SYMPLECTIC MANIFOLDS AND PARTIAL DIFFERENTIAL EQUATIONS

STUDFORSK PROJECT

JONAS P. VEAN - NTNU

1. Introduction

This article explores the intersection between manifold theory and partial differential equations (PDEs). In particular, we generalize the notion of symplectic vector spaces to symplectic manifolds in a self-contained exposition. The goal is to get the reader familiarized with a symplectic formalism on manifolds and function spaces using Hamiltonian functions and the theory surrounding such structures.

The recommended prerequisites for this article is multivariable calculus, linear algebra and basic group theory (permutation groups in particular). Prior exposure to the theory of manifolds and PDEs is also recommended, although most of the theory needed for understanding the article is reviewed in an effort to keep the material self-contained. That being said - this article will not be suitable or entirely sufficient as a first exposure to either manifold theory or the theory of PDEs. The author recommends the books by Loring W. Tu [10] and David Borthwick [2] as undergraduate level supplements for manifold theory and PDEs respectively.

Finally, an effort is made to keep notational conventions consistent throughout the text, although the fields of manifold theory and PDEs are somewhat disjoint in their conventions. Most of the notation is therefore lenient towards the notation used by topologists and geometers where the two fields intersect. In particular, our notation is similar to that of Tu [10].

2. Topological Preliminaries and Essential Manifold Theory

Generalizing concepts from a space that is locally flat and well-behaved like $\mathbb{R}^n$ or $\mathbb{C}^n$ to those that exhibit non-trivial local properties are thoroughly studied in differential geometry. The concept of manifolds as higher-dimensional analogues of surfaces in $\mathbb{R}^n$ is important not only in the study of differential geometry, but also in applications of PDE theory. Notable examples are that of the Einstein field equations from general relativity, any PDE modelled on or as a surface such as a minimal surface problem, etc. In any case, the reader is encouraged to sift through this section in order to refresh one's memory on topological aspects that will be needed in later sections.

Def. 2.1. A map $d\colon X \times X \to \mathbb{R}$ on a set X is called a *metric* if it satisfies

(i) $d(x, y) \geq 0 \quad \forall x, y \in X, \quad d(x, y) = 0$ iff. $x = y$

Date: January 2018 - July 2019.

(ii) $d(x,y) = d(y,x) \quad \forall x, y \in X$
(iii) $d(x,y) \leq d(x,z) + d(z,y) \quad \forall x, y, z \in X$

The pair (X, d) is called a *metric space.*

Metrics do not require anything of the set upon which it is defined. Consider the following example:

Example 2.1. Let $X = \{1, 2, 3\} \subset \mathbb{N}$. Define the function $d\colon X \times X \to \mathbb{R}$ by

$$d(x,y) = \left|\sin(x^2 - y^2)\right|.$$

This is a metric since it satisfies all of the conditions we placed on $d(x,y)$ being a metric on X. Note, however, that X need not be a vector space, $d(x,y)$ does not have to be continuous, surjective or injective as a map. We shall see that this weak requirement is integral in our definition of a *metric topology.* △

Def. 2.2. Consider $\mathbb{R}^n$ equipped with a metric $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. An *open ball* of $\mathbb{R}^n$ wrt. the metric d centered at $x_0 \in \mathbb{R}^n$ with radius $r > 0$ is a subset of $\mathbb{R}^n$ given by

$$B_r(x_0) = \{y \in \mathbb{R}^n \mid d(x_0, y) < r\}. \tag{1}$$

In our study of continuous multivariable functions, such as functions of the form $f\colon U \to \mathbb{R}^m$, where $U \subseteq \mathbb{R}^n$, we usually require some well-behaved properties on the set U. In particular, we may require that U be open in $\mathbb{R}^n$. The collection of open sets of $\mathbb{R}^n$ wrt. a metric d is usually denoted τ_d and is called the *metric topology* on $\mathbb{R}^n$.

Def. 2.3. (Metric topology of $\mathbb{R}^n$)
Let $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a metric on $\mathbb{R}^n$. The *metric topology* τ_d on $\mathbb{R}^n$ is the collection of the open sets described by the metric d.
A set $A \subseteq \mathbb{R}^n$ is called open in $\mathbb{R}^n$ if for every $x \in A$ there exists $r > 0$ such that $B_r(x) \subseteq A$, i.e. any open subset A contains an open ball centered at any point in A.

Remark. It should be noted that there are several possible topologies on any set X. In general, a topology $\mathcal{T}$ on X will have three properties:

(i) $\varnothing, X \in \mathcal{T}$, $\varnothing$ is the empty set
(ii) if $A_\alpha \in \mathcal{T}$ for $\alpha \in \mathscr{A}$, then $\bigcup_{\alpha \in \mathscr{A}} A_\alpha \in \mathcal{T}$
(iii) if $A_\alpha \in \mathcal{T}$ for $\alpha \in \mathscr{A}$, $\mathscr{A}$ finite, then $\bigcap_{\alpha \in \mathscr{A}} A_\alpha \in \mathcal{T}$.

The pair $(X, \mathcal{T})$ is called a *topological space.*

Of course, the metric topology $\mathcal{T}_d$ is in itself a topology, so open subsets of $\mathbb{R}^n$ exhibit the same three properties as above. Throughout the rest of the article, we shall consider open sets given by the standard (metric) topology, where the open balls are given as

$$B_r(x) = \left\{ y \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^n (x_i - y_i)^2 < r^2 \right\}.$$

Def. 2.4. A map $f\colon X \to Y$ between topological spaces X, Y is called a *continuous* function if the pre-image of any open subset $A \subseteq Y$ given by $f^{-1}(A) = \{x \in X \mid f(x) \in A\}$ is also open.

Def. 2.5. A function $f\colon U \to \mathbb{R}^m$, where $U \subseteq \mathbb{R}^n$ is open, is called of *class* $C^k(U, \mathbb{R}^m)$, $k > 0$ if all partial derivatives of each component $f_j\colon U \to \mathbb{R}$ of f up to order k exist and are continuous:

$$\frac{\partial^k f_j}{\partial x_1^{i_1} \partial x_2^{i_2} \dots \partial x_n^{i_n}} \in C(U), \quad 1 \leq j \leq n, \quad i_1 + i_2 + \cdots + i_n = k$$

Continuous scalar-valued functions on U are elements of the class $C(U)$.
Functions that are $C^k(U)$ for arbitrary k are called *smooth* on U.

Def. 2.6. A map $\phi\colon X \to Y$ between topological spaces X and Y is called a *homeomorphism* if it is bijective, continuous and with continuous inverse.
If there exists a homeomorphism between two topological spaces X, Y, then the spaces are called *homeomorphic* and we write $X \cong Y$.

One can think of the homeomorphic condition as topological spaces being "the same" in the similar sense that isomorphic vector spaces are considered equivalent as sets with vector space structure.
We are now ready to formulate the concept of a *manifold.*

Def. 2.7. (Topological Manifolds)
A topological space $(M, \mathcal{T})$ is called a *topological manifold* if for every $x \in M$, $x \in U \in \mathcal{T}$, there exists a homeomorphism $\phi\colon U \to \mathbb{R}^n$. The pair (U, ϕ) is often called a *coordinate chart* for M.

Remark. Some authors require that a topological manifold be Hausdorff, second countable and admit continuous transition functions. This is involved in the definition of an abstract (topological) manifold, and this is a generalization of our definition.

Recall that functions $f\colon S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ whose domain is not open in the domain space ($\mathbb{R}^n$ here) are called smooth if they may be extended to the function $\tilde{f}\colon U \subset \mathbb{R}^n \to \mathbb{R}^m$ such that $S \subset U$ and $\tilde{f}|_S = f$. In this way we may define smooth functions on arbitrary sets S of some larger Euclidean space. We will assume that all of our manifolds are contained in $\mathbb{R}^N$ for sufficiently large - but finite - N.

Def. 2.8. (Smooth Manifolds)
A manifold M is called *smooth* if for every $x \in M$ and open neighbourhood U around x, the map $\phi\colon U \to \mathbb{R}^n$ is a *diffeomorphism* onto its image $\phi(U)$, i.e. a homemorphism $U \cong \phi(U)$ that is smooth as a map.

A smooth map $F\colon M \to N$ between smooth manifolds M, N can be defined by way of smooth coordinate charts (U, ϕ), (V, ψ) for open sets $U \subseteq M, V \subseteq N$ such that the function is smooth in the Euclidean sense as in Def. 2.5 by the composition $\psi^{-1} \circ \eta \circ \phi$.

$$\begin{tikzcd}
M \arrow[r, "F"] \arrow[d, "\phi"'] & N \arrow[d, "\psi"] \\
\phi(U) \arrow[r, "\eta"] & \psi(V)
\end{tikzcd}$$

Remark. The diagram above commutes if we require, for instance, that $\phi(x) = 0$, $\psi(F(x)) = 0$ and $F(U) \subseteq V$. The first two are always valid assumptions since translations of points in $\mathbb{R}^n$ are diffeomorphisms on $\mathbb{R}^n$, and the latter condition is often sensible given that we can (quite often) assume the image of F to be non-empty in N. The map η as indicated in the diagram is called a *transition map* between $\phi(U)$ and $\psi(V)$.

Def. 2.9. The *tangent space* of a smooth manifold M at a point $p \in M$ is given by the image of the derivative of the *local parametrization* $\phi^{-1} \colon \phi(U) \subseteq \mathbb{R}^n \to M$ corresponding to the coordinate chart (U, ϕ) evaluated at the point which maps to $p \in M$, i.e. the tangent space at p denoted by T_pM is the image of

$$D\phi_0^{-1} = \phi_{*,0}^{-1} \colon T_0U \longrightarrow \mathbb{R}^n$$

where $T_0U = T_0\mathbb{R}^m$ is the tangent space based at $0 \in \mathbb{R}^n$. Here we set $\phi^{-1}(0) = p$ without loss of generality.

Proposition 2.1. The tangent space of $\mathbb{R}^n$ at a point $x \in \mathbb{R}^n$ denoted by $T_x\mathbb{R}^n$ is just the vector space $\mathbb{R}^n$.

Proof. It is easily seen that the identity map $\mathrm{id}_{\mathbb{R}^n} \colon \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism on $\mathbb{R}^n$. The Jacobian matrix of this map is just the identity matrix. If $x \in \mathbb{R}^n$, then

$$D\,\mathrm{id}_{\mathbb{R}^n,x} = [\delta^i_j]_{i,j}.$$

The image of this linear transformation is $T_x\mathbb{R}^n = D\,\mathrm{id}_{\mathbb{R}^n,x}(\mathbb{R}^n) = \mathbb{R}^n$. $\square$

Note that the derivative of the map ϕ^{-1} at $0 \in \mathbb{R}^n$ is the Jacobian matrix $D\phi$ evaluated at 0.

Def. 2.10. (Derivation of Smooth Maps)
Let $F \colon M \to N$ be a smooth map of smooth manifolds, (U, ϕ), (V, ψ) be coordinate charts of M and N respectively such that $\phi(x) = 0$ and $\psi(F(x)) = 0$. Then F admits a *differential at* x given locally by the map

$$F_{*,x} = \psi_{*,y}^{-1} \circ \eta_{*,0} \circ \phi_{*,x}.$$

$$\begin{tikzcd}
T_xM \arrow[r, "F_{*,x}"] \arrow[d, "\phi_{*,x}"'] & T_{F(x)}N \arrow[d, "\psi_{*,y}"] \\
T_0\phi(U) \arrow[r, "\eta_{*,0}"] & T_0\psi(V)
\end{tikzcd}$$

If $F(U) \subseteq V$ the diagram above commutes and the definition of the derivative makes sense. Indeed, one should check that this is well-defined, i.e. our choices of coordinate charts (U, ϕ), (V, ψ) leads to the same derivative.

2.1. **Tangent spaces.** We know what to do with functions $f \colon M \to \mathbb{R}$ where M is not necessarily open in $\mathbb{R}^n$, namely extend the function through bump functions or by finding an open $M \subset W \subseteq \mathbb{R}^n$ and $\tilde{f} \colon W \to \mathbb{R}$ such that $\tilde{f}|_M = f$.

Let $f \colon U \to \mathbb{R}$, $g : V \to \mathbb{R}$ be smooth functions on neighborhoods U, V of the point p. Then their gradients agree on some smaller open set $U \cap V$ if we have $f \equiv g$ on the entire set $U \cap V$, i.e. $\nabla f = \nabla g$ on $U \cap V$. We define the equivalence relation

$$(f, U) \sim (g, V) \iff f|_{U\cap V} = g|_{U\cap V}$$

whose equivalence classes we call *germs* of the functions f, g.

$$[(f, U)] = \{(g, V) \,|\, (f, U) \sim (g, V)\}$$

is the germ of f at p.

Example 2.2. The function $f : (-1, 1) \to \mathbb{R}$ defined by

$$f(x) = \sum_{n=0}^{\infty} x^{2n}$$

and the function $g : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ defined by

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

have equal germs $[f(x)] = [g(x)]$ on the open interval $(-1, 1)$. △

Our reason for introducing this equivalence relation is to view tangent vectors in T_pM of a smooth manifold M as sets of derivations, which we will now define:

Def. 2.11. Let $C_p^\infty(M)$ denote the set of all germs (f, M) of smooth functions $f : M \to \mathbb{R}$ at $p \in M$. A *derivation* at $p \in M$ is a linear map $X : C_p^\infty(M) \to \mathbb{R}$ that satisfies the Leibniz rule on functions $f, g \in C_p^\infty(M)$:

$$X(fg)|_p = X(f)\, g(p) + f(p)\, X(g) \quad \forall\, f, g \in C_p^\infty(M).$$

Def. 2.12. Let S be a product of sets: $S = S_1 \times S_2 \times \cdots \times S_n$. A *coordinate function* $x^i \colon S \to S_i$ picks out the i-th coordinate of a point $p \in S$:

$$p = (p_1, p_2, \dots, p_n) \in S : \quad x^i(p) = p_i \in S_i.$$

Example 2.3. Coordinate-wise derivatives are derivations: for $f, g \in C_p^\infty(U)$

$$\left.\frac{\partial}{\partial x^i}\right|_p f := \frac{\partial f}{\partial x^i}(p).$$

The standard Leibniz rule for derivatives gives us

$$\left.\frac{\partial}{\partial x^i}\right|_p (fg) = \left.\frac{\partial f}{\partial x^i}\right|_p g(p) + f(p) \left.\frac{\partial g}{\partial x^i}\right|_p,$$

which we can see satisfies the derivation property at $p \in U \subseteq M$. △

The reason for using germs lies within the fact that the choice of function for the following definition does not matter.

Def. 2.13. (Tangent Space via Germs)
The tangent space to a smooth manifold M at a point p defined by derivations is given by the set

$$\tilde{T}_pM = \{X \text{ derivation at } p \text{ wrt. } C_p^\infty(M)\}.$$

One can show that, for coordinates x^i around p in M, there exists a one-to-one correspondence of the vector space of point derivations $\tilde{T}_pM$ and the space of coordinate-wise derivations of the form

$$\tilde{T}_pM \ni D_v = \sum_{i=1}^{n} v^i \left.\frac{\partial}{\partial x^i}\right|_p : C_p^\infty(M) \longrightarrow \mathbb{R}.$$

A detailed outline of this correspondence is given in [10]. Using this identification, if $v \in T_pM$ then the corresponding derivation is given by

$$D_v = \sum_{i=1}^{n} v^i \left.\frac{\partial}{\partial x^i}\right|_p .$$

One remarkable fact about this identification is that the tangent space at a point p under this identification becomes the span of all coordinate-wise derivations at the point p. The following theorem summarizes the preceding information.

Theorem 2.1. Let M be a smooth manifold and let $\tilde{T}_pM$ be the space of all point derivations at $p \in M$. Then we have the following:

(i) elements $v \in \tilde{T}_pM$ can be written as

$$v = \sum_{i} v^i \left.\frac{\partial}{\partial x^i}\right|_p$$

where x^i are local coordinates around p.

(ii) there exists a one-to-one correspondence between $\tilde{T}_pM$ and T_pM, i.e. a tangent vector $v \in T_pM$ may be identified with a unique derivation at p.

(iii) under the one-to-one correspondence, we can think of the tangent space T_pM and $\tilde{T}_pM$ as the same space.

Henceforth, we view vectors $X_p \in T_pM$ through the identification of derivations as above. If $p \in U \subseteq M$ and $f \in C_p^\infty(U)$, then $X_p \in T_pM$ acts on f as $X_pf \in \mathbb{R}$.

We can now shift the perspective of the differential to that of tangent spaces of derivations. Let $F : M \to N$ be a smooth map between smooth manifolds. The differential at $p \in M$ is then the linear map $F_{*,p} \colon T_pM \to T_{F(p)}N$ given by

$$(F_{*,p}(X_p))f = X_p(f \circ F) \in \mathbb{R}$$

where f is a germ at $F(p)$: $f \in C^\infty_{F(p)}(N)$.

Def. 2.14. Let M be a smooth manifold. The *tangent bundle* TM of M is the disjoint union of all tangent spaces given by

$$\bigsqcup_{p\in M} T_pM = \bigcup_{p \in M} T_pM = \{(p, X_p) \in M \times T_pM \mid p \in M\}.$$

Def. 2.15. (Vector fields on smooth manifolds)
A *vector field* X on a smooth manifold M is a section $X \colon M \to TM$. A *section* s of the tangent bundle TM is a map $s \colon M \to TM$ such that $\pi \circ s = \mathrm{id}_M$ where $\pi : TM \to M$ is the projection map $\pi(p, X_p) = p$.

Example 2.4. If $M = \mathbb{R}^3$, then an example of a smooth vector field on $\mathbb{R}^3$ is

$$X = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}.$$

Note that for points $(x, y, z) \in \mathbb{R}^3$ we have $\pi \circ X(x, y, z) = (x, y, z)$. △

Def. 2.16. (Lie bracket)
Let X, Y be smooth vector fields on M. The *Lie bracket* of X and Y is the smooth vector field $[X, Y]$ defined by

$$[X, Y] = X \circ Y - Y \circ X = XY - YX$$

where we view the composition as being between derivations.

3. Symplectic Spaces and Symplectic Forms

In the following discussion we will consider linear spaces which may be endowed with a field $\mathbb{F}$, which may be taken to be either $\mathbb{R}$ or $\mathbb{C}$ - as is usually considered in topics within analysis.

We start this section by generalizing what we mean by linear maps on vector spaces over $\mathbb{F}$, and then look at already established notions of symplectic spaces and symplectic forms on vector spaces.

This section is heavily based on the text by Tu [10] and roughly follows the same progression. The material covered in this article will largely be surface level, and the preceding material covers only what is needed to understand the later discussions. Any confusion or deeper discussion of the material should be cleared up in Tu's seminal book.

3.1. **Multilinear Algebra.**

Def. 3.1. Let V be a vector space over $\mathbb{F}$. Denote by V^k the cartesian product $V^k := V \times V \times \cdots \times V$ (k copies). A map $f\colon V^k \to \mathbb{F}$ is called a *k-linear map* if it is linear in each of its k arguments, i.e. for $a, b \in \mathbb{F}$ and $v_i, u, w \in V$ for all i:

$$f(v_1, v_2, \ldots, au + bw, \ldots, v_k) = af(v_1, v_2, \ldots, u, \ldots, v_k) + bf(v_1, v_2, \ldots, w, \ldots, v_k).$$

In particular, for $k = 2$ the map is called *bilinear*, and for $k = 3$ *trilinear*.

Remark. k-linear maps on V are often called *k-tensors* on V, but usually with the restriction that $\mathbb{F} = \mathbb{R}$.

Def. 3.2. Let $(S_k, \circ)$ be the permutation group of order k with its group action $\circ$ as the composition of permutations. Let $f\colon V^k \to \mathbb{F}$ be a k-linear map. The map is called, $\forall\, \sigma \in S_k$

- *symmetric* if $f(v_1, v_2, \ldots, v_k) = f(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(k)})$
- *anti-symmetric* if $f(v_1, v_2, \ldots, v_k) = \operatorname{sgn}(\sigma) f(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(k)})$

When f is anti-symmetric, it has a non-trivial *signature* denoted by $\operatorname{sgn}(\sigma)$. The signature changes the prefactor to ± 1 based on whether the permutation is symmetric or anti-symmetric.

