be the identity map on $L^2(\mathbb{R})$. Then $\{T_t\}_{t\geq 0}$ is a strongly continuous, contractive semigroup of operators on $L^2(\mathbb{R})$ called the *heat semigroup*, or *heat flow*.

Let $\{T_t\}_{t\geq 0}$ be a C^0-semigroup of operators on a Banach space $\mathcal{E}$. We introduce an operator A, called the *infinitesimal generator* of $\{T_t\}_{t\geq 0}$, as follows. The domain of A is

$$\mathcal{D}(A) = \left\{ x \in \mathcal{E} : \lim_{t\to 0^+} \frac{T_t x - x}{t} \text{ exists in } \mathcal{E} \right\}.$$

Then $A : \mathcal{D}(A) \to \mathcal{E}$ is given, at each $x \in \mathcal{D}(A)$, by

$$Ax = \lim_{t\to 0^+} \frac{T_t x - x}{t}.$$

The following is Theorem 2, Section 7.4.1, of [Evans], Theorem 4, Section 34.1, of [Lax], and the Proposition in Section X.8 of [RS2].

Theorem B.0.1. *Let $\{T_t\}_{t\geq 0}$ be a C^0-semigroup of operators on a Banach space $\mathcal{E}$ and $A : \mathcal{D}(A) \to \mathcal{E}$ its infinitesimal generator. Then $\mathcal{D}(A)$ is a dense linear subspace of $\mathcal{E}$ and A is a closed linear operator on $\mathcal{D}(A)$.*

Remark B.0.1. Recall that A is *closed* if, whenever $x_n \in \mathcal{D}(A)$ for $n = 1, 2, \ldots$, $x_n \to x$ in $\mathcal{E}$ and $Ax_n \to y$ in $\mathcal{E}$, then $x \in \mathcal{D}(A)$ and $Ax = y$.

The following is a special case of Theorem 8.4.21 of [Nab5]. In the statement of the Theorem the derivative of $t \to T_t x$, for $x \in \mathcal{E}$, is defined to be the following limit in $\mathcal{E}$, provided the limit exists.

$$\frac{d}{dt} T_t x = \lim_{h\to 0} \frac{T_{t+h} x - T_t x}{h}$$

Theorem B.0.2. *Let $\{T_t\}_{t\geq 0}$ be a contractive C^0-semigroup of operators on a Banach space $\mathcal{E}$ and $A : \mathcal{D}(A) \to \mathcal{E}$ its infinitesimal generator. Let x be in $\mathcal{D}(A)$. Then*

1. *$T_t x$ is in $\mathcal{D}(A)$ for all $t \geq 0$,*
2. *$AT_t x = T_t Ax$ for all $t \geq 0$,*
3. *The map $t \to T_t x$ is continuously differentiable on $t > 0$, and*
4. *$\frac{d}{dt} T_t x = AT_t x$ for all $t > 0$.*

Remark B.0.2. One can regard the infinitesimal generator A as a vector field defined on $\mathcal{D}(A)$ and taking values in $\mathcal{E}$. Then, motivated by (4), we call $\{T_t\}_{t\geq 0}$ the *flow* of A. Also motivated by (4) we introduce the traditional notation for the semigroup generated by A, that is,

$$T_t = e^{tA}.$$

This notation is very suggestive and convenient. For example, $T_t T_s = T_{t+s}$ becomes $e^{tA}e^{sA} = e^{(t+s)A}$ and $\frac{d}{dt}T_t x = AT_t x$ becomes $\frac{d}{dt}e^{tA}x = Ae^{tA}x$. However, one should keep in mind that it is only under certain circumstances that e^{tA} is literally the exponential of an operator in the sense of the functional calculus; this is true, for example, if the infinitesimal generator A happens to be a bounded operator (Example B.0.1) and we will mention one other instance of this in a moment.

Typically, one is not given a flow (semigroup of operators) and asked to find the vector field that gives rise to it (its infinitesimal generator). Rather, one is given a vector field and would like to know that a flow exists. The crucial question then is, given an unbounded operator/vector field A how can one know that it is the infinitesimal generator for some C^0-semigroup of operators? This is the question addressed by the Hille-Yosida Theorem, to which we now turn.

We already know that the infinitesimal generator A of any C^0-semigroup $\{T_t\}_{t\geq 0}$ of operators on a Banach space $\mathcal{E}$ is a densely defined, closed operator on $\mathcal{E}$. If $\{T_t\}_{t\geq 0}$ is contractive, then A has two additional properties and, remarkably enough, these two characterize infinitesimal generators of contractive semigroups among the densely defined, closed operators. To describe these two properties we recall that $\lambda \in \mathbb{C}$ is in the *resolvent set* $\rho(A)$ of the closed operator A if and only if $\lambda - A : \mathcal{D}(A) \to \mathcal{E}$ is one-to-one and onto and that it follows from this that the resolvent operator $R_\lambda(A) = (\lambda - A)^{-1} : \mathcal{E} \to \mathcal{D}(A)$ is bounded (Theorem, Section VIII.1, [Yos]). One can then show that, if A is the infinitesimal generator of a contractive semigroup of operators on a Banach space, then

1. $(0, \infty) \subseteq \rho(A)$, and
2. $\|R_\lambda(A)\| = \|(\lambda - A)^{-1}\| \leq \frac{1}{\lambda} \;\; \forall \lambda > 0$.

For the proof of this one can consult Theorem 3(ii), Section 7.4.1, of [Evans], Section X.8 of [RS2], Section 34.1 of [Lax], or Section IX.3 of [Yos]). That these two properties alone characterize the infinitesimal generators of contractive semigroups of operators on a Banach space among the densely defined, closed operators is the content of the Hille-Yosida Theorem.

Theorem B.0.3. *(Hille-Yosida Theorem) Let $\mathcal{E}$ be a Banach space and $A : \mathcal{D}(A) \to \mathcal{E}$ a densely defined, closed operator on $\mathcal{E}$. Then A is the infinitesimal generator of a contractive semigroup of operators on $\mathcal{E}$ if and only if*

1. $(0, \infty) \subseteq \rho(A)$*, and*
2. $\|R_\lambda(A)\| = \|(\lambda - A)^{-1}\| \leq \frac{1}{\lambda} \;\; \forall \lambda > 0$.

Remark B.0.3. This is Theorem X.47a of [RS2] and Theorem 7, Section 34.2, of [Lax]. An extension of the result to arbitrary C^0-semigroups is generally called the Hille-Yosida-Phillips Theorem and is Theorem X.47b of [RS2]. A more general result for locally convex, sequentially complete topological vector spaces appears as the Theorem in Section IX.7 of [Yos].

Here is a consequence of Hille-Yosida that we will need; it is Theorem 8.4.23 of [Nab5].

Theorem B.0.4. *Let $\mathcal{H}$ be a Hilbert space and $T : \mathcal{D}(A) \to \mathcal{H}$ an operator on $\mathcal{H}$ that is self-adjoint and positive ($\langle T\psi, \psi \rangle \geq 0 \;\; \forall \psi \in \mathcal{D}(T)$). Then $-T$ generates a contractive C^0-semigroup of operators on $\mathcal{H}$.*

Remark B.0.4. We mention that in this case the semigroup operator e^{-tT} really is the exponential function of the operator $-tT$ in the sense of the functional calculus. That is, if

$$T = \int_{[0,\infty)} \lambda \, dE_\lambda$$

is the spectral decomposition of T, then

$$e^{-tT} = \int_{[0,\infty)} e^{-t\lambda} \, dE_\lambda.$$

Example B.0.3. Let $V : \mathbb{R} \to \mathbb{R}$ be a non-negative, measurable function. Then the corresponding multiplication operator on $L^2(\mathbb{R})$, which we will also denote V, is self-adjoint and positive on $\mathcal{D}(V) = \{\phi \in L^2(\mathbb{R}) : V\phi \in L^2(\mathbb{R})\}$. Consequently, $-V$ generates a contractive C^0-semigroup e^{-tV} on $L^2(\mathbb{R})$.

Example B.0.4. Let Δ denote the Laplace operator on $L^2(\mathbb{R})$. Its domain is the set of all ψ in $L^2(\mathbb{R})$ for which the distributional second derivative $\Delta\psi$ is also in $L^2(\mathbb{R})$ and this is precisely the Sobolev space $H^2(\mathbb{R})$ (see Appendix A). On this domain Δ is self-adjoint and satisfies $\langle \Delta\psi, \psi \rangle \leq 0 \;\; \forall \psi \in \mathcal{D}(\Delta)$. Consequently, the operator $-\Delta : \mathcal{D}(\Delta) \to L^2(\mathbb{R})$ is self-adjoint and positive so we conclude from Theorem B.0.4 that Δ generates a contractive semigroup

$$e^{t\Delta}$$

of operators on $L^2(\mathbb{R})$. In Example 8.4.8 of [Nab5] it is shown that this semigroup is precisely the heat semigroup on $L^2(\mathbb{R})$ with diffusion constant $D = 1$ (Example B.0.2).

References

AM. Abraham, R. and J.E. Marsden, *Foundations of Mechanics*, Second Edition, Addison-Wesley, Redwood City, CA, 1987 .

AMR. Abraham, R., J.E. Marsden and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, Third Edition, Springer, New York, NY, 2001.

AS. Aharonov, Y. and L. Susskind, Observability of the Sign Change of Spinors Under 2π Rotations, Phys. Rev., 158, 1967, 1237-1238.

AMS. Aitchison, I.J.R., D.A. MacManus, and T.M. Snyder, Understanding Heisenberg's "Magical" Paper of July 1925: A New Look at the Computational Details, Am. J. Phys. 72 (11), November 2004, 1370-1379.

Alb. Albert, D., *Quantum Mechanics and Experience*, Harvard University Press, Cambridge, MA, 1992.

AHM. Albeverio, S.A., R.J. Hoegh, and S. Mazzucchi, *Mathematical Theory of Feynman Path Integrals: An Introduction*, Second Edition, Springer, New York, NY, 2008.

Alex. Alexandrov, A.D., On Lorentz Transformations, Uspehi Mat. Nauk, 5, 3(37), 1950, 187 (in Russian)

AAR. Andrews, G.E., R. Askey, and R. Roy, *Special Functions*, Cambridge University Press, Cambridge, England, 2000.

Apos. Apostol, T.M., *Mathematical Analysis*, Second Edition, Addison-Wesley Publishing Co., Reading, MA, 1974.

Arn1. Arnol'd, V.I., *Ordinary Differential Equations*, MIT Press, Cambridge, MA, 1973.

Arn2. Arnol'd, V.I., *Mathematical Methods of Classical Mechanics*, Second Edition. Springer, New York, NY, 1989.

ADR. Aspect, A., J. Dalibard, and G. Roger, Experimental Test of Bell's Inequalities Using Time-Varying Analyzers, Physical Review Letters, Vol. 49, No. 25, 1982, 1804-1807.

Bal. Ballentine, L.E., The Statistical Interpretation of Quantum Mechanics, Rev. Mod. Phys., Vol 42, No 4, 1970, 358-381.

Barg. Bargmann, V., Note on Wigner's Theorem on Symmetry Operations, J. Math. Phys., Vol 5, No 7, 1964, 862-868.

BW. Bargmann, V. and E.P. Wigner, Group Theoretical Discussion of Relativistic Wave Equations, Proc. Nat. Acad. Sci., Vol 34, 1948, 211-223.

BB-FF. Barone, F.A., H.Boschi-Filho, and C. Farina, Three Methods for Calculating the Feynman Propagator, Am. J. Phys., 71, 5, 2003, 483-491.

Bar. Barut, A.O. and R. Raczka, *Theory of Group Representations and Applications*, Polish Scientific Publishers, Warszawa, Poland, 1977.