Example 3.1. An important example of an anti-symmetric map is the bilinear *wedge product*, which for $f, g\colon V^2 \to \mathbb{F}$, is defined by

$$(f \wedge g)(u, v) = f(u)g(v) - f(v)g(u).$$

Remark. From any multilinear function $f \in \mathcal{A}^k(V)$, we may construct its anti-symmetric or *alternating* counterpart by forcing permutations to give a signature prefactor. This can be achieved by the function

$$Af = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma)\, \sigma f.$$

We see that $\tau(Af) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma)\, \tau\sigma f = \operatorname{sgn}(\tau) \sum_{\sigma \in S_k} \operatorname{sgn}(\tau\sigma)\, \tau\sigma f = \operatorname{sgn}(\tau) Af$. Thus Af is anti-symmetric.

Remark. Being symmetric or anti-symmetric is group-theoretic in nature, i.e. a permutation σ from the permutation group S_k is symmetric or anti-symmetric depending on whether it can be expressed as a composition of even or odd many disjoint transpositions $\tau_i \colon i \mapsto \sigma(i)$ such that $\sigma = \tau_{j_m} \circ \cdots \circ \tau_{j_1}$. More details on groups, and in particular permutation groups can be found in Fraleigh's text [6].

Def. 3.3. The space of all k-linear maps on V^k are denoted $\mathcal{L}^k(V)$.
The space of all k-linear maps which are symmetric is denoted by $\mathcal{S}^k(V)$. Likewise, the space of all k-linear maps which are anti-symmetric (*alternating*) is denoted by $\mathcal{A}^k(V)$.

Def. 3.4. Let f be a k-linear, g be l-linear on V, then the *tensor product* $f \otimes g$ is a $(k+l)$-linear function $f \otimes g \colon V^{k+l} \to \mathbb{F}$, defined by

$$(f \otimes g)(v_1, \ldots, v_k, v_{k+1}, \ldots v_{k+l}) = f(v_1, \ldots, v_k) g(v_{k+1}, \ldots v_{k+l}).$$

The tensor product is intuitive since it produces as its "product" another multilinear function. We wish to formulate such a concept for alternating functions as well, since the tensor product of two alternating functions need not be alternating: Let $f, g \in \mathcal{A}^2(V)$, then we have $(f \otimes g)(v_1, v_2, v_3, v_4) = f(v_1, v_2) g(v_3, v_4)$. If we apply the transposition $\sigma = (2\,3)$, then $\operatorname{sgn}(\sigma) = -1$, but

$$\sigma(f \otimes g)(v_1, v_2, v_3, v_4) = f(v_1, v_3) g(v_2, v_4) \neq -f(v_1, v_2) g(v_3, v_4)$$

in general. This problem motivates the definition of a modified tensor product that preserves anti-symmetry:

Def. 3.5. (Wedge Product):
Let $f \in \mathcal{A}^k(V)$, $g \in \mathcal{A}^l(V)$. We define their *wedge product* to be the $(k+l)$-linear, alternating function $f \wedge g \in \mathcal{A}^{k+l}(V)$ given by

$$f \wedge g = \frac{1}{k!\, l!} A(f \otimes g)$$

where $A(f \otimes g)$ is given as in the remark from before.

Theorem 3.1. (Properties of the Wedge Product):
Let $f \in \mathcal{A}^k(V)$, $g \in \mathcal{A}^l(V)$, $h \in \mathcal{A}^n(V)$, then the following holds true for their wedge products:

(i) $f \wedge g = (-1)^{kl} g \wedge f$ (Anti-commutativity)
(ii) $f \wedge (g \wedge h) = (f \wedge g) \wedge h$ (Associativity)
(iii) if k is odd, then $f \wedge f = 0$

Proof. (i) Let $f \in \mathcal{A}^k(V)$, $g \in \mathcal{A}^l(V)$, then we have by definition

$$\begin{aligned} f \wedge g(v_1, \ldots, v_{k+l}) &= \frac{1}{k!\, l!} A(f \otimes g)(v_1, \ldots, v_{k+l}) \\ &= \frac{1}{k!\, l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma) f(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) g(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}), \\ g \wedge v(v_1, \ldots, v_{k+l}) &= \frac{1}{k!\, l!} A(g \otimes f)(v_1, \ldots, v_{k+l}) \\ &= \frac{1}{k!\, l!} \sum_{\tau \in S_{k+l}} \operatorname{sgn}(\tau) g(v_{\tau(1)}, \ldots, v_{\tau(k)}) f(v_{\tau(k+1)}, \ldots, v_{\tau(k+l)}) \end{aligned}$$

Then we can find a permutation $\eta \in S_{k+l}$ for each τ appearing in the preceding sum such that $\tau \circ \eta(k) = \sigma(k)$ for corresponding $\sigma \in S_{k+l}$ in the first sum. This permutation can be written in terms of transpositions

$$\begin{aligned} t_k &= (k\ k+l)(k\ k+l-1)\cdots(k\ k+1) \\ t_{k-1} &= (k-1\ k+l)(k-1\ k+l-1)\cdots(k-1\ k+1) \\ &\ \vdots \\ t_1 &= (1\ k+l)(1\ k+l-1)\cdots(1\ k+1) \end{aligned}$$

where $\eta = t_1 \circ t_2 \circ \cdots \circ t_k$. Since $\mathrm{sgn}(t_j) = (-1)^l$ for all $1 \leq j \leq k$, this means that $\mathrm{sgn}(\eta) = (-1)^{kl}$, and the equality holds.
(ii) This follows from associativity of the tensor product and a simple calculation.
(iii) This is an immediate corollary from (i). □

We are now ready to explore certain aspects of the vector space structure of $\mathcal{A}^k(V)$. Let $\dim(V) = n$. A collection $\{\alpha_j\}_{j\in J} \subset \mathcal{A}^k(V)$ is said to be linearly independent if for $a_j \in \mathbb{F}$

$$\sum_{j\in J} a_j\alpha_j = 0 \iff a_j = 0 \quad \forall j \in J,$$

where $0 \in \mathcal{A}^k(V)$ is the element which sends every vector to $0 \in \mathbb{F}$. We introduce canonical 1-covectors $\alpha^i \in \mathcal{A}^1(v)$ defined by $\alpha^i(e_j) = \delta^i_j$, where $\{e_1, e_2, \ldots, e_n\}$ is the canonical basis of V and δ^i_j is the Kronecker delta. An indexing set I of length k is a set $\{i_1, i_2, \ldots, i_k\}$ where $1 \leq i_j \leq n$ for $j = 1, 2, \ldots, k$. The k-covector β^I is defined as $\beta^I = \beta^{i_1} \wedge \beta^{i_2} \wedge \cdots \wedge \beta^{i_k}$ for $\beta \in \mathcal{A}^1(V)$. Thus we have that $\alpha^I(e_{j_1}, e_{j_2}, \ldots, e_{j_k}) = \delta^I_{\{j_1,j_2,\ldots,j_k\}}$, which leads us to the following result:

Proposition 3.1. Let V be a vector space. For $\alpha^i\colon V \to \mathbb{F}$ defined as above, we have that $\{\alpha^I\}_I$ for index sets I of length k form a basis for the vector space $\mathcal{A}^k(V)$.

Remark. We have not explicitly proven that $\mathcal{A}^k(V)$ is a vector space over $\mathbb{F}$, although this is easily verified.

Proof. We inspect the following sum, for $a_I \in \mathbb{F}$,

$$\sum_I a_I\alpha^I = 0$$

where $\alpha^I = \alpha^{i_1} \wedge \alpha^{i_2} \wedge \cdots \wedge \alpha^{i_k}$ when $I = \{i_1, i_2, \ldots, i_k\}$. We apply $e_J := (e_{j_1}, e_{j_2}, \ldots, e_{j_k})$ to this k-linear function

$$\sum_I a_I\alpha^I(e_J) = \sum_I a_I\delta^I_J = a_J = 0.$$

Hence $a_J = 0$ for all indexing sets J of length k since J was arbitrary. This proves that such a collection is linearly independent. Now, consider a k-linear function $f \in \mathcal{A}^k(V)$. Because of linearity it is enough to consider how f acts on e_J: $f(e_J) \in \mathbb{F}$.

$$\sum_I a_I\alpha^I(e_J) = a_J$$

Let $a_J = f(e_J)$. This implies that for $v_1, v_2, \ldots, v_k \in V$

$$f(v_1, v_2, \ldots, v_k) = \sum_J b_J f(e_J) = \sum_J b_J \Big(\sum_I a_I \alpha^I(e_J) \Big)$$
$$= \sum_I a_I \alpha^I \Big(\sum_J b_J e_J \Big) = \sum_I a_I \alpha^I (v_1, v_2, \ldots, v_k).$$

Both of these k-linear functions are thus the same, and the collection $\{\alpha_I\}_I$ span the vector space $\mathcal{A}^k(V)$. □

Remark. The attentive reader may have realized that in these sums of k-linear functions, some of the α^I are equal up to a sign for different I. An example for $f \in \mathcal{A}^2(\mathbb{F}^2)$ is, using the previous result and Theorem 3.1:

$$f = a_{12}\, \alpha^1 \wedge \alpha^2 + a_{21}\, \alpha^2 \wedge \alpha^1 = (a_{12} - a_{21})\, \alpha^1 \wedge \alpha^2.$$

This means that we may, and henceforth will, assume that every indexing set I is strictly ascending: $I = (i_1 < i_2 < \cdots < i_k)$.

Assuming we want k-vectors on manifolds, how would such a structure look like? So far we have used a vector space V in the definition of the space of k-vectors $\mathcal{A}^k(V)$. The tangent space is a natural vector space to consider, since T_pM is defined for any $p \in M$ and has an easy to compute local expression given a coordinate chart (U, ϕ) centered around p. This leads us to the construction of differential forms on M.

3.2. **Differential Forms.**

In the remainder of this section we set $\mathbb{F} = \mathbb{R}$.

Def. 3.6. Let M be a smooth manifold. The *cotangent space* T_p^*M is the dual of the tangent space T_pM for $p \in M$, i.e. the space of all linear functions

$$\omega_p \colon T_pM \to \mathbb{R}.$$

A 1-*form* on M is a function ω which assigns a 1-linear function ω_p for every point $p \in M$. The *cotangent bundle* T^*M is the disjoint union of cotangent spaces T_p^*M

$$\bigsqcup_{p \in M} T_p^*M = \bigcup_{p \in M} T_p^*M = \{(p, \omega_p) \in M \times T_p^*M \mid p \in M\}.$$

Def. 3.7. Let (U, ϕ) be a coordinate chart where $p \in U \subseteq M$. We write local coordinates on U as $x^i = r^i \circ \phi$ where r^i are coordinate functions on $\mathbb{R}^n$. We define the 1-form $\mathrm{d}x^i$, $1 \le i \le n$, pointwise as

$$(\mathrm{d}x^i)_p \left(\left. \frac{\partial}{\partial x^j} \right|_p \right) = \delta^i_j.$$

There is a natural way to define 1-forms out of 0-*forms*, which are smooth functions on M. If $f \in C^\infty(M)$, then the 1-form $\mathrm{d}f$ can be defined pointwise for $X_p \in T_pM$:

$$(\mathrm{d}f)_p(X_p) = X_p f.$$

We state two propositions pertaining to Def. 3.7 that will be useful to keep in mind. The proofs of these can be found in Tu [10], although it should not be difficult to prove this alone.

Proposition 3.2. Let M be a smooth manifold, $f \in C^\infty(M)$, $(U, x^1, \ldots, x^n)$ be a coordinate chart on M, then we have

(i) $\{(\mathrm{d}x^1)_p, (\mathrm{d}x^2)_p, \dots, (\mathrm{d}x^n)_p\}$ is a basis for T_p^*M for every $p \in M$.
(ii) The 1-form $\mathrm{d}f$ has a local expression on an open subset $U \subseteq M$ given by

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \mathrm{d}x^i.$$

Example 3.2. Consider the smooth manifold $S^1 \subseteq \mathbb{R}^2$ where U is the half-circle $U = \{\, (x, y) \in \mathbb{R}^2 \,|\, 0 < x \le 1, -1 < y < 1 \,\}$. A local parametrization for U is

$$\phi^{-1} \colon (-1, 1) \to U, \quad \phi^{-1}(\theta) = (\sqrt{1-\theta^2}, \theta).$$

Calculating the differential at θ using the local coordinates gives

$$\phi^{-1}_{*,\theta} = \frac{\theta}{\sqrt{1-\theta^2}} \frac{\partial}{\partial x}\bigg|_{\phi^{-1}(\theta)} + \frac{\partial}{\partial y}\bigg|_{\phi^{-1}(\theta)}.$$

Letting $\theta = 0 : \phi^{-1}(0) = (1, 0)$ we have

$$\phi^{-1}_{*,0} = \frac{\partial}{\partial y}\bigg|_{(1,0)}.$$

Thus every vector $X_{(1,0)}$ in the tangent space $T_{(1,0)}S^1$ has the form

$$X_{(1,0)} = a \frac{\partial}{\partial y}\bigg|_{(1,0)}, \quad a \in \mathbb{R}.$$

Now consider the function $f : S^1 \to \mathbb{R}$, $f(x, y) = x^2 + y^2$

$$\mathrm{d}f = 2x \,\mathrm{d}x + 2y \,\mathrm{d}y$$

Evaluating at the point $(1, 0) \in S^1$ and $X_{(1,0)} \in T_{(1,0)}S^1$ gives

$$\begin{aligned}(\mathrm{d}f)_{(1,0)}(X_{(1,0)}) &= 2(1)\mathrm{d}x_{(1,0)} \left(0 \frac{\partial}{\partial x}\bigg|_{(1,0)} \right) + 2(0)\mathrm{d}y_{(1,0)} \left(a \frac{\partial}{\partial y}\bigg|_{(1,0)} \right) \\ &= 2 \cdot 0 + 0 \cdot a = 0 \quad \forall a \in \mathbb{R}.\end{aligned}$$

△

Differential forms exhibit local behaviour similar to that of the global behaviour of k-linear functions on vector spaces.

Def. 3.8. Let $k \ge 1$ be a natural number, $(U, x^1, \dots, x^n)$ coordinate chart on M. We define a *differential k-form* ω on U to be k-linear functions on the tangent spaces when evaluated pointwise.

The space of differential k-forms on $U \subseteq M$ is denoted by $\Omega^k(U)$.

Remark. We note that $\mathrm{d}x^I := \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}$ is a k-form with the properties similar to that of the k-linear function α^I as before.

Proposition 3.3. Let M be a smooth manifold, $(U, x^1, \dots, x^n)$ a coordinate chart, then $\{\mathrm{d}x^I\}_I$ for strictly ascending indexing sets I of length k is a basis for $\Omega^k(U)$.

Def. 3.9. The *external derivative* of a k-form ω on M, is defined to be the $k+1$-form $\mathrm{d}\omega$ locally given by

$$\mathrm{d}\omega = \sum_I \mathrm{d}a_I \wedge \mathrm{d}x^I = \sum_I \left(\sum_j \frac{\partial a_I}{\partial x^j} \mathrm{d}x^j \right) \wedge \mathrm{d}x^I \in \Omega^{k+1}(U)$$

where $\omega = \sum_I a_I \mathrm{d}x^I$ for a coordinate chart $(U, x^1, \dots, x^n)$.

For our purposes we only really need the exterior derivative in the local perspective, however it can be shown that there exists a unique exterior derivative $\mathrm{d} : \Omega^*(M) \to \Omega^*(M)$, where

$$\Omega^*(M) = \bigoplus_{k=0}^{\infty} \Omega^k(M).$$

This exterior derivative acts globally and is not dependent on a coordinate chart. More details on this may be found in Section 19 in Tu [10].

A k-form ω on M is called a *top form* if $\deg(\omega) = \dim(M)$, i.e. if $\omega \in \Omega^{\dim(M)}(M)$. A k-form α is said to be *exact* if there exists a $(k-1)$-form β such that $\alpha = \mathrm{d}\beta$.

Proposition 3.4. (Properties of the exterior derivative)
Given a k-form $\omega \in \Omega^k(M)$, then the following properties hold:

(i) $\mathrm{d}(\mathrm{d}\omega) = 0$, equivalently: $\mathrm{d} \circ \mathrm{d} = 0$.
(ii) For exact 1-forms $\mathrm{d}f^1, \mathrm{d}f^2, \dots, \mathrm{d}f^k$ their wedge products are also zero: $\mathrm{d}f^1 \wedge \mathrm{d}f^2 \wedge \cdots \wedge \mathrm{d}f^k = 0$.

Proposition 3.5. Let $(U, x^1, \dots, x^n)$ be a coordinate chart on a smooth manifold M, and $f^1, f^2, \dots, f^k$ smooth functions on U. Then

$$\mathrm{d}f^1 \wedge \mathrm{d}f^2 \wedge \cdots \wedge \mathrm{d}f^k = \sum_I \frac{\partial(f^1, \dots, f^k)}{\partial(x^{i_1}, \dots, x^{i_k})} \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}$$

where

$$\frac{\partial(f^1, \dots, f^k)}{\partial(x^{i_1}, \dots, x^{i_k})} := \det \left[\frac{\partial f^m}{\partial x^{i_j}} \right]_{1 \leq m, j \leq k}.$$

3.3. **Digression on Vector Fields and Flows.**

A vector field X is called smooth if it is smooth as a section of the tangent bundle $X \colon M \to TM$. If $(U, x^1, \dots, x^n)$ is a coordinate chart containing $p \in M$, then

$$(X)|_p = X_p = \sum_i a_i(p) \frac{\partial}{\partial x^i}\bigg|_p, \quad a_i \in C^\infty(U)$$

is the pointwise evaluation of X at $p \in U$. Extrapolating to the entirety of $U \subseteq M$ gives

$$X = \sum_i a_i \frac{\partial}{\partial x^i}, \quad a_i \in C^\infty(U).$$

Def. 3.10. (Integral curves of vector fields)
Let X be a smooth vector field on M. An *integral curve of* X is a smooth curve $c \colon I \subseteq \mathbb{R} \to M$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} c(t) = X_{c(t)}, \quad t \in I.$$

Since the given vector field X has an expression given locally for a coordinate chart $(U, x^1, \dots, x^n)$ given by

$$X_p = \sum_i a_i(p) \frac{\partial}{\partial x^i}\bigg|_p,$$

then we may as well set $p = c(t)$ for $t \in I$, which leads to

$$X_{c(t)} = \sum_i a_i(c(t)) \frac{\partial}{\partial x^i}\bigg|_{c(t)}.$$

Using the condition of being an integral curve of the vector field X:

$$\frac{\mathrm{d}}{\mathrm{d}t} c(t) = \sum_i \dot{c}_i(t) \frac{\partial}{\partial x^i}\bigg|_{c(t)}.$$

Comparing components we have that integral curves have to locally satisfy $\dot{c}_i(t) = a_i(c(t))$ for all i. By the Picard-Lindelöf theorem we have that this system of ordinary differential equations admits a solution $c(t)$ defined on $(-\epsilon, \epsilon) \times W$ where $\epsilon > 0$, $(-\epsilon, \epsilon) \subseteq I$ and $W \subseteq U$ open. This guarantees the existence and uniqueness of integral curves given a locally defined smooth vector field X.

An integral curve $c\colon I \to M$ is said to start at the point $p \in M$ if $c(0) = p$. Again, the existence of such a curve is guaranteed by Picard-Lindelöf (see Brezis [3]). If we vary the starting point $p \in U$ smoothly, considering the solutions of

$$\frac{\mathrm{d}}{\mathrm{d}t} c(t) = X_{c(t)}, \quad c(0) = p$$

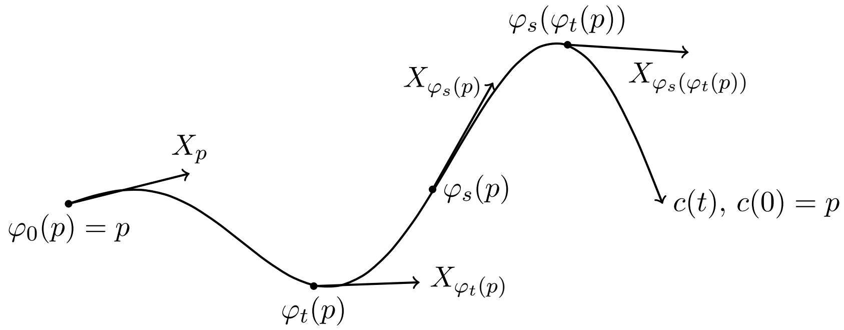
independently for each $p \in U$, then defining the function

$$\begin{aligned} \varphi_t :\ (-\epsilon, \epsilon) \times W &\longrightarrow U, \\ (t, q) \quad &\longmapsto \varphi_t(q) := \varphi(t, q) \end{aligned}$$

such that $\varphi(t, p) = c(t)$, the integral curve of X starting at the given point $p \in U$, then we call $\varphi_t(q)$ the *local flow* of X on U. This function is smooth on $(-\epsilon, \epsilon) \times W$. Note that $\varphi_0(q) = \varphi(0, q) = q$. The flow associated to the vector field X has special properties in regards to how it composes with itself at different parameter values $t, s \in (-\epsilon, \epsilon)$:

$$\varphi_{t+s}(q) = \varphi_s(\varphi_t(q)) \ \text{ whenever } \ t + s \in (-\epsilon, \epsilon).$$

To see this, it is helpful to look at a picture of an integral curve $c(t)$ of a particular vector field X where $c(0) = p$.