Bell. Bell, J.S., On the Einstein, Podolsky, Rosen Paradox, Physics 1, 3, 1964, 195-200.

BS. Berezin, F.A. and M.A. Shubin, *The Schrödinger Equation*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.

BGV. Berline, N., E. Getzler, and M. Vergne, *Heat Kernels and Dirac Operators*, Springer, New York, NY, 2004.

Bern. Berndt, R. *An Introduction to Symplectic Geometry*, American Mathematical Society, Providence, RI, 2001.

BP. Bernstein, H.J. and A.V. Phillips, Fiber Bundles and Quantum Theory, Scientific American, 245, 1981, 94-109.

BG. Bishop, R.L. and S. Goldberg, *Tensor Analysis on Manifolds*, Dover Publications, Inc., Mineola, NY, 1980.

BD1. Bjorken, J.D. and S.D. Drell, *Relativistic Quantum Mechanics*, McGraw-Hill Book Company, New York, 1964.

BD2. Bjorken, J.D. and S.D. Drell, *Relativistic Quantum Fields*, McGraw-Hill Book Company, New York, 1965.

BEH. Blank, J., Exner, P., and M. Havlicek, *Hilbert Space Operators in Quantum Physics*, Second Edition, Springer, New York, NY, 2010.

Blee. Bleecker, D., *Gauge Theory and Variational Principles*, Addison-Wesley Publishing Company, Reading, MA, 1981.

BLT. Bogolubov, N.N., A.A. Logunov, and I.T. Todorov, *Introduction to Axiomatic Quantum Field Theory*, W.A. Benjamin, Inc., Reading, MA, 1975.

Bohm. Bohm, D., *Quantum Theory*, Dover Publications, Inc., Mineola, NY, 1989.

BR. Bohr, N. and L. Rosenfeld, Zur Frage der Meßbarkeit der Electromagnetischen Feldgrößen, Danske Mat.-Fys. Meddr. XII **8**, 1933, 1-65.

Bolk. Bolker, E.D., The Spinor Spanner, Amer. Math. Monthly, Vol. 80, No. 9, 1973, 977-984.

Born1. Born, M., Nobel Lecture: The Statistical Interpretations of Quantum Mechanics, $http$: $www.nobelprize.org/nobel\,prizes/physics/laureates/1954/born - lecture.html$

Born2. Born, M., My Life: Recollections of a Nobel Laureate, Charles Scribner's Sons, New York, NY, 1975.

BJ. Born, M. and P. Jordan, Zur Quantenmechanik, Zeitschrift für Physik, 34, 858-888, 1925. [English translation available in [VDW].]

BHJ. Born, M., W. Heisenberg, and P. Jordan, Zur Quantenmechanik II, Zeitschrift für Physik, 35, 557-615, 1926. [English translation available in [VDW].]

BK. Braginsky, V.B. and F. Ya. Khalili, *Quantum Measurement*, Cambridge University Press, Cambridge, England, 1992.

BC. Brown, J.W. and R.V. Churchill, *Fourier Series and Boundary Value Problems*, Seventh Edition, McGraw-Hill, Boston, MA, 2008.

Br. Brown, L.M. (Editor), *Feynman's Thesis: A New Approach to Quantum Theory*, World Scientific, Singapore, 2005.

Busch. Busch, P., The Time-Energy Uncertainty Relation, http://arxiv.org/pdf/quant-ph/0105049v3.pdf.

Cam. Cameron, R.H., A Family of Integrals Serving to Connect the Wiener and Feynman Integrals, J. Math. and Phys., 39, 126-140, 1960.

Car1. Cartan, H., *Differential Forms*, Houghton-Mifflin, Boston, MA, 1970.

Car2. Cartan, H., *Differential Calculus*, Kershaw Publishing, London, England, 1971.

Cas. Casado, C.M.M., A Brief History of the Mathematical Equivalence of the Two Quantum Mechanics, Lat. Am. J Phys. Educ., Vol. 2, No. 2, 152-155, 2008.

Ch. Chernoff, P.R., Note on Product Formulas for Operator Semigroups, J. Funct. Anal., 2, 238-242, 1968.

ChM1. Chernoff, P.R. and J.E. Marsden, Some Basic Properties of Infinite Dimensional Hamiltonian Systems, Colloq. Intern., 237, 313-330, 1976. (PDF available at http://authors.library.caltech.edu/20406/1/ChMa1976.pdf)

ChM2. Chernoff, P.R. and J.E. Marsden, Some Remarks on Hamiltonian Systems and Quantum Mechanics, in *Problems of Probability Theory, Statistical Inference, and Statistical Theories of Science*, 35-43, edited by Harper and Hooker, D. Reidel Publishing Company, Dordrecht, Holland, 1976.

Chev. Chevalley, C., *Theory of Lie Groups*, Princeton University Press, Princeton, NJ, 1946.

CL. Cini, M. and J.M. Lévy-Leblond, *Quantum Theory without Reduction*, Taylor and Francis, Boca Raton, FL, 1990.

Corn. Cornish, F.H.J., The Hydrogen Atom and the Four-Dimensional Harmonic Oscillator, J. Phys. A: Math. Gen., 17, 1984, 323-327.

CM. Curtis, W.D. and F.R. Miller, *Differential Manifolds and Theoretical Physics*, Academic Press,Inc., Orlando, FL, 1985.

deJag. de Jager, E.M., The Lorentz-Invariant Solutions of the Klein-Gordon Equation, Siam J. Appl. Math., Vol. 15, No. 4, 1967, 944-963. For more details see http://oai.cwi.nl/oai/asset/7768/7768A.pdf.

Del. Deligne, P.,*et al*, Editors, *Quantum Fields and Strings: A Course for Mathematicians*, Volumes 1-2, American Mathematical Society, Providence, RI, 1999.

deAI. de Azcárraga, J. and J. Izquierdo, *Lie Groups, Lie Algebras, Cohomology and Some Applications in Physics*, Cambridge University Press, Cambridge, England, 1995.

deMJS. de Muynck, W.M., P.A.E.M. Janssen and A. Santman, Simultaneous Measurement and Joint Probability Distributions in Quantum Mechanics, Foundations of Physics, Vol. 9, Nos. 1/2, 1979, 71-122.