Assuming $t \in (-\epsilon, \epsilon)$, then we have $\varphi_t \circ \varphi_{-t} = \varphi_{-t} \circ \varphi_t = \mathrm{id}_M$. This means that a local flow φ_t is a local diffeomorphism of M onto M.

3.4. **Pullbacks of Differential Forms.**

As we have seen, it is possible to map vectors linearly from one tangent space to another tangent space through the differential of a smooth function between the two manifolds. In this subsection we will see the analogue case for differential k-forms.

The way we do this for 1-forms is through the pullback

$$F^*\colon T^*_{F(p)}N \longrightarrow T^*_pM$$

which sends $\omega_{F(p)}$ to the 1-form $F^*(\omega_{F(p)})$: $\omega_{F(p)}$ is *pulled back* by F.

Def. 3.11. (Pullback of a 1-form)
Let $F\colon M \to N$ be a smooth map between smooth manifolds. The *pullback*

$$F^*\colon T^*_{F(p)}N \longrightarrow T^*_pM$$

is the map of 1-forms defined pointwise by

$$F^*(\omega_{F(p)})(X_p) = \omega_{F(p)}(F_{*,p}(X_p))$$

where $X_p \in T_pM$. Using this, we may define the pullback between covector fields $F^*\colon T^*N \to T^*M$ pointwise by $(F^*\omega)_p = F^*(\omega_{F(p)})$ for all points $p \in M$.

Remark. Note the aesthetic similarities between the definitions of the differential F_* and the pullback of 1-forms F^*:

$$\begin{aligned} (F_{*,p}(X_p))f &= X_p(f \circ F), \quad f \in C^\infty_{F(p)}(N) \\ (F^*(\omega_{F(p)}))(X_p) &= \omega_{F(p)}(F_{*,p}(X_p)). \end{aligned}$$

Noting these similarities might make it easier to remember how the pullback acts on a vector X_p, just by remembering the definition of the differential!

What about pullbacks of 0-forms, i.e. pullbacks of smooth functions? Let $F : M \to N$ be a smooth function between smooth manifolds. If $g \in C^\infty(N)$, then $F^*g = g \circ F \in C^\infty(M)$ is the pullback of g by F.

Proposition 3.6. Let $F\colon M \to N$ be a smooth map between smooth manifolds. Then for any $h \in C^\infty(M)$ we have

$$F^*(\mathrm{d}h) = \mathrm{d}(F^*h).$$

Proof. Let $p \in M$. Evaluating the 1-form $F^*(\mathrm{d}h)$ pointwise we obtain

$$\begin{aligned} (F^*(\mathrm{d}h)_{F(p)})(X_p) &= \mathrm{d}h_{F(p)}(F_{*,p}(X_p)) \\ &= (F_{*,p}(X_p))h = X_p(h \circ F) \end{aligned}$$

for some $X_p \in T_pM$. Considering the right-hand side of the equality

$$(\mathrm{d}F^*h)_p(X_p) = X_p(F^*h) = X_p(h \circ F),$$

which establises the equality pointwise for arbitrary $p \in M$. □

It is a good exercise to see why the equalities in the preceding proof hold.

Proposition 3.7. For 1-forms ω, τ on a smooth manifold N, then for a smooth map $F\colon M \to N$ we have

$$F^*(\omega + \tau) = F^*\omega + F^*\tau.$$

Proof. Using the same principle as the previous proof, we evaluate pointwise for $X_p \in T_pM$:

$$\begin{aligned} F^*((\omega + \tau)_{F(p)})(X_p) &= (\omega + \tau)_{F(p)}(F_{*,p}(X_p)) \\ &= \omega_{F(p)}(F_{*,p}(X_p)) + \tau_{F(p)}(F_{*,p}(X_p)) \\ &= F^*\omega_{F(p)}(X_p) + F^*\tau_{F(p)}(X_p). \end{aligned}$$

The point $p \in M$ was arbitrary, and thus the equality holds. □

With pullbacks of 0- and 1-forms in mind, it remains to see how we are able to pull back differential k-forms. It turns out that this is not too different from the 1-form case.

It is customary to write $\bigwedge^k(T_p^*M)$ for k-linear functions $\omega_p = \omega|_p$ where $\omega \in \Omega^k(M)$ is a differential k-form on the smooth manifold M. Notice that in our conventions this means that $\bigwedge^k(T_p^*M) := \mathcal{A}^k(T_pM)$.

Def. 3.12. (Pullbacks of k-forms)
Let $F\colon M \to N$ be a smooth map between smooth manifolds. The pullback of a k-form $\omega \in \Omega^k(N)$ is the k-form $F^*\omega \in \Omega^k(M)$. The pullback as a map

$$F^*\colon \bigwedge^k(T_{F(p)}^*N) \longrightarrow \bigwedge^k(T_p^*M)$$

is defined through pointwise evaluation on a k-form $\omega_{F(p)} \in \bigwedge^k(T_{F(p)}^*N)$ as

$$F^*(\omega_{F(p)})(X_1, \dots, X_k) = \omega_{F(p)}(F_{*,p}(X_1), \dots, F_{*,p}(X_k))$$

where $X_1, \dots, X_k \in T_pM$.

Proposition 3.8. Let $F\colon M \to N$ be a smooth map between smooth manifolds. For $\omega, \tau \in \Omega^k(M)$ and $a \in \mathbb{R}$, we have

(i) $F^*(\omega + \tau) = F^*\omega + F^*\tau$,
(ii) $F^*(a\omega) = aF^*(\omega)$.

The proof of these claims is rather trivial and is a good check of one's understanding of what the pullback does.

Proposition 3.9. (Pullback of wedge products)
Let $F\colon M \to N$ be a smooth map between smooth manifolds. If ω, τ are differential forms on N, then

$$F^*(\omega \wedge \tau) = F^*(\omega) \wedge F^*(\tau).$$

Proof. Assuming $\omega \in \Omega^k(N)$, $\tau \in \Omega^l(N)$ for some k, l, we have that

$$\begin{aligned}
&(F^*(\omega \wedge \tau))(X_1, \dots, X_k, X_{k+1}, \dots, X_{k+l}) \\
&= (\omega \wedge \tau)_{F(p)}(F_{*,p}(X_1), \dots, F_{*,p}(X_k), F_{*,p}(X_{k+1}), \dots, F_{*,p}(X_{k+l})) \\
&= \omega_{F(p)} \wedge \tau_{F(p)}(F_{*,p}(X_1), \dots, F_{*,p}(X_k), F_{*,p}(X_{k+1}), \dots, F_{*,p}(X_{k+l})) \\
&= \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma)\omega_{F(p)}(F_{*,p}(X_{\sigma(1)}), \dots, F_{*,p}(X_{\sigma(k)})) \cdot \tau_{F(p)}(F_{*,p}(X_{\sigma(k+1)}), \dots, F_{*,p}(X_{\sigma(k+l)})) \\
&= \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma)(F^*\omega)_p(X_{\sigma(1)}, \dots, X_{\sigma(k)}) \cdot (F^*\tau)_p(X_{\sigma(k+1)}, \dots, X_{\sigma(k+l)}) \\
&= (F^*\omega \wedge F^*\tau)_p(X_1, \dots, X_{k+l}).
\end{aligned}$$

Since $p \in M$ and $X_1, \dots, X_{k+l} \in T_pM$ are all arbitrary, this means that $F^*(\omega \wedge \tau) = F^*\omega \wedge F^*\tau$ - as desired. ☐

Proposition 3.10. Let $F\colon M \to N$ be a smooth map between smooth manifolds, $\omega \in \Omega^*(M)$. Then the pullback and exterior derivative "commutes":

$$\mathrm{d}F^*\omega = F^*\mathrm{d}\omega.$$

Proof. We begin by considering the differential form ω locally on N. Let $(V, x^1, \dots, x^m)$ be a chart on N around $F(p)$, then

$$\omega = \sum_I a_I \, \mathrm{d}x^I, \quad a_I \in C^\infty(V).$$

Applying the pullback map to both sides and using Proposition 3.8:

$$F^*\omega = \sum_I F^*(a_I) F^*(\mathrm{d}x^I).$$

Then, note the formula holds especially for 0-forms:

$$\begin{aligned}
&(F^*\mathrm{d}x^i)_p(X_p) = (\mathrm{d}x^i)_{F(p)}(F_{*,p}(X_p)) = \sum_j \frac{\partial F^i}{\partial x^j}\bigg|_p \mathrm{d}x^j_p(X_p).\\
&\mathrm{d}F^*x^i = \mathrm{d}(F^*x^i) = \mathrm{d}(x^i \circ F) = \mathrm{d}F^i,\\
&\mathrm{d}F^i_p(X_p) = \sum_j \frac{\partial F^i}{\partial x^j}\bigg|_p \mathrm{d}x^j_p(X_p).
\end{aligned}$$

Here we have used the local form of the differential $F_{*,p}$. Then we have

$$\begin{aligned}
F^*\mathrm{d}x^I &= F^*(\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}) = F^*(\mathrm{d}x^{i_1}) \wedge \cdots \wedge F^*(\mathrm{d}x^{i_k})\\
&= \mathrm{d}F^{i_1} \wedge \cdots \wedge \mathrm{d}F^{i_k}
\end{aligned}$$

using Proposition 3.9 and the case for 0-forms. Having proven this, we establish that

$$\mathrm{d}F^*\omega = \sum_I \mathrm{d}(a_I \circ F) \wedge \mathrm{d}F^I.$$

Consider now the right-hand side of the formula:

$$\begin{aligned}
F^*(\mathrm{d}\omega) = F^*\left(\sum_I \mathrm{d}a_I \wedge \mathrm{d}x^I\right) &= \sum_I F^*\mathrm{d}a_I \wedge F^*\mathrm{d}x^I\\
&= \sum_I \mathrm{d}(F^*a_I) \wedge \mathrm{d}F^I = \sum_I \mathrm{d}(a_I \circ F) \wedge \mathrm{d}F^I.
\end{aligned}$$

From this we see that $\mathrm{d}F^*\omega = F^*\mathrm{d}\omega$ by comparison. □

3.5. The Lie Derivative and Interior Multiplication.

How do we differentiate objects over a smooth manifold M? We have seen curves $c \colon I \to M$ parametrized onto M being differentiated, but what about vector fields and differential forms? The Lie derivative happens to at least partially answer this question, and will be useful to us later.

A 1-parameter family of vector fields is a set $\{X_t\}_{t \in I}$ where $I \subseteq \mathbb{R}$. Similarly for a 1-parameter family of differential k-forms $\{\omega_t\}_{t \in I}$.

A limit $\lim_{t \to t_0} X_t$ exists if for every $p \in M$ admits a coordinate chart $(U, x^1, \dots, x^n)$ on which

$$(X_t)_p = \sum_i a_i(t,p) \frac{\partial}{\partial x^i}\bigg|_p$$

and $\lim_{t\to t_0} a_i(t,p)$ exists for all i. Given that the limit exists, we can write

$$\lim_{t\to t_0}(X_t)_p = \sum_i \lim_{t\to t_0} a_i(t,p)\,\frac{\partial}{\partial x^i}\bigg|_p .$$

A 1-parameter family $\{X_t\}_{t\in I}$ is said to be smoothly dependent on the parameter t if every point $p \in M$ admits a coordinate chart $(U, x^1, \ldots, x^n)$ such that we can write

$$(X_t)_p = \sum_i a_i(t,p)\,\frac{\partial}{\partial x^i}\bigg|_p, \quad (t,p) \in I \times U.$$

With smoothness of the 1-parameter family in mind we may define the derivative of this 1-parameter family of vector fields:

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=t_0} X_t\right)_p = \sum_i \frac{\partial a_i}{\partial t}(t_0,p)\,\frac{\partial}{\partial x^i}\bigg|_p$$

Of course, this means that the derivative is also a smooth varying 1-parameter family of vector fields on M.

Likewise, a 1-parameter family $\{\omega_t\}_{t\in I}$ is said to be smoothly dependent on t if every point $p \in M$ admits a coordinate chart $(U, x^1, \ldots, x^n)$ on which

$$(\omega_t)_p = \sum_J b_J(t,p)\,\mathrm{d}x^J\big|_p, \quad b_J \in C^\infty(I\times U).$$

We define its derivative at $t = t_0$ by

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=t_0} \omega_t\right)_p = \sum_J \frac{\partial b_J}{\partial t}(t_0,p)\,\mathrm{d}x^J\big|_p.$$

Proposition 3.11. Let $\{\omega_t\}_t$ and $\{\tau_t\}_t$ be smooth 1-parameter families of differential forms. Then derivation by t of their wedge product respects the product rule

$$\frac{\mathrm{d}}{\mathrm{d}t}(\omega_t \wedge \tau_t) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\omega_t\right) \wedge \tau_t + \omega_t \wedge \left(\frac{\mathrm{d}}{\mathrm{d}t}\tau_t\right).$$

Proof. Let $(U, x^1, \ldots, x^n)$ be a coordinate chart on M. Locally we may write

$$\omega_t = \sum_I a_I\,\mathrm{d}x^I, \quad \tau_t = \sum_J b_J\,\mathrm{d}x^J, \quad a_I,\, b_J \in C^\infty(I\times U)$$

where I, J in the sums are indexing sets. Differentiating by the parameter t for each of these differential forms gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\omega_t = \sum_I \frac{\partial a_I}{\partial t}\,\mathrm{d}x^I, \quad \frac{\mathrm{d}}{\mathrm{d}t}\tau_t = \sum_J \frac{\partial b_J}{\partial t}\,\mathrm{d}x^J.$$

In forming the wedge product of ω_t and τ_t we know that for non-disjoint indexing sets $I' \cap J' \neq \varnothing$, the wedge product $\mathrm{d}x^{I'} \wedge \mathrm{d}x^{J'}$ vanishes. However, the sum is not affected by a contribution of 0, so we keep these summands and write

$$\omega_t \wedge \tau_t = \sum_I \sum_J a_I \cdot b_J\,\mathrm{d}x^I \wedge \mathrm{d}x^J.$$

Differentiating with respect to t gives, by the usual product rule:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\omega_t \wedge \tau_t) &= \sum_I \sum_J \frac{\partial a_I}{\partial t} \cdot b_J \,\mathrm{d}x^I \wedge \mathrm{d}x^J + \sum_I \sum_J a_I \cdot \frac{\partial b_J}{\partial t} \,\mathrm{d}x^I \wedge \mathrm{d}x^J \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\omega_t\right) \wedge \tau_t + \omega_t \wedge \left(\frac{\mathrm{d}}{\mathrm{d}t}\tau_t\right).\end{aligned}$$

□

Proposition 3.12. Let $\{\omega_t\}_t$ be a smooth family of differential forms on M. If $\mathrm{d}\colon \Omega^*(M) \to \Omega^*(M)$ is the external derivative, then it "commutes" with differentiation by t, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{d}\omega_t) = \mathrm{d}\left(\frac{\mathrm{d}}{\mathrm{d}t}\omega_t\right).$$

Proof. Considering a coordinate chart $(U, x^1, \ldots, x^n)$ we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{d}\omega_t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_I \sum_j \frac{\partial a_I}{\partial x^j}\mathrm{d}x^j \wedge \mathrm{d}x^I\right) = \sum_I \sum_j \frac{\partial}{\partial x^j}\frac{\partial a_I}{\partial t}\mathrm{d}x^j \wedge \mathrm{d}x^I \\ &= \mathrm{d}\left(\sum_I \frac{\partial a_I}{\partial t}\,\mathrm{d}x^I\right) = \mathrm{d}\left(\frac{\mathrm{d}}{\mathrm{d}t}\omega_t\right)\end{aligned}$$

where the second equality is due to the smoothness of a_I. □

Our next task is to construct the Lie derivative of a vector field. Recall that a smooth vector field X on M may give rise to a local flow φ_t starting at $p \in M$ such that for $\epsilon > 0$ and $p \in U \subseteq M$ open,

$$\varphi_t\colon\; (-\epsilon, \epsilon) \times W \longrightarrow U$$

for an open subset $W \subseteq U$. By definition of a local flow, ϕ_t is also an integral curve of X:

$$\frac{\partial}{\partial t}\varphi_t(q) = X_{\varphi_t(q)}$$

where $\varphi_0(q) = q$ for all $q \in W$ and $\varphi_s \circ \varphi_t(q) = \varphi_{t+s}(q)$ whenever both sides are defined.

Suppose Y is another smooth vector field on M. In when we define a derivative A for vector valued functions $F\colon \mathbb{R}^n \to \mathbb{R}^m$, we usually require something along the lines of

$$\lim_{\|h\| \to 0} \frac{\|F(x+h) - F(x) - Ah\|}{\|h\|} = 0$$

being satisfied for the unique linear operator A being the derivative of F. Notice how we are in essence comparing the function at different points and taking some limit in order to get the derivative! How do we generalize this to vector fields Y on M? We need to compare Y at $\varphi_t(p)$ and p. We know that for fixed t the map $\varphi_{-t} : \varphi_t(W) \to W$ is a diffeomorphism. Since $Y_{\varphi_t(p)} \in T_{\varphi_t(p)}M$, we may use the differential $(\varphi_{-t})_{*,\varphi_t(p)}\colon T_{\varphi_t(p)}M \to T_pM$ to send $Y_{\varphi_t(p)}$ uniquely to a Y_p since φ_{-t} is a diffeomorphism.

Def. 3.13. (The Lie derivative of a vector field along a vector field)
Let X, Y be smooth vector fields on a smooth manifold M. Let $\varphi_t\colon (-\epsilon, \epsilon) \times W \to$

$U \subseteq M$ be the local flow of X where U is a neighborhood of p. Then the *Lie derivative of* Y *along* X is the vector field $\mathcal{L}_X Y$ given pointwise by

$$(\mathcal{L}_X Y)_p = \lim_{t\to 0} \frac{\varphi_{-t*}(Y_{\varphi_t(p)}) - Y_p}{t} = \lim_{t\to 0} \frac{(\varphi_{-t*}Y)_p - Y_p}{t} = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\varphi_{-t*}Y)_p.$$

Example 3.3. An example is in order to get a feel for what the Lie derivative does. Let $M = \mathbb{R}^3$ and let $X = x^2 \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$ be a smooth vector field on $\mathbb{R}^3$. The flow φ_t of X needs to satisfy the criteria of an integral curve

$$\frac{\partial}{\partial t}\varphi_t(q) = X_{\varphi_t(q)}.$$

Write $\varphi_t(q) = \varphi_t(q_1, q_2, q_3) = (c_1(t), c_2(t), c_3(t))$, and then we obtain

$$\dot{c_1}(t) = (c_1(t))^2, \quad \dot{c_2}(t) = c_2(t), \quad \dot{c_3}(t) = 0.$$

These are all easy to solve ordinary differential equations, and the solutions that satisfy $\varphi_0(q_1, q_2, q_3) = (q_1, q_2, q_3)$ are

$$\varphi_t(q_1, q_2, q_3) = \left(\frac{q_1}{q_1 t + 1}, q_2\, e^t, q_3 \right).$$

The differential $\varphi_t*\colon \mathbb{R}^3 \to \mathbb{R}^3$ is just the Jacobian of this map since we are working over Euclidean spaces:

$$\varphi_{t*} = \begin{pmatrix} \left(\frac{1}{q_1 t+1}\right)^2 & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & 1 \end{pmatrix} \Longrightarrow \varphi_{-t*} = \begin{pmatrix} \left(\frac{1}{-q_1 t+1}\right)^2 & 0 & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Now consider the general smooth vector field Y on $\mathbb{R}^3$ given by

$$Y = f(x,y,z)\,\frac{\partial}{\partial x} + g(x,y,z)\,\frac{\partial}{\partial y} + h(x,y,z)\,\frac{\partial}{\partial z}, \quad f, g, h \in C^\infty(\mathbb{R}^3).$$

We must evaluate this vector field at the point $\varphi_t(q)$ where $q = (q_1, q_2, q_3)$:

$$\begin{aligned} Y_{\varphi_t(q)} = f\left(\frac{q_1}{q_1 t+1}, q_2\, e^t, q_3\right) \left.\frac{\partial}{\partial x}\right|_{\varphi_t(q)} &+ g\left(\frac{q_1}{q_1 t+1}, q_2\, e^t, q_3\right) \left.\frac{\partial}{\partial y}\right|_{\varphi_t(q)} \\ + h\left(\frac{q_1}{q_1 t+1}, q_2\, e^t, q_3\right) \left.\frac{\partial}{\partial z}\right|_{\varphi_t(q)}&. \end{aligned}$$

Then we may compute the push forward of $Y_{\varphi_t(q)}$ by φ_{-t*}

$$\begin{aligned} (\varphi_{-t*}Y)_q = \frac{1}{(-q_1 t+1)^2} \cdot f\left(\frac{q_1}{q_1 t+1}, q_2\, e^t, q_3\right) \left.\frac{\partial}{\partial x}\right|_q \\ + g\left(\frac{q_1}{q_1 t+1}, q_2\, e^t, q_3\right) e^{-t} \left.\frac{\partial}{\partial y}\right|_q + h\left(\frac{q_1}{q_1 t+1}, q_2\, e^t, q_3\right) \left.\frac{\partial}{\partial z}\right|_q. \end{aligned}$$

Now we choose as examples $f(x,y,z) = x$, $g(x,y,z) = y$, $h(x,y,z) = z$, which implies that

$$(\varphi_{-t*}Y)_q = \frac{q_1}{(-q_1 t+1)^2(q_1 t+1)} \left.\frac{\partial}{\partial x}\right|_q + q_2 \left.\frac{\partial}{\partial y}\right|_q + q_3 \left.\frac{\partial}{\partial z}\right|_q$$

and we then differentiate this expression with respect to t and evaluate $t = 0$:

$$(\mathcal{L}_X Y)_q = -q_1^2 \left.\frac{\partial}{\partial x}\right|_q + 0 \left.\frac{\partial}{\partial y}\right|_q + 0 \left.\frac{\partial}{\partial z}\right|_q$$

This means that $\mathcal{L}_X Y = -x^2 \frac{\partial}{\partial x}$. What happens if we choose $Y = X$?

$$(\varphi_{-t*}X)_q = \frac{q_1^2}{(-q_1 t+1)^2(q_1 t+1)^2} \frac{\partial}{\partial x}\bigg|_q + q_2 \frac{\partial}{\partial y}\bigg|_q + q_3 \frac{\partial}{\partial z}\bigg|_q$$

$$\frac{\partial}{\partial t}\bigg|_{t=0} \frac{q_1^2}{(-q_1 t+1)^2(q_1 t+1)^2} = \frac{-4q_1^4 t}{(q_1 t+1)^3(q_1 t-1)^3}\bigg|_{t=0} = 0.$$

Thus $\mathcal{L}_X X = 0$ in this instance, but this makes sense since X does not change relative to the flow φ_t of X. We will see that this holds in general for any smooth vector field X. △

Locally we can write the differential of the flow φ_t of a smooth vector field X as

$$\varphi_{t*,p}\left(\frac{\partial}{\partial x^j}\bigg|_p\right) = \sum_i \frac{\partial \varphi^i}{\partial x^j}(t,p) \frac{\partial}{\partial x^i}\bigg|_p.$$

Assuming that the vector field Y may be written locally as

$$Y = \sum_j b_j \frac{\partial}{\partial x^j}, \text{ then at } \varphi_t(p) : Y_{\varphi_t(p)} = \sum_j b_j(\varphi_t(p)) \frac{\partial}{\partial x^j}\bigg|_{\varphi_t(p)}$$

which we apply to the local expression of the differential of the flow

$$\begin{aligned}\varphi_{-t*,p}\left(Y_{\varphi_t(p)}\right) &= \sum_j b_j(\varphi_t(p))\varphi_{-t*,p}\left(\frac{\partial}{\partial x^j}\bigg|_{\varphi_t(p)}\right) \\ &= \sum_j \sum_i b_j(\varphi_t(p)) \frac{\partial \varphi^i}{\partial x^j}(-t,p) \frac{\partial}{\partial x^i}\bigg|_p.\end{aligned}$$

By the definition of the Lie derivative of Y along X, we obtain

$$\begin{aligned}(\mathcal{L}_X Y)_p &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \varphi_{-t*,p}\left(Y_{\varphi_t(p)}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \sum_j \sum_i b_j(\varphi_t(p)) \frac{\partial \varphi^i}{\partial x^j}(-t,p) \frac{\partial}{\partial x^i}\bigg|_p \\ &= \sum_{i,j,k}\left(\frac{\partial b_j}{\partial x^k}(\varphi_t(p)) \frac{\partial \varphi_t^k}{\partial t}(p) \frac{\partial \varphi_{-t}^i}{\partial x^j}(p)\right) \frac{\partial}{\partial x^i}\bigg|_p\bigg|_{t=0} \\ &\quad - \sum_{i,j}\left(b_j(\varphi_t(p)) \frac{\partial}{\partial x^j} \frac{\partial \varphi_{-t}^i}{\partial t}(p) \frac{\partial}{\partial x^i}\bigg|_p\right)\bigg|_{t=0} \\ &= \sum_{i,j,k}\left(\frac{\partial b_j}{\partial x^k}(p) a_k(p) \frac{\partial \varphi_0^i}{\partial x^j}(p)\right) \frac{\partial}{\partial x^i}\bigg|_p - \sum_{i,j}\left(b_j(p) \frac{\partial a_i}{\partial x^j}(p)\right) \frac{\partial}{\partial x^i}\bigg|_p\end{aligned}$$

by the product rule and the fact that the flow is especially an integral curve of X and the fact that $\varphi_0(p) = p$ which implies that φ_{0*} is the identity transformation on T_pM. In other words:

$$\frac{\partial \varphi_0^i}{\partial x^j}(p) = \delta_j^i$$

Relabelling the indices to match the indices of the summands, we get

$$(\mathcal{L}_X Y)_p = \sum_{i,j}\left(a_j(p) \frac{\partial b_i}{\partial x^j}(p) - b_j(p) \frac{\partial a_i}{\partial x^j}(p)\right) \frac{\partial}{\partial x^i}\bigg|_p.$$

We have then given the local proof of the following theorem:

Theorem 3.2. Let X, Y be smooth vector fields on a smooth manifold M. Then the Lie derivative of Y along X is the Lie bracket of X and Y:

$$\mathcal{L}_X Y = [X, Y].$$

How do we define a Lie derivative for differential forms on M? The core principle is the same as that for vector fields.

Def. 3.14. (The Lie derivative of a differential form)
Let X be a smooth vector field on a smooth manifold M. Given a differential form $\omega \in \Omega^*(M)$, we define the *Lie derivative of* ω *along* X as the differential form given pointwise by

$$(\mathcal{L}_X \omega)_p = \lim_{t \to 0} \frac{\varphi_t^*(\omega_{\varphi_t(p)}) - \omega_p}{t} = \lim_{t \to 0} \frac{(\varphi_t^* \omega)_p - \omega_p}{t} = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} (\varphi_t^* \omega)_p.$$

where φ_t is the local flow of X and $p \in M$.

Proposition 3.13. If $f \in C^\infty(M)$ and X is a smooth vector field, then

$$\mathcal{L}_X f = Xf.$$

Proof. The definition of the Lie derivative of a differential form is given pointwise as

$$\begin{aligned} (\mathcal{L}_X f)_p &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} (\varphi_t^* f)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} (f \circ \varphi_t)(p) \\ &= \varphi_{t*}(p) \left(\frac{\partial}{\partial t}\bigg|_{t=0} \right) f = \frac{\mathrm{d}}{\mathrm{d}t} \varphi_t(p) \bigg|_{t=0} f = X_p f, \end{aligned}$$

since $\varphi_t(p) \colon (-\epsilon, \epsilon) \to M$ is a curve for fixed $p \in M$. Generalizing for all $p \in M$ establishes the equality. □

We now turn our attention to a map which lowers the degree of differential forms, in contrast of the exterior derivative. This operation is called the interior multiplication.

Def. 3.15. (Interior multiplication)
Let M be a smooth manifold, $X, X_2, \ldots, X_k$ smooth vector fields on M, and $\omega \in \Omega^k(M)$, then the *interior multiplication* of ω with X is $\iota_X \omega \in \Omega^{k-1}(M)$ defined by

$$\iota_X \omega(X_2, \ldots, X_k) = \omega(X, X_2, \ldots, X_k).$$

If $\omega \in \Omega^0(M)$, then we define $\iota_X \omega = 0$.
If $\omega \in \Omega^1(M)$, then we define $\iota_X \omega = \omega(X)$.

Remark. Some authors use different notation for interior multiplication, namely $X \lrcorner\, \omega = \iota_X \omega$. It is alternatively called a *contraction* of ω by the vector field X.

Next we state and prove some properties tied to interior multiplication that will prove useful later on.

Proposition 3.14. Let $\mathrm{d}f^1, \ldots, \mathrm{d}f^k$ be 1-forms on a smooth manifold M and let X be a smooth vector field on M. The interior multiplication of the wedge product $\mathrm{d}f^1 \wedge \cdots \wedge \mathrm{d}f^k$ with X is given by

$$\iota_X(\mathrm{d}f^1 \wedge \cdots \wedge \mathrm{d}f^k) = \sum_{i=1}^k (-1)^{i+1} \mathrm{d}f^i(X)\, \mathrm{d}f^1 \wedge \cdots \wedge \widehat{\mathrm{d}f^i} \wedge \cdots \wedge \mathrm{d}f^k$$

where the caret $\widehat{\ }$ means omitting that particular factor in the product.

Proof. Using the definition of the interior multiplication on $X_2, \dots, X_k$. For simplicity we denote $X_1 = X$.

$$\begin{aligned}
&\iota_X(\mathrm{d}f^1 \wedge \dots \wedge \mathrm{d}f^k)(X_2, \dots, X_k) = (\mathrm{d}f^1 \wedge \dots \wedge \mathrm{d}f^k)(X_1, X_2, \dots, X_k) \\
&= \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma) \mathrm{d}f^1(X_{\sigma(1)}) \cdots \mathrm{d}f^k(X_{\sigma(k)}) = \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma) \prod_{j=1}^{k} \mathrm{d}f^j(X_{\sigma(j)}) \\
&= \det([\mathrm{d}f^l(X_j)]_{l,j}
\end{aligned}$$

Expanding along the first column of this determinant gives us the desired formula, and we leave this as an exercise. □

Proposition 3.15. Let X be a smooth vector field on a smooth manifold M. Then

(i) $\iota_X \circ \iota_X = 0$,

(ii) For $\omega \in \Omega^k(M)$ and $\tau \in \Omega^l(M)$,

$$\iota_X(\omega \wedge \tau) = (\iota_X \omega) \wedge \tau + (-1)^k \omega \wedge (\iota_X \tau).$$

Proof. (i) Let $\omega \in \Omega^k(M)$ and consider smooth vector fields $X_3, \dots, X_k$:

$$\iota_X(\iota_X \omega)(X_3, \dots, X_k) = \iota_X \omega(X, X_3, \dots, X_k) = \omega(X, X, X_3, \dots, X_k) = 0.$$

(ii) Consider a coordinate chart $(U, x^1, \dots, x^n)$ on M. We express the differential forms locally by

$$\omega = \sum_I a_I \, \mathrm{d}x^I, \quad \tau = \sum_J b_J \, \mathrm{d}x^J.$$

Using these local expressions we may study a special case, namely for

$$\iota_X(\omega \wedge \tau) = \iota_X \left(\sum_I \sum_J a_I \cdot b_J \, \mathrm{d}x^I \wedge \mathrm{d}x^J \right) = \sum_I \sum_J a_I \cdot b_J \, \iota_X(\mathrm{d}x^I \wedge \mathrm{d}x^J),$$

then we instead study $\iota_X(\mathrm{d}x^{i_1} \wedge \dots \wedge \mathrm{d}x^{i_{k+l}})$. By Proposition 3.14 we have that this interior multiplication can be expanded as

$$\begin{aligned}
&\iota_X(\mathrm{d}x^{i_1} \wedge \dots \wedge \mathrm{d}x^{i_{k+l}}) = \sum_{j=1}^{k+l} (-1)^{j+1} \mathrm{d}x^{i_j}(X) \, \mathrm{d}x^{i_1} \wedge \dots \wedge \widehat{\mathrm{d}x^{i_j}} \wedge \dots \wedge \mathrm{d}x^{i_{k+l}} \\
&= \left(\sum_{j=1}^{k} (-1)^{j+1} \mathrm{d}x^{i_j}(X) \, \mathrm{d}x^{i_1} \wedge \dots \wedge \widehat{\mathrm{d}x^{i_j}} \wedge \dots \wedge \mathrm{d}x^{i_k} \right) \wedge \mathrm{d}x^{i_{k+1}} \dots \wedge \mathrm{d}x^{i_{k+l}} \\
&+ (-1)^k \mathrm{d}x^{i_1} \wedge \dots \wedge \mathrm{d}x^{i_k} \wedge \left(\sum_{j=1}^{l} (-1)^{j+1} \mathrm{d}x^{i_{k+j}}(X) \, \mathrm{d}x^{i_{k+1}} \wedge \dots \wedge \widehat{\mathrm{d}x^{i_{k+j}}} \wedge \dots \wedge \mathrm{d}x^{i_{k+l}} \right) \\
&= \iota_X(\mathrm{d}x^{i_1} \wedge \dots \wedge \mathrm{d}x^{i_k}) \wedge (\mathrm{d}x^{i_{k+1}}) \dots \wedge (\mathrm{d}x^{i_{k+l}}) \\
&+ (-1)^k (\mathrm{d}x^{i_1} \wedge \dots \wedge \mathrm{d}x^{i_k}) \wedge \iota_X(\mathrm{d}x^{i_{k+1}}) \dots \wedge (\mathrm{d}x^{i_{k+l}})
\end{aligned}$$

Linearity of ι_X immediately gives the general result as desired. □

Our next result will be very important in our further discourse, and the propositions thesmselves are also quite interesting in their own right.

Theorem 3.3. (Properties of the Lie derivative)
Let X be a smooth vector field, $\omega, \tau \in \Omega^*(M)$ be differential forms. Then the following properties hold

(i) $\mathcal{L}_X(\omega \wedge \tau) = (\mathcal{L}_X\omega) \wedge \tau + \omega \wedge (\mathcal{L}_X\tau)$,
(ii) $\mathcal{L}_X \mathrm{d}\omega = \mathrm{d}\mathcal{L}_X\omega$,
(iii) *The Cartan Homotopy Formula:* $\mathcal{L}_X = \mathrm{d}\iota_X + \iota_X\mathrm{d}$.

Proof. (i) Let φ_t be the flow of a smooth vector field X on M. Differentiating $(\varphi_t^*(\omega \wedge \tau))_p$ with respect to the parameter t amounts to

$$\frac{\mathrm{d}}{\mathrm{d}t}(\varphi_t^*(\omega \wedge \tau))_p = \frac{\mathrm{d}}{\mathrm{d}t}(\varphi_t^*\omega \wedge \varphi_t^*\tau)_p = \left(\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^*\omega\right)_p \wedge (\varphi_t^*\tau)_p + (\varphi_t^*\omega)_p \wedge \left(\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^*\tau\right)_p.$$

Evaluating the parameter $t = 0$ on both sides of the preceding equation then leads to

$$(\mathcal{L}_X(\omega \wedge \tau))_p = (\mathcal{L}_X\omega)_p \wedge \tau_p + \omega_p \wedge (\mathcal{L}_X\tau)_p.$$

(ii) Consider a coordinate chart $(U, x^1, \dots, x^n)$ on M. Then we write $\omega \in \Omega^*(M)$ as

$$\omega = \sum_I a_I \,\mathrm{d}x^I, \quad \mathrm{d}\omega = \sum_I \mathrm{d}a_I \wedge \mathrm{d}x^I.$$

Using this expression for the Lie derivative amounts to

$$(\mathcal{L}_X\mathrm{d}\omega)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(\varphi_t^*\mathrm{d}\omega)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(\mathrm{d}\varphi_t^*\omega)_p = \mathrm{d}\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(\varphi_t^*\omega)_p = (\mathrm{d}\mathcal{L}_X\omega)_p$$

where we have used Proposition 3.10 and 3.12.
(iii) Let $\omega \in \Omega^k(M)$ be a differential k-form. The Lie derivative is given by

$$(\mathcal{L}_X\omega)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(\varphi_t^*\omega)_p.$$

We first prove the case $k = 0$. Consider $f \in \Omega^0(M)$, then:

$$(\mathcal{L}_X f)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(\varphi_t^* f)_p = X_p f.$$

Based on the definition of interior multiplication on 0-forms and 1-forms:

$$\mathrm{d}\iota_X f = \mathrm{d}(0) = 0, \quad \iota_X \mathrm{d}f = \mathrm{d}f(X).$$

Note that $(\iota_X\mathrm{d}f)_p = (\mathrm{d}f_p)(X_p) = X_p f$ by definition of $\mathrm{d}f$ pointwise. This proves the case for $k = 0$.

Consider now the case $k = 1$. Let $\omega \in \Omega^1(M)$, then for a coordinate chart $(U, x^1, \dots, x^n)$ on M, we have

$$\omega = \sum_i a_i \,\mathrm{d}x^i = \sum_i a_i \wedge \mathrm{d}x^i, \quad a_i \in C^\infty(M).$$

The second equality is true since a_i are equivalently 0-forms on M. Using this we get

$$\begin{aligned}(\mathcal{L}_X\omega)_p &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(\varphi_t^*\omega)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\left(\varphi_t^*\left(\sum_i a_i \,\mathrm{d}x^i\right)\right)_p \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\left(\varphi_t^*\left(\sum_i a_i \wedge \mathrm{d}x^i\right)\right)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\left(\sum_i \varphi_t^* a_i \wedge \varphi_t^*\mathrm{d}x^i\right)_p \\ &= \sum_i \left((\mathcal{L}_X a_i \wedge \mathrm{d}x^i)_p + (a_i \wedge \mathcal{L}_X\mathrm{d}x^i)_p\right),\end{aligned}$$

where the last equality is due to the product rule of differentiation by t.
By property (ii) above and the fact that $\mathrm{d} \circ \mathrm{d} = 0$ we have

$$(\mathcal{L}_X \mathrm{d}x^i)_p = (\mathrm{d}\mathcal{L}_X x^i)_p = \mathrm{d}(X_p x^i) = \mathrm{d}(\mathrm{d}x^i(X))_p = 0.$$

Given that $(\mathcal{L}_X a_i)_p = X_p a_i = (\mathrm{d}a_i(X))_p$ we have that

$$(\mathcal{L}_X a_i \wedge \mathrm{d}x^i)_p = \mathrm{d}a_{i,p}(X_p)\, \mathrm{d}x^i_p.$$

$$\mathrm{d}\iota_X\omega = \mathrm{d}\left(\sum_i a_i\, \mathrm{d}x^i(X)\right) = \sum_i \mathrm{d}a_i \wedge \mathrm{d}x^i(X)$$

$$\iota_X \mathrm{d}\omega = \iota_X\left(\sum_i \mathrm{d}a_i \wedge \mathrm{d}x^i\right) = \left(\sum_i \mathrm{d}a_i(X) \wedge \mathrm{d}x^i - \mathrm{d}a_i \wedge \mathrm{d}x^i(X)\right)$$

from Proposition 3.15. Clearly we then have

$$\iota_X \mathrm{d}\omega + \mathrm{d}\iota_X\omega = \sum_i \mathrm{d}a_i(X) \wedge \mathrm{d}x^i.$$

This proves the case $k = 1$.