Dirac1. Dirac, P.A.M., The Fundamental Equations of Quantum Mechanics, Proc. Roy. Soc. London, Series A, Vol. 109, No. 752, 1925, 642-653.

Dirac2. Dirac, P.A.M., The Lagrangian in Quantum Mechanics, Physikalische Zeitschrift der Sowjetunion, Band 3, Heft 1, 1933, 64-72.

Ehren. Ehrenfest, P., Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik, Zeitschrift für Physik, 45, 1927, 455-457.

Ein1. Einstein, A., Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt, Annalen der Physik 17(6), 1905, 132-148. English translation in [terH]

Ein2. Einstein, A., Zur Elektrodynamik bewegter Körper, Annalen der Physik, 17(10), 1905, 891-921. English translation in [Ein3]

Ein3. Einstein, A., et al., *The Principle of Relativity*, Dover Publications, Inc., Mineola, NY, 1952.

Ein4. Einstein, A., *Investigations on the Theory of the Brownian Movement*, Edited with Notes by R. Fürth, Translated by A.D. Cowper, Dover Publications, Inc., Mineola, NY, 1956.

EPR. Einstein, A., B. Podolsky,and N. Rosen, Can the Quantum Mechanical Description of Reality be Considered Complete? Physical Review, 41, 1935, 777-780.

Eis. Eisberg, R.M., *Fundamentals of Modern Physics*, John Wiley and Sons, New York, NY, 1967.

Emch. Emch, G.G., *Mathematical and Conceptual Foundations of 20th-Century Physics*, North Holland, Amsterdam, The Netherlands, 1984.

Evans. Evans, L.C., *Partial Differential Equations*, Second Edition, American Mathematical Society, Providence, RI, 2010.

Fad. Fadell, E., Homotopy Groups of Configuration Spaces and the String Problem of Dirac, Duke Math. J., 29, 1962, 231-242.

FP1. Fedak, W.A. and J.J. Prentis, Quantum Jumps and Classical Harmonics, Am. J. Phys. 70 (3), 2002, 332-344.

FP2. Fedak, W.A. and J.J. Prentis, The 1925 Born and Jordan Paper "On Quantum Mechanics", Am. J. Phys. 77 (2), 2009, 128-139.

Fels. Felsager, B., *Geometry, Particles, and Fields*, Springer, New York, NY, 1998.

Ferm. Fermi, E., *Thermodynamics*, Dover Publications, Mineola, NY,1956.

FLS. Feynman, R.P., R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics, Vol I-III*, Addison-Wesley, Reading, MA, 1964.

Feyn. Feynman, R.P., Space-Time Approach to Non-Relativistic Quantum Mechanics, Rev. of Mod. Phys., 20, 1948, 367-403.

Flamm. Flamm, D., Ludwig Boltzmann-A Pioneer of Modern Physics, arXiv:physics/9710007.

Fol1. Folland, G.B., *Harmonic Analysis in Phase Space*, Princeton University Press, Princeton, NJ, 1989.

Fol2. Folland, G.B., *A Course in Abstract Harmonic Analysis*, CRC Press, Boco Ratan, FL, 1995.

Fol3. Folland, G. B., *Quantum Field Theory: A Tourist Guide for Mathematicians*, American Mathematical Society, Providence, RI, 2008.

FS. Folland, G.B. and A. Sitaram, The Uncertainty Principle: A Mathematical Survey, Journal of Fourier Analysis and Applications, Vol. 3, No. 3, 1997, 207-238.

FU. Freed, D.S. and K.K. Uhlenbeck, Editors, *Geometry and Quantum Field Theory*, American Mathematical Society, Providence, RI, 1991.

Fried. Friedman, A. *Foundations of Modern Analysis*, Dover Publications, Inc., Mineola, NY, 1982.

FrKo. Friesecke, G. and M. Koppen, On the Ehrenfest Theorem of Quantum Mechanics, J. Math. Phys., Vol. 50, Issue 8, 2009. Also see http://arxiv.org/pdf/0907.1877.pdf.

FrSc. Friesecke, G. and B. Schmidt, A Sharp Version of Ehrenfest's Theorem fo General Self-Adjoint Operators, Proc. R. Soc. A, 466, 2010, 2137-2143. Also see http://arxiv.org/pdf/1003.3372.pdf.

Fuj1. Fujiwara, D., A Construction of the Fundamental Solution for the Schrödinger Equations, Proc. Japan Acad., 55, Ser. A, 1979, 10-14.

Fuj2. Fujiwara, D., On the Nature of Convergence of Some Feynman Path Integrals I, Proc. Japan Acad., 55, Ser. A, 1979, 195-200.

Fuj3. Fujiwara, D., On the Nature of Convergence of Some Feynman Path Integrals II, Proc. Japan Acad., 55, Ser. A, 1979, 273-277.

Gaal. Gaal, S.A., *Linear Analysis and Representation Theory*, Springer, New York, NY, 1973.

Går. Gårding, L., Note on Continuous Representations of Lie Groups, Proc. Nat. Acad. Sci., Vol. 33, 1947, 331-332.

Gel. Gel'fand, I.M., *Representations of the Rotation and Lorentz Groups and their Applications*, Martino Publishing, Mansfield Centre, CT, 2012.

GN. Gel'fand, I.M. and M.A. Naimark, On the Embedding of Normed Rings into the Ring of Operators in Hilbert Space, Mat. Sbornik 12, 1943, 197-213.

GS. Gel'fand, I.M. and G.E. Shilov, *Generalized Functions, Volume I*, Academic Press, Inc., New York, NY, 1964.