For the case $k \geq 1$ we consider the following with $\omega \in \Omega^k(M)$:

$$\begin{aligned}
(\mathcal{L}_X\omega)_p &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} (\varphi_t^*\omega)_p = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \left(\varphi_t^*\left(\sum_I a_I\, \mathrm{d}x^I\right)\right)_p \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \left(\sum_I \varphi_t^*\left(a_I\, \mathrm{d}x^I\right)\right)_p = \sum_I \left(\mathcal{L}_X a_I \wedge \mathrm{d}x^I + a_I \wedge \mathcal{L}_X \mathrm{d}x^I\right)_p \\
&= \sum_I \left(\mathrm{d}a_I(X) \wedge \mathrm{d}x^I + a_I \wedge \mathcal{L}_X \mathrm{d}x^I\right)_p.
\end{aligned}$$

From property (i) which states $\mathcal{L}_X(\mathrm{d}x^i \wedge \mathrm{d}x^j) = (\mathcal{L}_X \mathrm{d}x^i) \wedge \mathrm{d}x^j + \mathrm{d}x^i \wedge (\mathcal{L}_X \mathrm{d}x^j)$. Given $I = (i_1, \dots, i_k)$, then we calculate

$$\mathcal{L}_X \mathrm{d}x^I = \sum_{j=1}^k \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathcal{L}_X \mathrm{d}x^{i_j} \wedge \cdots \wedge \mathrm{d}x^{i_k}.$$

From the case $k = 1$ we have, because of $\mathrm{d} \circ \mathrm{d} = 0$:

$$\mathcal{L}_X \mathrm{d}x^{i_j} = \mathrm{d}\iota_X \mathrm{d}x^{i_j} + \iota_X \mathrm{d}(\mathrm{d}x^{i_j}) = \mathrm{d}(\mathrm{d}x^{i_j}(X)) + 0 = 0.$$

Thus we have $\mathcal{L}_X \mathrm{d}x^I = 0$. Like in the case for $k = 1$ we consider now the sum $\iota_X \mathrm{d}\omega + \mathrm{d}\iota_X\omega$:

$$\iota_X \mathrm{d}\omega = \iota_X\left(\sum_I \mathrm{d}a_I \wedge \mathrm{d}x^I\right) = \sum_I \left(\mathrm{d}a_I(X) \wedge \mathrm{d}x^I - \mathrm{d}a_I \wedge (\iota_X \mathrm{d}x^I)\right)$$

$$\mathrm{d}\iota_X\omega = \mathrm{d}\left(\sum_I a_I\, \iota_X \mathrm{d}x^I\right) = \sum_I \left(\mathrm{d}a_I \wedge \iota_X \mathrm{d}x^I + a_I \wedge \mathrm{d}\iota_X \mathrm{d}x^I\right)$$

Calculating $\mathrm{d}\iota_X \mathrm{d}x^I$ by using Proposition 3.14 gives

$$\mathrm{d}\iota_X \mathrm{d}x^I = \mathrm{d}\left(\sum_{j=1}^k (-1)^{j+1} \mathrm{d}x^{i_j}(X) \mathrm{d}x^{i_1} \wedge \cdots \wedge \widehat{\mathrm{d}x^{i_j}} \wedge \cdots \wedge \mathrm{d}x^{i_k}\right) = 0,$$

which then finally gives us the desired result

$$\iota_X \mathrm{d}\omega + \mathrm{d}\iota_X \omega = \sum_I \mathrm{d}a_I(X) \wedge \mathrm{d}x^I = \mathcal{L}_X \omega.$$

□

4. Symplectic Manifolds

In this section we tackle the subject of symplectic manifolds. The main goal of this section will be to expand on the informative article by Henry Cohn "Why symplectic geometry is the natural setting for classical mechanics" [4]. So far we have established the necessary machinery to understand this.

Def. 4.1. (Symplectic vector spaces)
Let V be a finite-dimensional vector space over the field $\mathbb{R}$. A *symplectic form on V* is a map $\omega \in \mathcal{A}^2(V)$ that is non-degenerate, i.e. if $\omega(u, w) = 0$ for all $w \in V$ then necessarily $u = 0$. The pair (V, ω) is called a *symplectic vector space.*

From our treatment of multilinear algebra, we may see vector spaces V over $\mathbb{R}$ as being diffeomorphic to some tangent space based at a point p since an n-dimensional vector space over $\mathbb{R}$ satisfies

$$V \cong \mathbb{R}^n = T_p\mathbb{R}^n.$$

Def. 4.2. (Symplectic manifolds)
Let M be a smooth manifold. We call $\omega \in \Omega^2(M)$ a *symplectic form* on M if it is both non-degenerate and closed, i.e. $\mathrm{d}\omega = 0$. A smooth manifold M endowed with a symplectic form is called a *symplectic manifold* and is often denoted by the pair (M, ω).

Remark. Notice how symplectic vector spaces immediately specialize from this definition, apart from the condition that ω be closed! If $\omega \in \Omega^2(\mathbb{R}^n)$, then we may express the form globally as

$$\omega = \sum_{i<j} a_{ij}\, \mathrm{d}x^i \wedge \mathrm{d}x^j$$

which then implies that

$$\mathrm{d}\omega = \sum_{i<j} \mathrm{d}a_{ij} \wedge \mathrm{d}x^i \wedge \mathrm{d}x^j.$$

Does this need to vanish in general? Judging by this equation alone one would not expect this without further inspection.

Def. 4.3. Let (M, ω) be a symplectic manifold. Any smooth function $H : M \to \mathbb{R}$ is called a *Hamiltonian function* or simply a *Hamiltonian.* Associated to this Hamiltonian is a *symplectic gradient* or *symplectic vector field*, which is a smooth vector field X_H on M satisfying

$$\mathrm{d}H = \iota_{X_H}\omega.$$

Proposition 4.1. If (M, ω) is a symplectic manifold, then $\dim M$ is even.

Proof. Since M is endowed with a symplectic form ω, we may express this form locally in terms of a coordinate chart $(U, x^1, \dots, x^n)$. The open set $U \subseteq M$ is diffeomorphic to some open $V \subseteq \mathbb{R}^m$ where $m = \dim M$. Let this diffeomorphism be $\phi\colon V \to U$, then the 2-form on V is given by $\omega_V = \phi^*\omega$. On V we may express ω_V as

$$\omega_V = \sum_{i<j} a_{ij}\, \alpha^i \wedge \alpha^j.$$

In fact, we may write ω_V as a matrix acting on column vectors $u, w \in \mathbb{R}^m$:

$$\omega_V(u, w) = u^T\, [\omega_V]\, w$$

where $[\omega_V]$ is the matrix representation of ω_V and T denotes the transpose. Since ω_V is a symplectic form on $V \subseteq \mathbb{R}^m$, we have that $[\omega_V]$ is skew-symmetric:

$$[\omega_V]^T = -[\omega_V].$$

Using the properties of the determinant we have

$$\det([\omega_V]) = \det([\omega_V]^T) = \det(-[\omega_V]) = (-1)^m \det([\omega_V]),$$

hence $m = \dim M$ must be even. □

Example 4.1. Let $M = \mathbb{R}^{2n}$ and ω be the canonical symplectic form given by

$$\omega = \sum_{i<j} \mathrm{d}x^i \wedge \mathrm{d}x^j.$$

Consider the Hamiltonian given by $H(x^1, \dots, x^{2n}) = x^1 + \cdots + x^{2n}$, then

$$\mathrm{d}H = \mathrm{d}x^1 + \cdots + \mathrm{d}x^{2n} = \iota_{X_H}\omega = \sum_{i<j} (X_i \mathrm{d}x^j - X_j \mathrm{d}x^i).$$

As for our example we may let $n = 2$, then

$$\begin{aligned}\sum_{i=1}^{4} \mathrm{d}x^i &= X_1\mathrm{d}x^4 - X_4\mathrm{d}x^1 + X_1\mathrm{d}x^3 - X_3\mathrm{d}x^1 + X_1\mathrm{d}x^2 - X_2\mathrm{d}x^1 \\ &\quad + X_2\mathrm{d}x^4 - X_4\mathrm{d}x^2 + X_2\mathrm{d}x^3 - X_3\mathrm{d}x^2 + X_3\mathrm{d}x^4 - X_4\mathrm{d}x^3 \\ &= (-X_2 - X_3 - X_4)\mathrm{d}x^1 + (X_1 - X_3 - X_4)\mathrm{d}x^2 \\ &\quad + (X_1 + X_2 - X_4)\mathrm{d}x^3 + (X_1 + X_2 + X_3)\mathrm{d}x^4\end{aligned}$$

Solving this linear system gives $X_1 = -1, X_2 = 1, X_3 = 1, X_4 = -3$. Thus the symplectic vector field associated to H is given by

$$X_H = -\frac{\partial}{\partial x^1} + \frac{\partial}{\partial x^2} + \frac{\partial}{\partial x^3} - 3\frac{\partial}{\partial x^4}.$$

△

How are the laws of physics expressed on manifolds? It turns out that our formalism gives rise to easy to interpret dynamics. In classical mechanics, the Hamiltonian may represent the total energy of the physical system. We neglect the phenomenological aspects of the physics in this article, although the mathematical aspects of classical mechanics is quite interesting. More details on these topics can be found in the seminal text "Mathematical Methods of Classical Mechanics" by V. I. Arnold.[1]

Dynamics on symplectic manifolds is given by flows of symplectic gradient fields. Conventionally one separates the $2n$ coordinates of a coordinate chart on M as the *canonical coordinates* $(q^1, \dots, q^n, p^1, \dots, p^n)$. In physics one usually calls $q^1, \dots, q^n$ the generalized coordinates and $p^1, \dots, p^n$ the canonical momenta.

In physics one calls a system closed if the physical energy of the system is conserved - the total energy of the system is constant in time. When a system is closed the Hamiltonian is the kinetic energy plus the potential energy of that system. In general this looks like

$$H(q^1,\dots,q^n,p^1,\dots,p^n)=\sum_{i=1}^{n}\frac{(p^i)^2}{2m}+V(q^1,\dots,q^n)$$

where m is the mass of the particle which the system describes, and V is the potential function. Then we calculate

$$\mathrm{d}H=\sum_{i=1}^{n}\frac{p^i}{m}\,\mathrm{d}p^i+\sum_{i=1}^{n}\frac{\partial V}{\partial q^i}\,\mathrm{d}q^i$$

A symplectic form on M can be written as

$$\omega=\sum_{i=1}^{n}\mathrm{d}q^i\wedge\mathrm{d}p^i,$$

then the relation between the Hamiltonian and the symplectic form gives

$$\mathrm{d}H=\iota_{X_H}\omega=\sum_{i=1}^{n}(\mathrm{d}q^i(X_H)\mathrm{d}p^i-\mathrm{d}p^i(X_H)\mathrm{d}q^i)=\sum_{i=1}^{n}\left(\frac{p^i}{m}\,\mathrm{d}p^i+\frac{\partial V}{\partial q^i}\,\mathrm{d}q^i\right).$$

Comparing the coefficients of the 1-forms:

$$\mathrm{d}q^i(X_H)=\frac{p^i}{m},\quad \mathrm{d}p^i(X_H)=-\frac{\partial V}{\partial q^i}.$$

This means that the symplectic gradient has the form

$$X_H=\frac{p^1}{m}\frac{\partial}{\partial q^1}+\cdots+\frac{p^n}{m}\frac{\partial}{\partial q^n}-\frac{\partial V}{\partial q^1}\frac{\partial}{\partial p^1}-\cdots-\frac{\partial V}{\partial q^n}\frac{\partial}{\partial p^n}.$$

The flow φ_t of this symplectic gradient is

$$\frac{\partial}{\partial t}\varphi_t=(X_H)_{\varphi_t}.$$

Example 4.2. Let the potential V be given by

$$V(q^1,\dots,q^n)=\sum_{i=1}^{n}\frac{1}{2}k(q^i)^2,\quad k\in\mathbb{R}^+,$$

then the symplectic gradient is

$$X_H=\frac{p^1}{m}\frac{\partial}{\partial q^1}+\cdots+\frac{p^n}{m}\frac{\partial}{\partial q^n}-kq^1\frac{\partial}{\partial p^1}-\cdots-kq^n\frac{\partial}{\partial p^n}.$$

Let φ_t be the flow of the symplectic gradient. Coordinate-wise we denote the time-derivative of the flow by

$$\frac{\partial}{\partial t}\varphi_t=\dot{\phi}_1\left.\frac{\partial}{\partial q^1}\right|_{\varphi_t}+\cdots+\dot{\phi}_n\left.\frac{\partial}{\partial q^n}\right|_{\varphi_t}+\dot{\phi}_{n+1}\left.\frac{\partial}{\partial p^1}\right|_{\varphi_t}+\cdots+\dot{\phi}_{2n}\left.\frac{\partial}{\partial p^n}\right|_{\varphi_t}.$$

$$(X_H)_{\varphi_t}=\frac{\phi_{n+1}}{m}\left.\frac{\partial}{\partial q^1}\right|_{\varphi_t}+\cdots+\frac{\phi_{2n}}{m}\left.\frac{\partial}{\partial q^n}\right|_{\varphi_t}-k\phi_1\left.\frac{\partial}{\partial p^1}\right|_{\varphi_t}-\cdots-k\phi_n\left.\frac{\partial}{\partial p^n}\right|_{\varphi_t}$$

Comparing the coefficients we have

$$\dot{\phi}_1=\frac{\phi_{n+1}}{m},\dots,\dot{\phi}_n=\frac{\phi_{2n}}{m},$$

$$\dot{\phi}_{n+1} = -k\phi_1, \ldots, \dot{\phi}_{2n} = -k\phi_n.$$

Differentiating the first set equations with respect to t gives

$$\ddot{\phi}_1 = -\frac{k}{m}\phi_1, \ldots, \ddot{\phi}_n = -\frac{k}{m}\phi_n$$

which we may solve, for $\varphi_0(x_1, \ldots, x_n, y_1, \ldots, y_n) = (x_1, \ldots, x_n, y_1, \ldots, y_n)$:

$$\phi_i(t) = \frac{y_i}{\sqrt{km}} \sin\sqrt{\frac{k}{m}}t + x_i \cos\sqrt{\frac{k}{m}}t, \quad 1 \leq i \leq n,$$

$$\phi_j(t) = y_j \cos\sqrt{\frac{k}{m}}t - x_j\sqrt{km}\sin\sqrt{\frac{k}{m}}t, \quad n+1 \leq j \leq 2n.$$

Thus the flow of this particular symplectic gradient on M gives rise to an oscillating system $\varphi_t = (\phi_1, \ldots, \phi_n, \phi_{n+1}, \ldots, \phi_{2n})$ locally around φ_0. $\triangle$

What if we wanted to study mechanics and dynamics on a manifold of odd dimension such as the 1-dimensional circle $S^1 \subseteq \mathbb{R}^2$? Then we may for instance study the cotangent bundle T^*S^1. The reason why this works is because for every smooth manifold M, the cotangent bundle T^*M is a smooth manifold of dimension $2\dim M$. Because of this one may endow the cotangent bundle with a symplectic structure and consequently interpret points $(q, \omega) \in T^*M$ as generalized coordinates $q = (q^1, \ldots, q^n)$ and generalized momenta $\omega = (\mathrm{d}p^1, \ldots, \mathrm{d}p^n)$. We will not dive further into this construction, and leave the interested reader to study more in Arnold's text.[1]

Now we turn to explaining the requirements we place on symplectic forms. First of all we require that a symplectic form $\omega \in \Omega^2(M)$ is non-degenerate. Consider $\omega \in \Omega^2(M)$ and X, Y smooth vector fields on M. We may represent ω locally on $U \subseteq M$ as

$$\omega = \sum_{i=1}^{n} a_i \, \mathrm{d}q^i \wedge \mathrm{d}p^i, \quad a_i \in C^\infty(U).$$

For any 1-form $\eta \in \Omega^1(M)$ we wish to solve for the vector field X in the expression

$$\eta(Y) = \omega(X, Y)$$

for any vector field Y. Recall that we may express $\omega(X, Y)$ locally as a matrix $[\omega]$ multiplied on vectors X, Y:

$$\omega(X, Y) = X^T[\omega]Y.$$

We write the 1-form η component-wise as

$$\eta = \sum_{i=1}^{n} \eta_i \, \mathrm{d}q^i + \sum_{j=1}^{n} \eta_j \, \mathrm{d}p^j$$

and the vector fields X and Y as

$$X = \sum_{i=1}^{n} X_i \frac{\partial}{\partial q^i} + \sum_{j=1}^{n} X_{n+j} \frac{\partial}{\partial p^j}, \quad Y = \sum_{i=1}^{n} Y_i \frac{\partial}{\partial q^i} + \sum_{j=1}^{n} Y_{n+j} \frac{\partial}{\partial p^j}.$$

Calculating the action of ω on X and Y gives

$$\begin{aligned}\omega(X, Y) &= \sum_{i=1}^{n} a_i(\mathrm{d}q^i(X)\mathrm{d}p^i(Y) - \mathrm{d}q^i(Y)\mathrm{d}p^i(X)) \\ &= a_1X_1Y_{n+1} + \cdots + a_nX_nY_{2n} - a_1X_{n+1}Y_1 - \cdots - a_nX_{2n}Y_n.\end{aligned}$$

Representing X and Y as column vectors $X = (X_1, \dots, X_{2n})^T$ and $Y = (Y_1, \dots, Y_{2n})^T$ then enables us to write ω as the $2n \times 2n$-matrix $[\omega]$ given as

$$[\omega] = \begin{pmatrix} \mathbf{Z} & \mathbf{a} \\ -\mathbf{a} & \mathbf{Z} \end{pmatrix}$$

where $\mathbf{Z}$ is the $n \times n$ matrix of only zeroes and $\mathbf{a}$ is the diagonal matrix $\mathbf{a} = \mathrm{diag}(a_1, \dots, a_n)$. Representing the 1-form η as $\eta = (\eta_1, \dots, \eta_{2n})^T$ we obtain the linear system relation

$$\eta^T Y = X^T[\omega]Y.$$

This should hold for any vector field Y, thus

$$(\eta^T - X^T[\omega])\, Y = 0 \ \text{ implies that } \ \eta^T = X^T[\omega].$$

Using the skew-symmetry of the matrix representation, we have

$$X = -[\omega]^{-1}\, \eta,$$

since we wish to solve for the vector field X. The inverse of this matrix exists if and only if $\det([\omega]) \neq 0$, and due to linear algebra results this is equivalent to our condition of non-degeneracy

$$X^T[\omega]Y = 0 \text{ for all } Y, \text{ then } X = 0.$$

Our next condition for symplectic forms states that all symplectic forms should be closed. If $\omega \in \Omega^2(M)$ is a symplectic form on a smooth manifold M, then we have for a Hamiltonian $H \colon M \to \mathbb{R}$ a symplectic gradient X_H satsifying

$$\mathrm{d}H = \iota_{X_H}\omega.$$

Denote the flow of the symplectic vector field X_H by φ_t. The flow has to obey certain conditions in regards to the dynamics of the system. Consider the pullback $\varphi_t^*\omega$, and then the relation

$$\mathrm{d}H = \iota_{Y_H}(\varphi_t^*\omega).$$

But here we are comparing 2-forms ω and $\varphi_t^*\omega$ along the curve φ_t on M, so given that φ_t is an integral curve of X_H, we should have that the symplectic gradient Y_H is the same as X_H. This means that we are actually requiring $\varphi_t^*\omega = \omega$ for all t. Taking the derivative of $\varphi_t^*\omega$ at $t = t_0$ is given by

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(\varphi_t^*\omega)_p(t_0) &= \lim_{t \to t_0} \frac{(\varphi_t^*\omega)_p - (\varphi_{t_0}^*\omega)_p}{t - t_0} \\
&= \lim_{t \to t_0} \frac{\varphi_{t_0}^*(\varphi_{t-t_0}^*\omega)_{\varphi_{t_0}(p)} - \varphi_{t_0}^*(\omega_{\varphi_{t_0}(p)})}{t - t_0} \\
&= \varphi_{t_0}^*\left(\lim_{t-t_0 \to 0} \frac{(\varphi_{t-t_0}^*\omega)_{\varphi_{t_0}(p)} - \omega_{\varphi_{t_0}(p)}}{t - t_0}\right) \\
&= \varphi_{t_0}^*(\mathcal{L}_{X_H}\omega)_{\varphi_{t_0}(p)} = (\varphi_{t_0}^*\mathcal{L}_{X_H}\omega)_p.
\end{aligned}$$

For general t we then have that the derivative of $\varphi_t^*\omega$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^*\omega = \varphi_t^*\mathcal{L}_{X_H}\omega.$$

Using Cartan's homotopy formula on ω we have

$$\mathcal{L}_{X_H}\omega = \mathrm{d}\iota_{X_H}\omega + \iota_{X_H}\mathrm{d}\omega,$$

but we know that $\mathrm{d}H = \iota_{X_H}\omega$, thus $\mathrm{d}\iota_{X_H}\omega = \mathrm{d}(\mathrm{d}H) = 0$. This implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^*\omega = 0 \text{ if and only if } \mathrm{d}\omega = 0.$$

Our requirement of consistent symplectic gradients on the flow of X_H gives us that ω needs to be closed.

Def. 4.4. Let $H, E\colon M \to \mathbb{R}$ be Hamiltonians on M, then the *Poisson bracket* of H and E is given locally by

$$\{H, E\} = \sum_{i=1}^{n}\left(\frac{\partial H}{\partial q^i}\frac{\partial E}{\partial p^i} - \frac{\partial H}{\partial p^i}\frac{\partial E}{\partial q^i}\right),$$

where $(U, q^1, \dots, q^n, p^1, \dots, p^n)$ is a coordinate chart on M.

Theorem 4.1. (Noether's theorem)
Let (M, ω) be a symplectic manifold with the canonical symplectic form ω given by

$$\omega = \sum_{i=1}^{n} \mathrm{d}q^i \wedge \mathrm{d}p^i.$$

Then

$$\{H, E\} = 0 \text{ if and only if } \frac{\mathrm{d}}{\mathrm{d}t}E(\varphi_t) = 0$$

where φ_t is the flow of the symplectic gradient associated to H.

Proof. Assuming $(U, q^1, \dots, q^n, p^1, \dots, p^n)$ is a coordinate chart on M we may calculate

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\varphi_t) = E_{*,\varphi_t}(\varphi_{t*}) = \mathrm{d}E_{\varphi_t}((X_H)_{\varphi_t}).$$

In a previous example we saw that the symplectic gradient satisfies what are called Hamilton's equations:

$$\mathrm{d}q^i(X_H) = \frac{\partial H}{\partial p^i}, \quad \mathrm{d}p^i(X_H) = -\frac{\partial H}{\partial q^i}.$$

Writing out the differential of E in terms of the 1-form

$$\mathrm{d}E_{\varphi_t} = \sum_{i=1}^{n}\frac{\partial E}{\partial q^i}\,\mathrm{d}q^i_{\varphi_t} + \sum_{j=1}^{n}\frac{\partial E}{\partial p^j}\,\mathrm{d}p^j_{\varphi_t}$$

and using the property of the flow

$$\varphi_{t*} = \frac{\partial}{\partial t}\varphi_t = (X_H)_{\varphi_t}$$

we then calculate the time derivative of $E(\varphi_t)$ as

$$\begin{aligned}
\mathrm{d}E_{\varphi_t}(\varphi_{t*}) &= \sum_{i=1}^{n}\frac{\partial E}{\partial q^i}\,\mathrm{d}q^i_{\varphi_t}(X_H) + \sum_{j=1}^{n}\frac{\partial E}{\partial p^j}\,\mathrm{d}p^j_{\varphi_t}(X_H) \\
&= \sum_{i=1}^{n}\frac{\partial E}{\partial q^i}\frac{\partial H}{\partial p^i}(\varphi_t) - \sum_{j=1}^{n}\frac{\partial E}{\partial p^j}\frac{\partial H}{\partial q^j}(\varphi_t) \\
&= \sum_{i=1}^{n}\left(\frac{\partial E}{\partial q^i}\frac{\partial H}{\partial p^i} - \frac{\partial E}{\partial p^i}\frac{\partial H}{\partial q^i}\right)\bigg|_{\varphi_t} = -\{H, E\}(\varphi_t).
\end{aligned}$$

Thus the Poisson bracket vanishes on φ_t if and only if the time derivative of $E(\varphi_t)$ vanishes. $\square$

Remark. Noether's theorem is equivalent to stating that E is a conserved quantity on the flow of the symplectic gradient of the Hamiltonian H if and only if the Poisson bracket of E and H vanishes.

Example 4.3. Let $(\mathbb{R}^2, \omega)$ be a symplectic manifold with $\omega = \mathrm{d}q \wedge \mathrm{d}p$ defined globally on $\mathbb{R}^2$. Let $H\colon \mathbb{R}^2 \to \mathbb{R}$ be the Hamiltonian given by $H(q,p) = p^2$. Let E be the Hamiltonian given by $E(q,p) = p$. The symplectic gradient of H is

$$X_H = 2p\,\frac{\partial}{\partial q}$$

so the flow associated to H is then

$$\varphi_t(x,y) = (2yt + x, y).$$

The quantity $E(q,p)$ is conserved on φ_t by Noether's theorem.
We may easily extrapolate to the $2n$-dimensional case, where if

$$H((q^1,\dots q^n) + (\delta^1,\dots,\delta^n), p^1,\dots,p^n) = H(q^1,\dots,q^n,p^1,\dots,p^n)$$

is satisfied for any $(\delta^1,\dots,\delta^n) \in \mathbb{R}^n$ then the quantity

$$E(q^1,\dots,q^n,p^1,\dots,p^n) = \sum_{i=1}^{n} p^i$$

is conserved on the flow associated to the Hamiltonian H. We say that H is translation invariant and that its corresponding symmetry is conservation of momentum.

5. Sobolev Spaces and Fourier Analysis

Def. 5.1. Let $f\colon C^k(\Omega) \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$ open, be a k times continuously differentiable function. Then we introduce *multi-index notation*

$$D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}$$

where α is a multi-index of order k: $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n \leq k$. It is customary to write α as an n-tuple of non-negative integers:

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}_0^n, \quad \mathbb{N}_0 = \mathbb{N} \cup \{0\}.$$

Example 5.1. The derivative of $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$\frac{\partial^3 f}{\partial x^2 \partial y}(x,y,z) = D^{(2,1,0)} f$$

with multi-index $\alpha = (2,1,0)$ illustrates the compactness of the notation. △

We recall some basic definitions from functional analysis. Further information on these subjects may be found in Brezis.[3]

Def. 5.2. (Cauchy sequences and Banach spaces): A sequence $\{x_k\}_{k\in\mathbb{N}}$ in a normed space $(X, \|\cdot\|_X)$ is called a *Cauchy sequence* if it satisfies the following:

$$\forall \epsilon > 0\, \exists N \in \mathbb{N} : \|x_k - x_m\| < \epsilon \text{ for } k,m > N.$$

A normed space $(X, \|\cdot\|_X)$ in which all Cauchy sequences converge is called a *Banach space.*

Def. 5.3. An inner product space $(X, \langle\cdot,\cdot\rangle)$ may inherit the structure of a normed space by considering the *induced norm* defined by $\|x\| = \sqrt{\langle x, x\rangle}$. Inner product spaces which are Banach spaces with respect to the induced norm are called *Hilbert spaces.*

We state without proof an important theorem from functional analysis.

Theorem 5.1. (Completion theorem for normed spaces)
Every normed space is densely and isometrically embedded in a Banach space. In other words, there exists an isomorphism ϕ which preserves distances between the normed space $(X, \|\cdot\|_X)$ and a subset of a Banach space $(Y, \|\cdot\|_Y)$ such that $\mathrm{clos}(\phi(X)) = Y$, where "clos" denotes the closure.

Def. 5.4. (L^p-spaces)
The space of p-integrable functions are called the *L^p-spaces.* Let $\Omega \subseteq \mathbb{R}^n$. Then for $1 \le p < \infty$:

$$L^p(\Omega) = \left\{ f \in C(\Omega, \mathbb{C}) \,\middle|\, \left(\int_\Omega |f(x)|^p \,\mathrm{d}x \right)^{1/p} < \infty \right\}$$

where functions $f \in L^p(\Omega)$ have an associated p-norm given by

$$\|f\|_p = \left(\int_\Omega |f(x)|^p \,\mathrm{d}x \right)^{1/p}.$$

The case $p = \infty$ defines a space $L^\infty(\Omega)$ of functions where

$$\|f\|_\infty = \inf\{\, C \ge 0 \mid |f(x)| \le C \text{ for a.e. } x \in \Omega \,\} < \infty.$$

Remark. Here we have defined L^p-spaces for continuous functions, but there is a weaker requirement called *Lebesgue integrable* which is measure-theoretic in nature. "*a.e.*" means "almost every" or "almost everywhere" (on a set) and a property holds almost everywhere if it holds everywhere except for on subsets with *measure zero.* More information on this may be found in Brezis.[3]

Def. 5.5. (Schwartz space)
The space of rapidly decreasing functions on $\mathbb{R}^n$ is called the *Schwartz space* over $\mathbb{R}^n$ and is defined by the set

$$\mathscr{S}(\mathbb{R}^n) = \{\, u \in C^\infty(\mathbb{R}^n) \mid x^\beta D^\alpha u \in L^\infty(\mathbb{R}^n) \quad \forall\, \alpha, \beta \in \mathbb{N}_0^n \,\}.$$

Note that $x^\beta := x^{\beta_1} \cdots x^{\beta_n}$ for a multi-index β.

Recall that a function $f : \mathbb{R}^n \to \mathbb{C}$ has a Fourier transform given by

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

The inverse Fourier transform is given by

$$\mathscr{F}^{-1}(\hat{f}) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \hat{f}(\xi)\, e^{ix\cdot\xi} \,\mathrm{d}\xi.$$

Proposition 5.1. The following properties hold for the Fourier transformation for functions on $\mathscr{S}(\mathbb{R}^n)$

(i) The inverse Fourier transform is well-defined and unique,
(ii) $\mathscr{F}(D^\alpha f)(\xi) = (i\xi)^\alpha \hat{f}(\xi)$,
(iii) $\mathscr{F}(x^\alpha f)(\xi) = (iD_\xi)^\alpha \hat{f}(\xi)$,

where D_ξ is the multi-index derivative with respect to ξ.

Theorem 5.2. (Symmetries of the Fourier transform)
Let $f \in \mathscr{S}(\mathbb{R})$.

(i) If f is an even function then $\hat{f}$ is an even function.
(ii) If f is an odd function then $\hat{f}$ is an odd function.
(iii) If f is real and even then $\hat{f}$ is real and even.
(iv) If f is real and odd then $\hat{f}$ is imaginary and odd.

Proof. (i) Recall that if f is even then $f(-x) = f(x)$ for all $x \in \mathbb{R}$. Consider then

$$\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)\, e^{-ix\xi}\, \mathrm{d}x$$

where taking $\xi \mapsto -\xi$ amounts to

$$\begin{aligned}\hat{f}(-\xi) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)\, e^{ix\xi}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{-\infty} f(-x)\, e^{-ix\xi}\, (-\mathrm{d}x) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)\, e^{-x\xi}\, \mathrm{d}x = \hat{f}(\xi)\end{aligned}$$

thus $\hat{f}(\xi)$ is also an even function.
(ii) Similarly as (i), since if f is odd then $f(x) = -f(-x)$ for all x.
(iii) Assume f is real and even. Recall that $e^{-ix\xi} = \cos x\xi - i \sin x\xi$ by Euler's formula. Using this on the Fourier transform we get

$$\sqrt{2\pi}\hat{f}(\xi) = \int_{\mathbb{R}} f(x)\, e^{-ix\xi}\, \mathrm{d}x = \int_{\mathbb{R}} f(x) \cos x\xi - i \sin x\xi\, \mathrm{d}x = \int_{\mathbb{R}} f(x) \cos x\xi\, \mathrm{d}x$$

since $f(x) \sin x\xi$ is even times odd which integrates to zero over $\mathbb{R}$. Thus $\hat{f}$ is real and even since f is real and even.
(iv) Similarly as (iii). □

Theorem 5.3. The Fourier transformation is an automorphism on $\mathscr{S}(\mathbb{R}^n)$.

Proof. Let $f(x) \in \mathscr{S}(\mathbb{R}^n)$. Then we have

$$x^\beta D^\alpha f(x) \in L^\infty(\mathbb{R}^n)$$

for any multi-indices α, β. Consider the Fourier transform of f

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

We need to show that $\hat{f}(\xi)$ is in the Schwartz space $\mathscr{S}(\mathbb{R}^n)$, so we need to show that $\|\xi^\beta D_\xi^\alpha \hat{f}(\xi)\|_{L^\infty(\mathbb{R}^n)}$ is finite for any multi-indices α, β. Expanding this norm according to Proposition 5.1 amounts to

$$\|\xi^\beta D_\xi^\alpha \hat{f}(\xi)\|_{L^\infty(\mathbb{R}^n)} = \left\| \frac{i^{|\alpha|+|\beta|}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} x^\alpha (D^\beta f(x))\, e^{-ix\cdot\xi}\, \mathrm{d}x \right\|_{L^\infty(\mathbb{R}^n)}.$$

Consider then the following with $C = i^{|\alpha|+|\beta|}/(2\pi)^{\frac{n}{2}}$:

$$\begin{aligned}
|\xi^\beta D_\xi^\alpha \hat{f}(\xi)| &= |C| \left| \int_{\mathbb{R}^n} x^\alpha D^\beta f(x) e^{-ix\cdot\xi}\,\mathrm{d}x \right| \\
&\leq |C| \int_{\mathbb{R}^n} |x^\alpha D^\beta f(x) e^{-ix\cdot\xi}|\,\mathrm{d}x \\
&\leq |C| \int_{\mathbb{R}^n} |x^\alpha D^\beta f(x)|\,\mathrm{d}x \\
&= |C| \int_{\mathbb{R}^n} \left| x^\alpha D^\beta f(x) \frac{(1+\|x\|_2^2)^m}{(1+\|x\|_2^2)^m} \right| \mathrm{d}x, \quad m \in \mathbb{N}.
\end{aligned}$$

Note that $|x^\alpha D^\beta f(x)(1+\|x\|_2^2)^m|$ is bounded since $f(x) \in \mathscr{S}(\mathbb{R}^n)$, then this same function is bounded with respect to a L^∞-norm:

$$\|x^\alpha D^\beta f(x)(1+\|x\|_2^2)^m\|_{L^\infty(\mathbb{R}^n)} < \infty.$$

Using this finite norm we establish the inequality

$$|\xi^\beta D_\xi^\alpha \hat{f}(\xi)| \leq |C| \cdot \|x^\alpha D^\beta f(x)(1+\|x\|_2^2)^m\|_{L^\infty(\mathbb{R}^n)} \int_{\mathbb{R}^n} \frac{1}{(1+\|x\|_2^2)^m}\,\mathrm{d}x$$

which is indeed finite for appropriate $m \in \mathbb{N}$. Thus we have established that

$$\|\xi^\beta D_\xi^\alpha \hat{f}(\xi)\|_{L^\infty(\mathbb{R}^n)} < \infty$$

for any multi-indices α, β. This shows that $\hat{f}(\xi) \in \mathscr{S}(\mathbb{R}^n)$.

Since $\mathscr{F} : \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ is clearly linear, and furthermore bijective due to it clearly being injective by

$$\mathscr{F}(f) = \mathscr{F}(g) \iff \int_{\mathbb{R}^n} (f(x) - g(x))\, e^{-ix\cdot\xi}\,\mathrm{d}x = 0 \iff f(x) = g(x)$$

and surjective due to the inverse being well-defined by Proposition 5.1. Then $\mathscr{F} \colon \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ is an isomorphism of vector spaces, and therefore an automorphism on $\mathscr{S}(\mathbb{R}^n)$. □

Remark. The Fourier transformation exhibits especially nice properties as an automorphism since we can look at the derivative $D^\alpha f(x) \in \mathscr{S}(\mathbb{R}^n)$ and take Fourier transform to $(i\xi)^\alpha \hat{f}(\xi)$, work with this expression and transform back using the inverse Fourier transformation. This is often a powerful tool in the analysis of PDEs.

Theorem 5.4. (Plancherel's theorem)
Let $\Omega \subseteq \mathbb{R}^n$. Then the Fourier transformation can be extended to a unitary transformation on the Hilbert space $L^2(\Omega)$, satisfying

$$\langle \hat{f}, \hat{g} \rangle_{L^2(\Omega)} = \langle f, g \rangle_{L^2(\Omega)} \quad \text{for } f, g \in L^2(\Omega).$$

Def. 5.6. (Homogeneous Sobolev spaces)
The *homogeneous Sobolev space* $\dot{H}^s(\mathbb{R}^n)$ is the completion of $\mathscr{S}(\mathbb{R}^n)$ with respect to the *homogeneous norm* $\|\cdot\|_{\dot{H}^s(\mathbb{R}^n)}$. Explicitly this space is

$$\dot{H}^s(\mathbb{R}^n) = \operatorname{clos}(\mathscr{S}(\mathbb{R}^n), \|f\|_{\dot{H}^s}),$$

where the homogeneous norm is given by

$$\|f\|_{\dot{H}^s(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |\xi|^{2s}\, |\hat{f}(\xi)|^2\,\mathrm{d}\xi.$$

6. Outline of Basic Techniques for Linear PDEs

In order to understand the next section, it is useful to gather some intuition for why analysis of PDEs may still be useful even if one does not solve the equations explicitly. This section focuses on illustrating some basic techniques for analyzing *linear* PDEs.

Consider a fluid flow through a pipe. We wish to model how the density of water $u(t,x)$ dispersed in the pipe changes with respect to time t and position x when we model the pipe as 1-dimensional. We are interested in $u(t,x)$ for $t \geq 0$ and $x \in [0,L] \subset \mathbb{R}$. What physical considerations are relevant when analyzing fluid flow through such a pipe? We may for instance impose that the pipe has a constant volume of fluid within it at all times, so if $V(t)$ is the total volume of fluid in the pipe at time t we have

$$V(t) = \int_0^L u(t,x)\,\mathrm{d}x \quad \text{and} \quad \frac{\mathrm{d}V}{\mathrm{d}t} = \int_0^L \frac{\partial u}{\partial t}(t,x)\,\mathrm{d}x = 0.$$

Given that the pipe may also have fluid going in at $x = 0$ and out at $x = L$, we may define the flux of fluid at time t and position x to be $Q(t,x)$. If the fluid has a homogeneous rate of flow such that the velocity at each point is uniformly $v \in \mathbb{R}$, we may write $Q(t,x) = v \cdot u(t,x)$. Considering that the change in the total volume of fluid is the fluid going in at $x = 0$ plus the fluid going out at $x = L$, we have that

$$0 = \frac{\mathrm{d}V}{\mathrm{d}t} = Q(t,0) - Q(t,L) = -\int_0^L \frac{\partial Q}{\partial x}(t,x)\,\mathrm{d}x = -v\int_0^L \frac{\partial u}{\partial x}(t,x)\,\mathrm{d}x.$$

Combining our two equations we get

$$\int_0^L \left(\frac{\partial u}{\partial t} + v\,\frac{\partial u}{\partial x}\right)\mathrm{d}x = 0,$$

which in turn leaves us with the homogeneous, linear transport equation

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

Summarizing our work so far, we have shown that physical considerations have naturally given us a first order, homogeneous and linear PDE. How do we solve for the function $u(t,x)$? We may guess the constant solution $u(t,x) = C$ or an exponential solution $u(t,x) = e^{x-vt}$, but which of these, among the infinite family of possible solutions, is the one we are interested in? This leads to an important part of PDE analysis, namely that of boundary conditions and initial conditions - we need more information to solve PDE problems uniquely!

Consider now the dimensionally generalized version of the homogeneous transport equation

$$\partial_t u + v \cdot \nabla u = 0$$

where $t \geq 0$, $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ and $x \in \Omega \subseteq \mathbb{R}^n$. For the transport equation we wish to consider initial conditions $u(0,x) = g(x) \in C^1(\Omega)$, since we wish to have solutions $u(t,x) \in C^1([0,\infty) \times \Omega)$. So our problem is now

$$\partial_t u + v \cdot \nabla u = 0, \quad u(0,x) \in C^1(\Omega).$$

Another formulation of this equation is through the dot product of vectors given by

$$(1,v) \cdot (\partial_t u, \nabla u) = 0$$

for all solutions $u(t,x)$. The vector $(1,v) \in \mathbb{R}^{n+1}$ is based at the origin in $\mathbb{R}^{n+1}$ but we may translate this vector parallel to $(1,v)$ by a linear transformation $P : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ such that $(P(1,v)) \cdot (\partial_t u, \nabla u) = 0$. Thus any solution to the PDE has orthogonal derivatives to vectors parallel to $(1,v)$. This observation motivates parametrizing straight lines $(t+s, x+sv)$, $s \in \mathbb{R}$ and defining the function

$$z(s) := u(t+s, x+sv) \quad s \in \mathbb{R}.$$

If we differentiate with respect to the parameter s we obtain

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \partial_t u(z(s)) + v \cdot \nabla u(z(s)) = 0$$

which means that the function $z(s)$ is constant everywhere where it is defined. Of course, this means that $u(t+s, x+sv) = u(t,x)$ for all s such that the left hand side is defined. In particular, for $s = -t$, this means that $u(0, x-tv) = u(t,x)$. But $u(0, x-tv) = g(x-tv)$ where $g(x)$ is given! Our solution of the linear homogeneous transport equation is then

$$u(t,x) = g(x-tv) \in C^1(\Omega)$$

where $t \geq 0$, $x \in \Omega$, $u(0,x) = g(x) \in C^1(\Omega)$. What if the right hand side of the transport equation is non-homogeneous? Given the PDE

$$\partial_t u + v \cdot \nabla u = f(t,x)$$

and we may define, as before, the function

$$z(s) := u(t+s, x+sv)$$

and take its derivative with respect to the parameter s:

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \partial_t u(z(s)) + v \cdot \nabla u(z(s)) = f(z(s)).$$

Integrating with respect to the parameter s and including the homogeneous solution:

$$u(t+s, x+sv) = g(x-tv) + \int_{-t}^{s} f(t+s', x+s'v)\,\mathrm{d}s'$$

which means that the general solution is given by

$$u(t,x) = g(x-tv) + \int_{-t}^{0} f(t+s', x+s'v)\,\mathrm{d}s'$$

where $u(0,x) = g(x) \in C^1(\Omega)$, $(t,x) \in [0,\infty) \times \Omega$. The function $f(t,x)$ needs to be continuous for the PDE to make sense in our context.