GY. Gel'fand, I.M. and A.M. Yaglom, Integration in Functional Spaces and its Applications in Quantum Physics, J. Math. Phys., Vol. 1, No. 1, 1960, 48-69.

GJ. Glimm, J. and A. Jaffe, *Quantum Physics: A Functional Integral Point of View*, Second Edition, Springer, New York, NY, 1987.

Gold. Goldstein, H., C. Poole and J. Safko, *Classical Mechanics*, Third Edition. Addison-Wesley, Reading, MA, 2001.

Good. Goodman, R.W., Analytic and Entire Vectors for Representations of Lie Groups, Trans. Amer. Math. Soc., 143, 55-76, 1969.

Got. Gotay, M.J., On the Groenewold-Van Hove Problem for $\mathbb{R}^{2n}$, J. Math. Phys., Vol. 40, No. 4, 2107-2116, 1999.

GGT. Gotay, M.J., H.B. Grundling, and G.M. Tuynman, Obstruction Results in Quantization Theory, J. Nonlinear Sci., Vol. 6, 469-498, 1996.

GR. Gradshteyn, I.S. and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Seventh Edition, Academic Press, Burlington, MA, 2007.

GG. Gravel, P. and C. Gauthier, Classical Applications of the Klein-Gordon Equation, Am. J. Phys. 79(5), 447-453, May, 2011.

Gre. Greenberg, M., *Lectures on Algebraic Topology*, W.A. Benjamin, New York, NY, 1967.

Gri. Greiner, W., *Relativistic Quantum Mechanics: Wave Equations*, 3rd Edition Springer-Verlag, Berlin, 2000.

Groe. Groenewold, H.J., On the Principles of Elementary Quantum Mechanics, Physica (Amsterdam), 12, 405-460, 1946.

GS1. Guillemin, V. and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, England, 1984.

GS2. Guillemin, V. and S. Sternberg, *Supersymmetry and Equivariant de Rham Theory*, Springer, New York, NY, 1999.

Gurtin. Gurtin, M.E,, *An Introduction to Continuum Mechanics*, Academic Press, New York, NY, 1981.

HK. Hafele, J.C. and R.E. Keating, Around-the-World Atomic Clocks: Observed Relativistic Time Gains, Science 177 (4044), 168-170.

Hall. Hall, B.C., *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, Springer, New York, NY, 2003.

Hal1. Halmos, P.R., *Measure Theory*, Springer, New York, NY, 1974.

Hal2. Halmos, P.R., Foundations of Probability Theory, Amer. Math. Monthly, Vol 51, 1954, 493-510.

Hal3. Halmos, P.R., What Does the Spectral Theorem Say?, Amer. Math. Monthly, Vol 70, 1963, 241-247.

Ham. Hamilton, R.S., The Inverse Function Theorem of Nash and Moser, Bull. Amer. Math. Soc., Vol 7, No 1, July, 1982.

Hardy. Hardy, G.H., *Divergent Series*, Clarendon Press, Oxford, England, 1949.

Heis1. Heisenberg, W., Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen, Zeitschrift für Physik, 33, 1925, 879-893. [English translation available in [VDW].]

Heis2. Heisenberg, W., *The Physical Principles of the Quantum Theory*, Dover Publications, Inc., Mineola, NY, 1949.

Helg. Helgason, S., *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, NY, 1978.

HP. Hille, E. and R.S. Phillips, *Functional Analysis and Semi-Groups*, American Mathematical Society, Providence, RI, 1957.

Horv. Horvathy, P.A., The Maslov Correction in the Semiclassical Feynman Integral, http://arxiv.org/abs/quant-ph/0702236

Howe. Howe, R., On the Role of the Heisenberg Group in Harmonic Analysis, Bull. Amer. Math. Soc., 3, No. 2, 1980, 821-843.

Iyen. Iyengar, K.S.K., A New Proof of Mehler's Formula and Other Theorems on Hermitian Polynomials, Proc. Indian Acad. Sci., Section A, 10 (3), 1939, 211-216. (See http://repository.ias.ac.in/29465/

Jaffe. Jaffe, A., High Energy Behavior in Quantum Field Theory, Strictly Localizable Fields, Phys. Rev. 158, 1967, 1454-1461.

Jamm. Jammer, M., *The Conceptual Development of Quantum Mechanics*, Tomash Publishers, Los Angeles, CA, 1989.

JL. Johnson, G.W. and M.L. Lapidus, *The Feynman Integral and Feynman's Operational Calculus*, Oxford Science Publications, Clarendon Press, Oxford, England, 2000.

Kal. Kallianpur, G., *Stochastic Filtering Theory*, Springer, New York, NY, 1980.

Kant. Kantorovitz, S., *Topics in Operator Semigroups* Birkhäuser Boston, Springer, New York, NY, 2010.

Kato. Kato, T., Fundamental Properties of Hamiltonian Operators of Schrödinger Type, Trans. Amer. Math. Soc., Vo. 70, No. 2, 1951, 195-211.

KM. Keller, J.B. and D.W. McLaughlin, The Feynman Integral, Amer. Math. Monthly, 82, 1975, 451-465.

Kenn. Kennard, E.H., Zür Quantenmechanik einfacher Bewegungstypen, Zeitschrift füphysik, Vol 44, 1927, 326-352.

Kiri. Kirillov, A.A., *Lectures on the Orbit Method*, American Mathematical Society, Providence, RI, 2004.

Knapp. Knapp, A.WW., *Lie Groups: Beyond an Introduction*, Second Edition, Birkhauser, Boston, MA, 2002.

KN1. Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry*, Volume 1, Wiley-Interscience, New York, NY, 1963.

KN2. Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry*, Volume 2, Wiley-Interscience, New York, NY, 1969.

Koop. Koopman, B.O., Hamiltonian Systems and Transformations in Hilbert Spaces, Proc. Nat. Acad. Sci., Vol 17, 1931, 315-318.