In our example of the homogeneous and non-homogeneous transport equation, we get a taste of what is involved when analyzing PDEs, in this instance the analysis of symmetry and regularity. This particular method does not extend generally to non-linear PDEs, and may not even be effective for all types of linear PDEs. It does however motivate a "nice" first method for analyzing linear PDEs, which is the *method of characteristics*. A treatment of this method in a general setting may be found in Evans [5]. We turn our attention to a simpler treatment as found in Borthwick [2].

Consider the first-order PDE given by

$$\partial_t u + v \cdot \nabla u + w = 0$$

where $v = v(t, x)$ and $w = w(t, x, u)$ for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. The *characteristics* associated to this equation are the solutions of

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = v(t, x(t)).$$

We then define the *Lagrangian derivative* (also called the *material derivative*) of $u(t, x)$ by

$$\frac{Du}{Dt} = \frac{\mathrm{d}}{\mathrm{d}t} u(t, x(t)).$$

Theorem 6.1. Consider the equation given by

$$\partial_t u + v \cdot \nabla u + w = 0$$

for $v = v(t, x)$ and $w = w(t, x, u)$ on a domain $(t, x) \in I \times \Omega \subseteq \mathbb{R}^{n+1}$.
Then on every characteristic curve $x(t)$ this PDE reduces to the ODE given by

$$\frac{Du}{Dt} + w|_{x(t)} = 0.$$

Proof. Let $x(t)$ be a characteristic curve associated to the given PDE, then by the chain rule applied to the Lagrangian derivative one has

$$\frac{Du}{Dt} = \partial_t u(t, x(t)) + \nabla u(t, x(t)) \cdot \frac{\mathrm{d}x}{\mathrm{d}t}.$$

But the characteristic curve satisfies $\dot{x}(t) = v(t, x(t))$, so adding $w|_{x(t)}$ to this becomes

$$\frac{Du}{Dt} + w|_{x(t)} = \partial_t u(t, x(t)) + v \cdot \nabla u(t, x(t)) + w|_{x(t)} = 0.$$

□

Example 6.1. Let $\Omega \subseteq \mathbb{R}$ and consider the PDE

$$\partial_t u + x \partial_x u + tx = 0.$$

Then a characteristic curve $x(t)$ starting at $a \in [0, 1]$ satisfying

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = x(t)$$

has the solution given as $x(t) = a\, e^t$. The Lagrangian derivative is then

$$\frac{Du}{Dt} + t \cdot x(t) = 0,$$

which we may integrate with respect to t to obtain

$$u(t, x(t)) = u(0, a) + \int_0^t -at\, e^t \, \mathrm{d}t = u(0, a) - at\, e^t + a(e^t - 1).$$

Every $a \in \Omega$ may be given as $a = x(t)\, e^{-t}$, furthermore:

$$u(t, x(t)) = u(0, x(t)\, e^{-t}) - tx(t) + x(t) - x(t)\, e^{-t}$$

Provided with the initial condition $u(0, x) = g(x) \in C^1(\Omega)$ and the identification $x = x(t)$, we obtain

$$u(t, x) = g(x\, e^{-t}) - xt + x - e^{-t} x$$

which one can verify is a solution to our intial value problem. △

7. Analyzing PDEs using the symplectic formalism

7.1. Using symplecticity for PDEs over Sobolev spaces.

Loosely and informally speaking, a set B which is an abstract C^k-manifold where Euclidean spaces are replaced with Banach spaces and derivatives replaced with Frechét derivatives is called a *Banach Manifold* of class C^k. More information on more general manifolds can be found in Serge Lang's "Differential Manifolds" [7].

Following the result of Alan Weinstein [11], we know that a Banach manifold can be equipped with a symplectic form Ω. How does the symplectic form look like in coordinates? Weinstein proves that if B is a Hilbert space and $\{\xi_i\}_{i\in I} \cup \{\eta_i\}_{i\in I}$ is a basis for the dual space B^*, then the symplectic form can be written as

$$\Omega = \sum_{i\in I} \xi_i \wedge \eta_i$$

If $B = \dot{H}^s(\mathbb{R})$, then the symplectic form Ω is given by the wedge product of basis-elements of $B^* = \dot{H}^{-s}(\mathbb{R})$. As one can imagine, finding such a basis and computing the wedge product as above proves to be a difficult task. Therefore, we resort to an implicit treatment of this symplectic form.

Working with $\dot{H}^s(\mathbb{R})$ and $s = 1/2$ gives a Sobolev space and its dual $\dot{H}^{-1/2}(\mathbb{R})$. We consider the bilinear function given by

$$\Omega(u, v) = 2\int_{\mathbb{R}} (\partial_x^{-1} u(x))(v(x))\,\mathrm{d}x$$

where ∂_x^{-1} is the inverse of ∂_x. One can show that this is a symplectic form on $\dot{H}^{1/2}(\mathbb{R})$. So then we may consider the map $\partial_x \colon \dot{H}^{1/2}(\mathbb{R}) \to \dot{H}^{-1/2}(\mathbb{R})$ by inspecting the Sobolev norm of $\partial_x u$

$$\int_{\mathbb{R}} \frac{|i\xi \hat{u}(\xi)|^2}{|\xi|}\,\mathrm{d}\xi = \int_{\mathbb{R}} \frac{|\xi|^2 |\hat{u}(\xi)|^2}{|\xi|}\,\mathrm{d}\xi$$

given that $u \in \dot{H}^{1/2}(\mathbb{R})$, then the right hand side is finite and thus $\partial_x u \in \dot{H}^{-1/2}(\mathbb{R})$. Considering the Fourier transform of $\partial_x u$ it is natural to define the map ∂_x^{-1} by

$$\partial_x^{-1} \colon \hat{u}(\xi) \longmapsto \frac{\hat{u}(\xi)}{i\xi}.$$

For $u \in \dot{H}^{-1/2}(\mathbb{R})$ we have that

$$\int_{\mathbb{R}} |\xi| \left|\left(\frac{\hat{u}(\xi)}{i\xi}\right)\right|^2 \mathrm{d}\xi = \int_{\mathbb{R}} \frac{|\hat{u}(\xi)|^2}{|\xi|}\,\mathrm{d}\xi < \infty.$$

This shows that $\partial_x^{-1} u \in \dot{H}^{1/2}(\mathbb{R})$. By Plancherel's theorem and Proposition 5.1 we have that the map ∂_x^{-1} is indeed the inverse of ∂_x, and that

$$\partial_x^{-1} u(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\hat{u}(\xi)}{i\xi}\, e^{ix\xi}\,\mathrm{d}\xi.$$

Example 7.1. Consider the Hamiltonian given by the integral

$$H(u) = \int_{\mathbb{R}} u^2\,\mathrm{d}x.$$

The differential of such a function is given by the *Gateaux derivative*

$$\mathrm{d}H(u;v) = \frac{\mathrm{d}}{\mathrm{d}t}H(u+tv)\bigg|_{t=0}.$$

So then, given the symplectic form as before

$$\Omega(u,v) = 2\int_{\mathbb{R}} (\partial_x^{-1}u(x))v(x)\,\mathrm{d}x$$

we can solve for the symplectic gradient X_H by calculating for the function $u(t,\cdot) \in \dot{H}^{1/2}(\mathbb{R})$, where we regard t as a fixed parameter,

$$\begin{aligned}\mathrm{d}H(u;v) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} (u+tv)^2\,\mathrm{d}x\bigg|_{t=0} = \int_{\mathbb{R}} \frac{\partial}{\partial t}(u+tv)^2\,\mathrm{d}x\bigg|_{t=0} \\ &= \int_{\mathbb{R}} 2(u+tv)v\,\mathrm{d}x\bigg|_{t=0} = 2\int_{\mathbb{R}} uv\,\mathrm{d}x.\end{aligned}$$

Solving for X_H requires that

$$\mathrm{d}H(u;v) = 2\int_{\mathbb{R}} (\partial_x^{-1}X_H(u))v\,\mathrm{d}x$$

which after comparing the integrals furthermore implies that

$$\partial_x^{-1}X_H(u) = u$$

using the invertibility of ∂_x we finally obtain

$$X_H(u) = \partial_x u,\ X_H = \frac{\partial}{\partial x}.$$

The flow associated to this symplectic gradient may be written as

$$\partial_t u = \partial_x u$$

where t is now varying as a parameter. Note that the solution to this PDE has to be in our Sobolev space $\dot{H}^{1/2}(\mathbb{R})$, so an initial value problem with this PDE needs initial conditions in $\dot{H}^{1/2}(\mathbb{R})$. This rules out constant solutions and other solutions which do not decay sufficiently fast over $\mathbb{R}$. $\triangle$

Recall that Noether's theorem gives that flows $u(t,x)$ of symplectic gradients associated to a Hamiltonian H are in particular constant when H is evaluated along $u(t,x)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}H(u(t,x)) = 0.$$

For PDEs which are more difficult to analyze, the Hamiltonian formulation proves to be very useful. The symplectic formalism works for non-linear equations as well. We demonstrate this with the example

Example 7.2. Let H be a Hamiltonian given by

$$H(u) = \int_{\mathbb{R}} u^3 + \partial_x u\,\mathrm{d}x.$$

Then we calculate

$$\begin{aligned}
\mathrm{d}H(u;v) &= \int_{\mathbb{R}} \frac{\partial}{\partial t}\left[(u+tv)^3 + \partial_x u + t\partial_x v\right] \,\mathrm{d}x\bigg|_{t=0} \\
&= \int_{\mathbb{R}} 3u^2 v + \partial_x v \,\mathrm{d}x = \int_{\mathbb{R}} (3u^2 - x)v \,\mathrm{d}x \\
&= \int_{\mathbb{R}} (\partial_x^{-1} X_H(u))v \,\mathrm{d}x
\end{aligned}$$

which means that the symplectic gradient satisfies

$$X_H(u) = \partial_x(3u^2 - x) = 6u\,\partial_x u - 1.$$

The flow is then given by

$$\partial_t u = 6u\,\partial_x u - 1$$

which is a quasi-linear, non-homogeneous, first order PDE. △

Other PDEs like the Korteweg-de Vries (KdV) equation which models shallow water waves is of great importance in the study of non-linear dispersive PDEs and hydrodynamics and is classically given by

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

This equation admits several Hamiltonians as the generator of the flow $\partial_t u$. Let us consider a few of these given by

$$H(u) = \int_{\mathbb{R}} 2u^3 + (\partial_x u)^2 \,\mathrm{d}x$$

where we then calculate

$$\begin{aligned}
\mathrm{d}H(u;v) &= \int_{\mathbb{R}} \frac{\partial}{\partial t}[2(u+tv)^3 + (\partial_x u + t\partial_x v)^2] \,\mathrm{d}x\bigg|_{t=0} \\
&= \int_{\mathbb{R}} 6(u+tv)^2 v + 2(\partial_x u + t\partial_x v)\partial_x v \,\mathrm{d}x\bigg|_{t=0} \\
&= \int_{\mathbb{R}} 6u^2 v + 2\partial_x u\,\partial_x v \,\mathrm{d}x = \int_{\mathbb{R}} (6u^2 - 2\partial_x^2 u)v \,\mathrm{d}x
\end{aligned}$$

comparing with the symplectic form we have

$$\partial_x^{-1} X_H(u) = 3u^2 - \partial_x^2 u$$

which then means that the flow of the symplectic gradient is given by

$$\partial_t u = 6u\partial_x u - \partial_x^3 u$$

so the flow of the symplectic gradient is indeed the KdV equation. What other Hamiltonians may be interesting for this PDE? Generally, as indicated in Section 6, we are interested in analyzing symmetries of the solution to the PDE. Employing Noether's theorem is a good strategy for analyzing what quantities are conserved on the flow of the symplectic gradient.

Consider the Hamiltonian E_1 given by

$$E_1(u) = \int_{\mathbb{R}} u \,\mathrm{d}x.$$

The flow of the symplectic gradient X_{E_1} is found to satisfy

$$\partial_t u = 0$$

because of the relation

$$\mathrm{d}E_1(u;v) = \int_{\mathbb{R}} v \, \mathrm{d}x = 2\int_{\mathbb{R}} \partial_x^{-1} X_{E_1}(u) v \, \mathrm{d}x$$

which implies that its flow is $u(t,x) = f(x)$ for a function $f \in \dot{H}^{1/2}(\mathbb{R})$.
We can then calculate the Poisson bracket $\{H, E_1\}$ by

$$\begin{aligned}\{H, E_1\}(u) &= \Omega(X_H(u), X_{E_1}(u)) = 2\int_{\mathbb{R}} (\partial_x^{-1} X_H(u)) X_{E_1}(u) \, \mathrm{d}x \\ &= 2\int_{\mathbb{R}} (3u^2 - \partial_x^2 u) X_{E_1}(u) \, \mathrm{d}x = 0 \qquad (X_{E_1}(u) = 0)\end{aligned}$$

which by Noether's theorem implies that the quantity E_1 is conserved along the flow of the symplectic gradient X_H.

Consider another Hamiltonian E_2 given by

$$E_2(u) = \int_{\mathbb{R}} u^2 \, \mathrm{d}x$$

as examined in a previous example, we have that the flow of X_{E_2} is

$$\partial_t u = \partial_x u.$$

The Poisson bracket is, knowing that $X_{E_2}(u) = \partial_x u$:

$$\begin{aligned}\{H, E_2\}(u) &= \Omega(X_H(u), X_{E_2}(u)) = 2\int_{\mathbb{R}} (\partial_x^{-1} X_H(u)) X_{E_2}(u) \, \mathrm{d}x \\ &= 2\int_{\mathbb{R}} (3u^2 - \partial_x^2 u)\partial_x u \, \mathrm{d}x = \int_{\mathbb{R}} \partial_x (2u^3 - (\partial_x u)^2) \, \mathrm{d}x = 0\end{aligned}$$

since for $u(t,\cdot) \in \dot{H}^{1/2}(\mathbb{R})$ the integrand decays sufficiently fast:

$$\int_{\mathbb{R}} \partial_x (2u^3 - (\partial_x u)^2) \, \mathrm{d}x = 2u^3 - (\partial_x u)^2 \Big|_{x=-\infty}^{\infty} = 0.$$

Knowing that the corresponding symmetry of the conserved quantity E_2 is spatial translation, one may assert that $E(u(t, x + v)) = E(u(t, x))$ for all t, x. Again, this is useful for when classical techniques fail or when, for example, physical quantities that are naturally conserved are considered.

What if one needs to reverse engineer from a PDE to a Hamiltonian H? As an example, consider the setting

$$\partial_t u + u\partial_x u + L\partial_x u = 0$$

where $L \colon \dot{H}^{1/2}(\mathbb{R}) \to \dot{H}^{-1/2}(\mathbb{R})$ is a linear operator. As a further specialization, we may assume $L = \partial_x^2$. This means that

$$\partial_t u + u\partial_x u + \partial_x^3 u = 0$$

which after rearranging gives the flow corresponding to X_H by

$$\partial_t u = -u\partial_x u - \partial_x^3 u = X_H(u).$$

Using the inverse operator ∂_x^{-1} on the symplectic gradient gives

$$\partial_x^{-1} X_H(u) = -\frac{1}{2}u^2 - \partial_x^2 u$$

which furthermore can be used in the expression

$$\mathrm{d}H(u;v) = 2\int_{\mathbb{R}} \partial_x^{-1} X_H(u) v \,\mathrm{d}x = \int_{\mathbb{R}} (-u^2 - 2\partial_x^2 u)\,\mathrm{d}x.$$

Consider now the sufficiently general Hamiltonian for $p, k, l \in \mathbb{N}_0$:

$$H(u) = \int_{\mathbb{R}} u^p + (\partial_x u)^k + (\partial_x^2 u)^l \,\mathrm{d}x$$

thus

$$\begin{aligned}
\mathrm{d}H(u;v) &= \int_{\mathbb{R}} p(u+tv)^{p-1} v + k(\partial_x u + t\partial_x v)^{k-1}\partial_x v + l(\partial_x^2 u + t\partial_x^2 v)^{l-1}\partial_x^2 v \,\mathrm{d}x \Big|_{t=0} \\
&= \int_{\mathbb{R}} p u^{p-1} v + k(\partial_x u)^{k-1}\partial_x v + l(\partial_x^2 u)^{l-1}\partial_x^2 v \,\mathrm{d}x \\
&= \int_{\mathbb{R}} \big[p u^{p-1} - k(k-1)(\partial_x u)^{k-2}\partial_x^2 u \\
&\qquad + l(l-1)(l-2)(\partial_x^2 u)^{l-3}\,(\partial_x^3 u)^2 + l(l-1)(\partial_x^2 u)^{l-2}(\partial_x^4 u)\big]\, v \,\mathrm{d}x
\end{aligned}$$

where the last equality is due to integration by parts and the product rule. Comparing the expressions for $\mathrm{d}H(u;v)$ and scaling appropriately we find that necessarily $p = 3$, $k = 2$, $l = 0$ and that the Hamiltonian H is given by

$$H(u) = \int_{\mathbb{R}} -\frac{1}{3}u^3 + (\partial_x u)^2 \,\mathrm{d}x.$$

Another Hamiltonian of the same PDE happens to be

$$H(u) = \int_{\mathbb{R}} -\frac{1}{3}u^3 + (\partial_x u)^2 + \partial_x^2 u \,\mathrm{d}x$$

since the numbers $p = 3$, $k = 2$, $l = 1$ also work in our previous examination. Notice how these are equivalent Hamiltonian functions for the KdV equation. These two Hamiltonians are actually exactly the same, since the last term in the latter Hamiltonian evaluates to zero when integrated.

Consider now a more general linear operator

$$L = \sum_{i=0}^{m} a_i \frac{\partial^i}{\partial x^i} = \sum_{i=0}^{m} a_i\, \partial_x^i.$$

The symplectic gradient satisfies

$$\mathrm{d}H(u;v) = 2\int_{\mathbb{R}} \partial_x^{-1} X_H(u) v \,\mathrm{d}x = \int_{\mathbb{R}} \left[-u^2 - 2\sum_{i=0}^{m} a_i\, \partial_x^i u\right] v \,\mathrm{d}x$$

which is then equivalent to

$$\mathrm{d}H(u;v) = \int_{\mathbb{R}} (-u^2 - 2Lu)\, v \,\mathrm{d}x$$

The reader is not encouraged to try to find this general Hamiltonian, since it happens to be quite more difficult than first meets the eye. Consider first the linear operator $L = \frac{1}{2}\,\partial_x$. To this end, consider the more general Hamiltonian H given by

$$H(u) = \int_{\mathbb{R}} F(u, \partial_x u)\,\mathrm{d}x$$

for some function $F(u, \partial_x u)$, which furthermore means that

$$\begin{aligned}
\mathrm{d}H(u; v) &= \int_{\mathbb{R}} \partial_t F(u + tv, \partial_x u + t\partial_x v)\,\mathrm{d}x\Big|_{t=0} \\
&= \int_{\mathbb{R}} \frac{\partial F}{\partial u} v + \frac{\partial F}{\partial(\partial_x u)} \partial_x v\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \left[\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\frac{\partial F}{\partial(\partial_x u)}\right] v\,\mathrm{d}x.
\end{aligned}$$

We wish to have a Hamiltonian H which, up to a scalar, satisfies

$$\mathrm{d}H(u; v) = \int_{\mathbb{R}} \partial_x u\, v\,\mathrm{d}x.$$

Thus we have

$$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\frac{\partial F}{\partial(\partial_x u)} = \partial_x u.$$

Based on this equation we propose the following Hamiltonian function as a solution to our problem given our operator L.

Claim: The Hamiltonian given by

$$H(u) = \int_{\mathbb{R}} -\frac{1}{2}\partial_x^{-1}(\partial_x u)^2\,\mathrm{d}x$$

satisfies the property that

$$\mathrm{d}H(u; v) = \int_{\mathbb{R}} \partial_x u\, v\,\mathrm{d}x.$$

Proof. Let $H(u)$ be given as

$$H(u) = \int_{\mathbb{R}} \partial_x^{-1}(\partial_x u)^2\,\mathrm{d}x.$$

Then we calculate

$$\begin{aligned}
\mathrm{d}H(u; v) &= \int_{\mathbb{R}} \partial_t \partial_x^{-1}(\partial_x u + tv)^2\,\mathrm{d}x\Big|_{t=0} = \int_{\mathbb{R}} \partial_x^{-1}\partial_t(\partial_x u + tv)^2\,\mathrm{d}x\Big|_{t=0} \\
&= \int_{\mathbb{R}} \partial_x^{-1}(2(\partial_x u + t\partial_x v)\partial_x v)\,\mathrm{d}x\Big|_{t=0} = \int_{\mathbb{R}} \partial_x^{-1}(2\partial_x u\, \partial_x v)\,\mathrm{d}x.
\end{aligned}$$

There exists a function $G(u, \partial_x u, v, \partial_x v)$ such that

$$\int_{\mathbb{R}} \partial_x^{-1}(2\partial_x u\,\partial_x v)\,\mathrm{d}x = \int_{\mathbb{R}} G(u, \partial_x u, v, \partial_x v)\,\mathrm{d}x$$

which after applying the operator ∂_x on both integrands gives

$$\begin{aligned}
\int_{\mathbb{R}} 2\partial_x u\,\partial_x v\,\mathrm{d}x &= \int_{\mathbb{R}} \partial_x G(u, \partial_x u, v, \partial_x v)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \frac{\partial G}{\partial u}\partial_x u + \frac{\partial G}{\partial(\partial_x u)}\partial_x^2 u + \frac{\partial G}{\partial v}\partial_x v + \frac{\partial G}{\partial(\partial_x v)}\partial_x^2 v\,\mathrm{d}x.
\end{aligned}$$

Integration by parts on the first integral gives

$$\int_{\mathbb{R}} -2(\partial_x^2 u)\, v\,\mathrm{d}x = \int_{\mathbb{R}} \frac{\partial G}{\partial u}\partial_x u + \frac{\partial G}{\partial(\partial_x u)}\partial_x^2 u + \frac{\partial G}{\partial v}\partial_x v + \frac{\partial G}{\partial(\partial_x v)}\partial_x^2 v\,\mathrm{d}x$$

which when comparing each term and applying integration by parts when necessary amounts to the equations

$$\frac{\partial G}{\partial u} = 0, \quad \frac{\partial G}{\partial(\partial_x u)} = -2v, \quad \frac{\partial G}{\partial v} = -2\partial_x u, \quad \frac{\partial G}{\partial(\partial_x v)} = 0.$$

These equations imply that $G(u, \partial_x u, v, \partial_x v) = -2\partial_x u\, v$, so

$$\mathrm{d}H(u; v) = \int_{\mathbb{R}} -2\partial_x u\, v\, \mathrm{d}x$$

which after scaling the Hamiltonian gives the desired property. $\square$

Remark. The observation to make from this example of L is that our conventional method for finding a suitable Hamiltonian for a given PDE might not be optimal. Even finding a term linear in the derivative of u is quite the ordeal, and finding Hamiltonians for other linear operators L might be more difficult, perhaps even impossible in certain cases.

We turn our attention to the case where the linear operator L in

$$\partial_t u + u\partial_x u + L\partial_x u = 0$$

is a *Fourier multiplier*, which is an operator satisfying

$$\mathscr{F}(Lu)(\xi) = m(\xi)\hat{u}(\xi)$$

for every $u(t, x)$ and a given complex-valued function $m(\xi)$.
In our setting we have the symplectic gradient given by

$$X_H(u) = -\frac{1}{2}\partial_x(u)^2 - L\partial_x u$$

applying the anti-derivative and multiplying by 2 gives

$$2\partial_x^{-1} X_H(u) = -u^2 - 2\partial_x^{-1}(L\partial_x u).$$

Using the definition and properties of the anti-derivative we have

$$\partial_x^{-1}(L\partial_x u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{i\xi} \mathscr{F}(L\partial_x u)(\xi)\, e^{ix\xi}\, \mathrm{d}\xi$$

where $\mathscr{F}(L\partial_x u)(\xi) = m(\xi)\mathscr{F}(\partial_x u)(\xi) = i\xi\, m(\xi)\hat{u}(\xi)$, so then

$$\partial_x^{-1}(L\partial_x u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} m(\xi)\hat{u}(\xi)\, e^{ix\xi}\, \mathrm{d}\xi.$$

The appropriate Hamiltonian generating the symplectic gradient needs to then satisfy

$$\mathrm{d}H(u; v) = \int_{\mathbb{R}} -u^2 v - \frac{2}{\sqrt{2\pi}} \int_{\mathbb{R}} m(\xi)\hat{u}(\xi)\, e^{ix\xi}\, \mathrm{d}\xi\, v\, \mathrm{d}x$$

where the first term is easily generated by a cubic term u^3, however the second term needs more examination to determine its generating Hamiltonian. In particular, we wish to find a Hamiltonian H_m such that

$$\mathrm{d}H_m(u; v) = \int_{\mathbb{R}} \frac{2}{\sqrt{2\pi}} \int_{\mathbb{R}} m(\xi)\hat{u}(\xi)\, e^{ix\xi}\, \mathrm{d}\xi\, v\, \mathrm{d}x$$

which due to the inverse Fourier transform is equivalent to the equation given by

$$\mathrm{d}H_m(u; v) = 2\int_{\mathbb{R}} (Lu) \cdot v\, \mathrm{d}x.$$

Proposition 7.1. The Hamiltonian H_m given by

$$H_m(u) = \int_{\mathbb{R}} (Lu) \cdot u \, \mathrm{d}x$$

satisfies the desired property that its differential is

$$\mathrm{d}H(u; v) = 2 \int_{\mathbb{R}} (Lu) \cdot v \, \mathrm{d}x.$$

Proof. We first treat the case where $m(\xi), u, v$ are all real-valued. Given $H_m(u)$ as above we obtain

$$\begin{aligned}
\mathrm{d}H_m(u; v) &= \int_{\mathbb{R}} \partial_t (Lu + tLv)(u + tv) \, \mathrm{d}x \bigg|_{t=0} \\
&= \int_{\mathbb{R}} Lv \cdot (u + tv) + (Lu + tLv) \cdot v \, \mathrm{d}x \bigg|_{t=0} \\
&= \int_{\mathbb{R}} (Lv) \cdot u + (Lu) \cdot v \, \mathrm{d}x
\end{aligned}$$

the latter term is the one we want, so we inspect the first term

$$Lv \cdot u = (\mathscr{F}^{-1}\{m(\xi)\, \hat{v}(\xi)\}) \cdot u = \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} m(\xi)\, \hat{v}(\xi)\, e^{ix\xi} \, \mathrm{d}\xi \right) \cdot u.$$

Integrating over x and using Fubini's theorem we have

$$\begin{aligned}
\int_{\mathbb{R}} Lv \cdot u \, \mathrm{d}x &= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} m(\xi)\, \hat{v}(\xi)\, u(x)\, e^{ix\xi} \, \mathrm{d}\xi \, \mathrm{d}x \\
&= \int_{\mathbb{R}} m(\xi)\, \hat{v}(\xi) \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u(x)\, e^{ix\xi} \, \mathrm{d}x \right) \mathrm{d}\xi.
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u(x)\, e^{ix\xi} \, \mathrm{d}x &= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{-\infty} u(-x)\, e^{-ix\xi} \, (-\mathrm{d}x) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u(-x)\, e^{-ix\xi} \, \mathrm{d}x = \hat{u}(-\xi)
\end{aligned}$$

which furthermore implies that we have

$$\int_{\mathbb{R}} Lv \cdot u \, \mathrm{d}x = \int_{\mathbb{R}} m(\xi)\, \hat{v}(\xi)\, \hat{u}(-\xi) \, \mathrm{d}\xi.$$

It is natural for us to consider the symmetric and anti-symmetric parts of $\hat{u}$ denoted by $\hat{u}_+$, $\hat{u}_-$ respectively such that $\hat{u}(\xi) = \hat{u}_+(\xi) + \hat{u}_-(\xi)$ where $\hat{u}_+(-\xi) = \hat{u}_+(\xi)$ and $\hat{u}_-(-\xi) = -\hat{u}_-(\xi)$ for all ξ. Then

$$\int_{\mathbb{R}} Lv \cdot u \, \mathrm{d}x = \int_{\mathbb{R}} m(\xi)\, \hat{v}(\xi)\, (\hat{u}_+(\xi) - \hat{u}_-(\xi)) \, \mathrm{d}\xi.$$

Considering the inner product on $L^2(\mathbb{R})$ for real-valued functions u, v, we have by Plancherel's theorem (Theorem 5.4)

$$\langle \hat{v}, m\hat{u}_\pm \rangle_{L^2(\mathbb{R})} = \int_{\mathbb{R}} \overline{m(\xi)\, \hat{u}_\pm(\xi)}\, \hat{v}(\xi) \, \mathrm{d}\xi = \int_{\mathbb{R}} \overline{(Lu_\pm)(x)}\, v(x) \, \mathrm{d}x = \langle v, (Lu_\pm) \rangle_{L^2(\mathbb{R})}.$$

By Theorem 5.2 we know that when u is a real-valued function the anti-symmetric (odd) part u_- has an imaginary and odd Fourier transform, whereas the symmetric (even) part u_+ has a real and even Fourier transform, meaning that we have

$$\int_{\mathbb{R}} (Lv) \cdot u \,\mathrm{d}x = \int_{\mathbb{R}} (Lu_+ + Lu_-)\, v \,\mathrm{d}x = \int_{\mathbb{R}} (Lu) \cdot v \,\mathrm{d}x$$

thus establishing the result for real-valued functions u, v and real-valued multiplier $m(\xi)$. Generalizing to complex-valued functions is then straightforward given that complex-valued functions u may be represented as $u = \Re(u)+i\Im(u)$ where $\Re(u), \Im(u)$ are real-valued functions. Consider now complex-valued u, v and the same Hamiltonian as before

$$\mathrm{d}H_m(u; v) = \int_{\mathbb{R}} (Lv) \cdot u + (Lu) \cdot v \,\mathrm{d}x$$

which due to our result for real-valued functions gives us

$$\begin{aligned}
&\int_{\mathbb{R}} (Lv) \cdot u \,\mathrm{d}x = \int_{\mathbb{R}} (L\Re(v) + iL\Im(v))(\Re(u) + i\Im(u)) \,\mathrm{d}x \\
&= \int_{\mathbb{R}} (L\Re(v)) \cdot \Re(u) + i(L\Im(v)) \cdot \Re(u) + i(L\Re(v)) \cdot \Im(u) - (L\Im(v)) \cdot \Im(u) \,\mathrm{d}x \\
&= \int_{\mathbb{R}} (L\Re(u)) \cdot \Re(v) + i(L\Re(u)) \cdot \Im(v) + i(L\Im(u)) \cdot \Re(v) - (L\Im(u)) \cdot \Im(v) \,\mathrm{d}x \\
&= \int_{\mathbb{R}} (L\Re(u) + iL\Im(u))(\Re(v) + i\Im(v)) \,\mathrm{d}x = \int_{\mathbb{R}} (Lu) \cdot v \,\mathrm{d}x.
\end{aligned}$$

This shows that our Hamiltonian H_m works for complex-valued functions as well. Now we examine complex-valued multipliers expresssed as $m(\xi) = \Re(m)(\xi)+i\Im(m)(\xi)$ with real-valued functions u, v in the expression

$$\begin{aligned}
&\int_{\mathbb{R}} m(\xi)\, \overline{\hat{u}_\pm(\xi)} \hat{v}(\xi) \,\mathrm{d}\xi = \int_{\mathbb{R}} (\Re(m)(\xi) + i\Im(m)(\xi))\overline{\hat{u}_\pm(\xi)} \hat{v}(\xi) \,\mathrm{d}\xi \\
&= \int_{\mathbb{R}} \Re(m)(\xi)\, \overline{\hat{u}_\pm(\xi)}\, \hat{v}(\xi) \,\mathrm{d}\xi + i \int_{\mathbb{R}} \Im(m)(\xi)\, \overline{\hat{u}_\pm(\xi)}\, \hat{v}(\xi) \,\mathrm{d}\xi \\
&= \int_{\mathbb{R}} \overline{\Re(m)(\xi)\hat{u}_\pm(\xi)}\, \hat{v}(\xi) \,\mathrm{d}\xi + (2i - i) \int_{\mathbb{R}} \overline{\Im(m)(\xi)\hat{u}_\pm(\xi)}\, \hat{v}(\xi) \,\mathrm{d}\xi \\
&= \int_{\mathbb{R}} \overline{m(\xi)\hat{u}_\pm(\xi)}\, \hat{v}(\xi) \,\mathrm{d}\xi + 2i \int_{\mathbb{R}} \overline{\Im(m)(\xi)\hat{u}_\pm(\xi)}\, \hat{v}(\xi) \,\mathrm{d}\xi \\
&= \int_{\mathbb{R}} \overline{Lu_\pm(x)}\, v(x) \,\mathrm{d}x + 2i \int_{\mathbb{R}} \mathscr{F}^{-1}(\Im(m)\hat{u}_\pm)(x)\, v(x) \,\mathrm{d}x
\end{aligned}$$

where the last equality is due to Plancherel's theorem. Note that

$$\overline{Lu_\pm(x)} = \mathscr{F}^{-1}(\Re(m)\hat{u}_\pm) - i\mathscr{F}^{-1}(\Im(m)\hat{u}_\pm)$$

since our function $u = u_+ + u_-$ was assumed to be real. Ultimately this means that

$$\int_{\mathbb{R}} m(\xi)\overline{\hat{u}_\pm(\xi)}\hat{v}(\xi) \,\mathrm{d}\xi = \int_{\mathbb{R}} Lu_\pm(x)\, v(x) \,\mathrm{d}x$$

which furthermore implies that

$$\begin{aligned}\int_{\mathbb{R}} (Lv) \cdot u \,\mathrm{d}x &= \int_{\mathbb{R}} m(\xi)\,(\hat{u}_+(\xi) - \hat{u}_-(\xi))\hat{v}(\xi)\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}} m(\xi)\,(\overline{\hat{u}_+(\xi)} + \overline{\hat{u}_-(\xi)})\hat{v}(\xi)\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}} (Lu_+(x) + Lu_-(x))\,v(x)\,\mathrm{d}x = \int_{\mathbb{R}} (Lu) \cdot v \,\mathrm{d}x.\end{aligned}$$

Complex-valued multipliers $m(\xi)$ are then permissible given that u, v are real-valued functions. The proof of the case where the Fourier multiplier and functions u, v are complex is analogous to the calculation based on the separation into real and imaginary parts as before. □

7.2. **Natural Spaces for the Symplectic Formalism.**
Why did we consider the Schwartz space or Sobolev spaces for our discussion of symplecticity on function spaces? It turns out that our options are limited, and few are so natural as these "special" spaces in our context. We begin this section by reviewing some aspects of topology on $\mathbb{R}^n$.

By the Heine-Borel theorem (cf. Munkres [8]), *compact subsets* of $\mathbb{R}^n$ are precisely the subsets which are bounded and closed in $\mathbb{R}^n$ with the standard metric topology. *Connected subsets* of $\mathbb{R}^n$ are subsets which cannot be written non-trivially as disjoint unions of open subsets of $\mathbb{R}^n$.

Example 7.3. Considering the real line we have the following examples:

- The subset $[a, b] \subset \mathbb{R}$ for finite $a, b \in \mathbb{R}$ is both compact and connected.
- The subset $[a, b) \subset \mathbb{R}$ for any $a, b \in \mathbb{R}$ is not compact since it is not closed, but it is connected.
- The subset $[0, 1] \cup [2, 3] \subset \mathbb{R}$ is compact but not connected.

△

The *support* of a function $f\colon X \subseteq \mathbb{R}^n \to \mathbb{R}$ is the set

$$\operatorname{supp}(f) = \overline{\{\, x \in X \mid f(x) \neq 0 \,\}}$$

where the bar denotes the closure of the set. A function $f\colon X \to \mathbb{R}$ with compact support is said to be compactly supported on X. A continuous function compactly supported on X will always have a finite integral over X because of the extreme value theorem.

The reader is encouraged to try to think of symplectic forms on various function spaces besides the one we have discussed in the previous sub-section. This turns out to be a difficult task, especially considering that we wish to connect the symplectic forms to some physical quantities. Consider the following integral form K for $\Omega \subset \mathbb{R}$ connected and compact, L a linear operator:

$$K(f, g) = \int_{\Omega} (Lf) \cdot g \,\mathrm{d}x$$

We easily see that K is bilinear, although it is not skew-symmetric a priori. The condition we require is then that

$$\int_\Omega (Lf) \cdot g \,\mathrm{d}x = -\int_\Omega f \cdot (Lg) \,\mathrm{d}x.$$

No natural choice of operator L possesses the property that we obtain a sign change when permuting the functions in the integrand, so we rely integration by parts which granted skew-symmetry for the symplectic form on the Sobolev spaces. Let L be the differential operator ∂_x. Then we have for $\Omega = [a, b]$

$$\int_\Omega (\partial_x f) \cdot g \,\mathrm{d}x = f \cdot g\big|_a^b - \int_\Omega f \cdot (\partial_x g) \,\mathrm{d}x$$

where we need the first term on the right hand side to vanish in order for the bilinear function to be skew-symmetric. Granted that Ω is compact and connected this condition is met. Notice however that $\partial_x f = 0$ whenever f is constant. This violates our non-degeneracy, so a more appropriate linear operator is $L = \partial_x^{-1}$ whenever this is defined in a meaningful way. In addition, we need to have a unique inverted counterpart to L so that we may solve for symplectic gradients.

All of these considerations point toward the Sobolev spaces as being the natural spaces for considering symplecticity on function spaces. The homogeneous Sobolev spaces are especially natural since one can define equivalence relations between functions up to additive constants, such that $f \equiv g \iff f = g + c$ for some constant function c. The functions and their derivatives ought to decay rapidly so that we may use integration by parts and have the bothersome term vanish. Keep in mind that we think of the Schwartz space as being embedded in a given Sobolev space, so that our discussion above applies to the Schwartz space as well.

8. Further Reading

The book "Applications of Lie Groups to Differential Equations" by Peter J. Olver [9] further explores how one may use Lie groups, smooth manifolds which are also groups, to analyze PDEs. Our exposition to manifold theory is good enough to get a grasp on the beginning chapters of this book.

A good beginner's exposition to PDEs is Borthwick's "Introduction to Partial Differential Equations" [2]. There are many more aspects to the analysis of PDEs than indicated in this article, and Borthwick gives an undergraduate level invitation to the field.

The reader should also be aware that Tu's book "Introduction to Manifolds" [10] contains the generalized, nicer versions of Stokes', Green's and Gauss' theorems from multivariable calculus, and that this article more or less covers the prerequisites for understanding the final subchapters pertaining this theory.

Acknowledgements

The author wishes to thank Professor Mats Ehrnström (NTNU) and Gabriele Brüll (Karlsruhe Institute of Technology) for their inclusiveness and introducing a hopeful student to the mathematical community, and Glen M. Wilson (Rutgers University) for his lectures on differential topology and insight regarding the equivalences of tangent spaces. This StudForsk-project is supported by the Olav Thon foundation.

References

[1] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag New York, 2nd edition, 1989.

[2] D. Borthwick. *Introduction to Partial Differential Equations*. Universitext. Springer International Publishing, 2018.

[3] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2011.

[4] H. Cohn. Why symplectic geometry is the natural setting for classical mechanics. URL: `http://math.mit.edu/~cohn/Thoughts/symplectic.html`. (Retrieved July 6th 2019).

[5] L. C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, 2nd edition, 2010.

[6] J. B. Fraleigh. *A First Course in Abstract Algebra*. Pearson Education Limited, 7th edition, 2013.

[7] S. Lang. *Differential Manifolds*. Springer-Verlag New York, 2nd edition, 1985.

[8] J. Munkres. *Topology*. Pearson Education Limited, 2nd edition, 2014.

[9] P. J. Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag New York, 1986.

[10] L. W. Tu. *Introduction to Manifolds*. Universitext. Springer-Verlag New York Inc., 2nd edition, 2011.

[11] A. Weinstein. Symplectic structures on banach manifolds. *Bull. Amer. Math. Soc.*, 75:1040–1041, 1969.