Lands. Landsman, N.P., Between Classical and Quantum, http://arxiv.org/abs/quant-ph/0506082

LaLi. Landau, L.D. and E.M. Lifshitz, *The Classical Theory of Fields*, Third Revised English Edition, Pergamon Press, Oxford, England, 1971.

Lang1. Lang, S., *Linear Algebra*, Second Edition, Addison-Wesley Publishing Company, Reading, MA, 1971.

Lang2. Lang,S., $SL_2(R)$, Springer, New York, NY, 1985.

Lang3. Lang, S., *Differential and Riemannian Manifolds*, Springer, New York, NY, 1995.

Lang4. Lang, S., *Introduction to Differentiable Manifolds*, Second Edition. Springer, New York, NY, 2002.

Langm. Langmann, E., Quantum Theory of Fermion Systems: Topics Between Physics and Mathematics, in *Proceedings of the Summer School on Geometric Methods for Quantum Field Theory*, edited by H.O.Campo, A.Reyes, and S. Paycha, World Scientific Publishing Co., Singapore, 2001.

Lax. Lax, P.D., *Functional Analysis*, Wiley-Interscience, New York, NY, 2002.

Lee. Lee, J.M., *Introduction to Smooth Manifolds*, Springer, New York, NY, 2003.

LGM. Liang, J-Q., B-H. Guo, and G. Morandi, Extended Feynman Formula for Harmonic Oscillator and its Applications, Science in China (Series A), Vol. 34, No. 11, 1346-1353, 1991.

LL. Lieb, E.H. and M. Loss, *Analysis*, American Mathematical Society, Providence, RI, 1997.

Lop. Lopuszánski, J., *An Introduction to Symmetry and Supersymmetry in Quantum Field Theory*, World Scientific, Singapore, 1991.

Lucr. Lucritius, *The Nature of Things*, Translated with Notes by A.E. Stallings, Penguin Classics, New York, NY, 2007.

Mack1. Mackey, G.W., Quantum Mechanics and Hilbert Space, Amer. Math. Monthly, Vol 64, No 8, 1957, 45-57,

Mack2. Mackey, G.W., *Mathematical Foundations of Quantum Mechanics*, Dover Publications, Mineola, NY, 2004.

MacL. Mac Lane, S., Hamiltonian Mechanics and Geometry, Amer. Math. Monthly, Vol. 77, No. 6, 1970, 570-586.

Mar1. Marsden, J., Darboux's Theorem Fails for Weak Symplectic Forms, Proc. Amer. Math. Soc., Vol. 32, No. 2, 1972, 590-592.

Mar2. Marsden, J., *Applications of Global Analysis in Mathematical Physics*, Publish or Perish, Inc., Houston, TX, 1974.

Mar3. Marsden, J., *Lectures on Geometrical Methods in Mathematical Physics*, SIAM, Philadelphia, PA, 1981.

Mazz1. Mazzucchi, S., Feynman Path Integrals, in *Encyclopedia of Mathematical Physics*, Vol. 2, 307-313, J-P Françoise, G.L. Naber, and ST Tsou (Editors), Academic Press (Elsevier), Amsterdam, 2006.

Mazz2. Mazzucchi, S., *Mathematical Feynman Path Integrals and Their Applications* World Scientific, Singapore, 2009.

MS. McDuff, D. and D. Salamon, *Introduction to Symplectic Topology*, Oxford University Press, Oxford, England, 1998.

MR. Mehra, J. and H. Rechenberg, *The Historical Development of Quantum Theory, Volume 2*, Springer, New York, NY, 1982.

Mess1. Messiah, A., *Quantum Mechanics, Volume I*, North Holland, Amsterdam, 1961.

Mess2. Messiah, A., *Quantum Mechanics, Volume II*, North Holland, Amsterdam, 1962.

MTW. Misner, C.W., K.S. Thorne and J.A. Wheeler, *Gravitation*, W.H. Freeman and Company, San Francisco, CA, 1973.

MP. Mörters, P. and Y. Peres, *Brownian Motion*, Cambridge University Press, Cambridge, England, 2010.

Mos. Moser, J., On the Volume Elements on a Manifold, Trans. Amer. Math. Soc., 120, 1965, 286-294.

Nab1. Naber, G., *Spacetime and Singularities: An Introduction*, Cambridge University Press, Cambridge, England, 1988.

Nab2. Naber, G., *Topology, Geometry and Gauge Fields: Foundations*, Second Edition, Springer, New York, NY, 2011.

Nab3. Naber, G., *Topology, Geometry and Gauge Fields: Interactions*, Second Edition, Springer, New York, NY, 2011.

Nab4. Naber, G., *The Geometry of Minkowski Spacetime: An Introduction to the Mathematics of the Special Theory of Relativity*, Second Edition, Springer, New York, NY, 2012.

Nab5. Naber, G., *Foundations of Quantum Mechanics: An Introduction to the Physical Background and Mathematical Structure*, Available at http://gregnaber.com.

Nab6. Naber, G., *Positive Energy Representations of the Poincaré Group: A Sketch of the Positive Mass Case and its Physical Background*, Available at http://gregnaber.com.

Nash. Nash, C., *Differential Topology and Quantum Field Theory*, Academic Press (Elsevier), San Diego, CA, 2003.

Nel1. Nelson, E., Analytic Vectors, Ann. Math., Vol. 70, No. 3, 1959, 572-615.

Nel2. Nelson, E., Feynman Integrals and the Schroedinger Equation, J. Math. Phys., 5, 1964, 332-343.

Nel3. Nelson, E., *Dynamical Theories of Brownian Motion*, Second Edition, Princeton University Press, Princeton, NJ, 2001. Available online at https://web.math.princeton.edu/ nelson/books/bmotion.pdf

O'DV. O'Donnell, K. and M. Visser, Elementary Analysis of the Special Relativistic Combination of Velocities, Wigner Rotation, and Thomas Precession, arXiv:1102.2001v2 [gr-qc] 11 June 2011.

Olv. Olver, P.J., *Applications of Lie Groups to Differential Equations*, Second Edition, Springer, New York, NY, 2000.

O'N. O'Neill, B., *Semi-Riemannian Geometry with Applications to Relativity*, Academic Press, San Diego, CA, 1983.

Oz. Ozawa, M., Universally Valid Reformulation of the Heisenberg Uncertainty Principle on Noise and Disturbance in Measurement, Phys. Rev. A., Vol 67, Issue 4, 042105(2003), 1-6, http://arxiv.org/pdf/quant-ph/0207121v1.pdf.

Pais. Pais, A., *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*, Oxford University Press, Oxford, England, 1982.

PM. Park, J.L. and H. Margenau, Simultaneous Measurability in Quantum Theory, International Journal of Theoretical Physics, Vol. 1, No. 3, 1968, 211-283.

Pater. Paternain, G.P., *Geodesic Flows*, Birkhäuser, Boston, MA, 1999.

Pazy. Pazy, A., *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, NY, New York, 1983.

Per. Perrin, J., *Les Atomes* Flammarion, Paris, New Edition, 1991.

Planck. Planck, M., Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum, Verhandlungen der Deutschen Physikalischen Gesellschaft, 2, 1900, 237-245. English translation available in [terH].

Prug. Prugovečki, E., *Quantum Mechanics in Hilbert Space*, Academic Press, Inc., Orlando, FL, 1971.

Ratc. Ratcliffe, J.G., *Foundations of Hyperbolic Manifolds*, Second Edition, Springer, New York, NY, 2006.

RS1. Reed, M. and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, Inc., Orlando, FL, 1980.

RS2. Reed, M. and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness*, Academic Press, Inc., Orlando, FL, 1975.

RiSz.N. Riesz, F. and B. Sz.-Nagy, *Functional Analysis* Dover Publications, Inc., Mineola, NY, 1990.

Rob. Robertson, H.P., A General Formulation of the Uncertainty Principle and its Classical Interpretation, Phys. Rev. 34(1929), 163-164.

Ros. Rosay, J-P., A Very Elementary Proof of the Malgrange-Ehrenpreis Theorem, Amer. Math. Monthly, Vol. 98, No. 6, Jun.-Jul., 1991, 518-523.

Rosen,S. Rosenberg, S., *The Laplacian on a Riemannian Manifold*, Cambridge University Press, Cambridge, England, 1997.

Rosen,J. Rosenberg, J., A Selective Hisory of the Stone-von Neumann Theorem, Contemporary Mathematics, 365, Amer. Math. Soc., Providence, RI, 2004, 331-353.

Roy. Royden, H.L., *Real Analysis*, Second Edition, Macmillan Co., New York, NY, 1968.

Roz. Rozema, L.A., A. Darabi, D.H. Mahler, A. Hayat, Y. Soudagar, and A.M. Steinberg, Violation of Heisenbergs Measurement-Disturbance Relationship by Weak Measurements, Physical Review Letters, 109, 100404(2012), 1-5.

Rud1. Rudin, W., *Fourier Analysis on Groups*, Interscience Publishers, New York, NY, 1962.

Rud2. Rudin, W., *Functional Analysis*, McGraw-Hill, New York, NY, 1973.

Rud3. Rudin, W., *Real and Complex Analysis*, Third Edition, McGraw-Hill Book Company, New York, 1987.

Ryd. Ryder, L.H., *Quantum Field Theory*, Second Edition, Cambridge University Press, Cambridge, England, 2005.

Sak. Sakuri, J.J. and J. Napolitano, *Modern Quantum Mechanics*, Second Edition, Addison-Wesley, Boston, MA, 2011.

Schl. Schlosshauer, M., Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics, Rev. Mod. Phys., 76(4), 2005, 1267-1305.

Schm1. Schmüdgen, K., On the Heisenberg Commutation Relation I, J. Funct. Analysis, 50, 1983, 8-49.

Schm2. Schmüdgen, K., On the Heisenberg Commutation Relation II, Publ. RIMS, Kyoto Univ, 19, 1983, 601-671.

Schm3. Schmüdgen, K., *Unbounded Self-Adjoint Operators on Hilbert Space*, Springer, New York, NY, 2012.

Schrö1. Schrödinger, E., Quantisierung als Eigenwertproblem, Annalen der Physik, 4, Volume 79, 1926, 273-376. English translation available in [Schrö2].

Schrö2. Schrödinger, E., *Collected Papers on Wave Mechanics*, Third Edition, AMS Chelsea Publishing, American Mathematical Society, Providence, RI, 1982.

Schul. Schulman, L.S., *Techniques and Applications of Path Integration*, Dover Publications, Inc., Mineola, NY, 2005.

Segal1. Segal, I.E., Postulates for General Quantum Mechanics, Annals of Mathematics, Second Series, Vol. 48, No. 4, 1947, 930-948.

Segal2. Segal, I.E., A Class of Operator Algebras Which Are Determined By Groups, Duke Math. J., 18, 1951, 221-265.

SDBS. Sen, D., S.K. Das, A.N. Basu, and S. Sengupta, Significance of Ehrenfest's Theorem in Quantum-Classical Relationship, Current Science, Vol. 80, No. 4, 2001, 536-541.

Simm1. Simmons, G.F., *Topology and Modern Analysis*, McGraw-Hill, New York, NY, 1963.

Simm2. Simmons, G.F., *Differential Equations with Applications and Historical Notes*, McGraw-Hill, New York, NY, 1972.

Simms. Simms, D.J., *Lie Groups and Quantum Mechanics*, Lectures Notes in Mathematics 52, Springer, New York, NY, 1968.

Simon1. Simon, B., The Theory of Semi-Analytic Vectors: A New Proof of a Theorem of Masson and McClary, Indiana Univ. Math. J., 20, 1971, 1145-1151.

Simon2. Simon, B., *Functional Integration and Quantum Physics*, Academic Press, New York, NY, 1979.

Smir. Smirnov, V.A., *Feynman Integral Calculus*, Springer, New York, NY, 2006.

Smith. Smith, M.F., The Pontrjagin Duality Theorem in Linear Spaces, Ann. of Math. 2, 56, 1952, 248-253.

Sp1. Spivak, M., *Calculus on Manifolds*, W.A. Benjamin, New York, NY, 1965.

Sp2. Spivak, M., *A Comprehensive Introduction to Differential Geometry, Vol. I-V*, Third Edition, Publish or Perish, Houston, TX, 1999.

Sp3. Spivak, M., *Physics for Mathematicians, Mechanics I*, Publish or Perish, Houston, TX, 2010.

Srin. Srinivas, M.D., Collapse Postulate for Observables with Continuous Spectra, Commun. Math. Phys. 71 (2), 131-158.

Str. Strange, P., *Relativistic Quantum Mechanics with Applications in Condensed Matter Physics and Atomic Physics*, Cambridge University Press, Cambridge, England, 1998.

SW. Streater, R.F. and A.S. Wightman, *PCT, Spin and Statistics, and All That*, W.A. Benjamin, New York, NY, 1964.

Synge. Synge, J.L., *Relativity: The Special Theory*, North-Holland Publishing Company, Amsterdam, The Netherlands, 1956.

Szab. Szabo, R.J., *Equivariant Cohomology and Localization of Path Integrals*, Springer, New York, NY, 2000. Also see http://arxiv.org/abs/hep-th/9608068.

Sz. Szegö, G., *Orthogonal Polynomials*, American Mathematical Society, Providence, RI, 1939.

Takh. Takhtajan, L.A., *Quantum Mechanics for Mathematicians*, American Mathematical Society, Providence, RI, 2008.

TaylA. Taylor, A.E., *Introduction to Functional Analysis*, John Wiley and Sons, Inc., London, England, 1967.

TaylM. Taylor, M.E., *Partial Differential Equations I*, Second Edition, Springer, New York, NY, 2011.

TW. Taylor, E.F. and J.A. Wheeler, *Spacetime Physics*, W.H. Freeman, San Francisco, CA, 1963.

terH. ter Haar, D., *The Old Quantum Theory*, Pergamon Press, Oxford, England, 1967.

't Ho1. 't Hooft, G., Gauge Theories of the Forces between Elementary Particles, Scientific American, 242, 1980, 90-116.

't Ho2. 't Hooft, G., How a Wave Function can Collapse Without Violating Schrödinger's Equation, and How to Understand Born's Rule, arXiv:1205.4107v2 [quant-ph] 21 May 2012.

Trau. Trautman, A., Noether Equations and Conservation Laws, Commun. Math. Phys. 6, 248-261, 1967.

TAE. Twareque Ali, S. and M. Engliš, Quantization Methods: A Guide for Physicists and Analysts, http://arxiv.org/abs/math-ph/0405065

Tych. Tychonoff, A., Théorèmes d'Unicité pour l'Equation de la Chaleur, Mat. Sb., Vol 42, No 2, 1935, 199-216.

VDB. van den Ban, E., Representation Theory and Applications in Classical Quantum Mechanics, Lectures for the MRI Spring School, Utrecht, June, 2004, v4: 20/6. Available at http://www.staff.science.uu.nl/ ban00101/lecnotes/repq.pdf

VDW. van der Waerden, B.L., Ed., *Sources of Quantum Mechanics*, Dover Publications, Mineola, NY, 2007.

VH. Van Hove, L., Sur Certaines Représentations Unitaires d'un Groupe Infini de Transformations, Proc. R. Acad. Sci. Belgium, 26, 1-102, 1951.

Vara. Varadarajan, V.S., *Geometry of Quantum Theory*, *Second Edition*, Springer, New York, NY, 2007.

Vog. Vogan, D.A., Review of "Lectures on the Orbit Method," by A.A. Kirillov, Bull. Amer. Math. Soc., Vol.42, No. 4, 1997, 535-544.

v.Neu. von Neumann, J., *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, NJ, 1983.

Wald. Wald, R.M., *General Relativity*, University of Chicago Press, Chicago, IL, 1984.

Warn. Warner, F., *Foundations of Differentiable Manifolds and Lie Groups*, Springer, New York, NY, 1983.

Water. Waterhouse, W.C., Dual Groups of Vector Spaces, Pacific Journal of Math., Volume 26, No. 1, 1968, 193-196.

Wat. Watson, G., Notes on Generating Functions of Polynomials, Hermite Polynomials, J. Lond. Math. Soc., 8, 1933, 194-199.

Weinb. Weinberg, S., *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*, Vintage Books, New York, NY, 1993.

Weins. Weinstein, A., Symplectic Structures on Banach Manifolds, Bull. Amer. Math. Soc., Vol. 75, No. 5, 1969, 1040-1041.

Wien. Wiener, N., Differential Space, J. Math. Phys., 2, 1923, 131-174.

Wight. Wightman, A.S, How it was Learned that Quantized Fields are Operator-Valued Distributions, Fortchr. Phys. **44** (2), 1996, 143-178.

Wig1. Wigner, E.P., On Unitary Representations of the Inhomogeneous Lorentz Group, Annals of Mathematics **40** (1), 1939, 149-204.
Wig2. Wigner, E.P., *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*, Academic Press, New York, NY, 1959.
Wig3. Wigner, E.P., Phenomenological Distinction Between Unitary and Antiunitary Symmetry Operators, J. Math. Phys., Volume 1, Issue 5 (1960), 414-416.
Wilcz. Wilczek, F., Origins of Mass, arXiv:1206.7114v2 [hep-ph] 22 Aug 2012, 1-35.
Witt. Witten, E., Supersymmetry and Morse Theory, J. Diff. Geo., 17 (1982), 661-692.
Yos. Yosida, K., *Functional Analysis*, Springer, New York, NY, 1980.
Z. Zeeman, E.C., Causality Implies the Lorentz Group, J. Math. Phys. 5, 1964, 490-493.

Index