Differential Geometry, Analysis and Physics

Jeffrey M. Lee

Contents

0.1 Preface

In this book I present differential geometry and related mathematical topics with the help of examples from physics. It is well known that there is something strikingly mathematical about the physical universe as it is conceived of in the physical sciences. The convergence of physics with mathematics, especially differential geometry, topology and global analysis is even more pronounced in the newer quantum theories such as gauge field theory and string theory. The amount of mathematical sophistication required for a good understanding of modern physics is astounding. On the other hand, the philosophy of this book is that mathematics itself is illuminated by physics and physical thinking.

The ideal of a truth that transcends all interpretation is perhaps unattainable. Even the two most impressively objective realities, the physical and the mathematical, are still only approachable through, and are ultimately inseparable from, our normative and linguistic background. And yet it is exactly the tendency of these two sciences to point beyond themselves to something transcendentally real that so inspires us. Whenever we interpret something real, whether physical or mathematical, there will be those aspects which arise as mere artifacts of our current descriptive scheme and those aspects that seem to be objective realities which are revealed equally well through any of a multitude of equivalent descriptive schemes-"cognitive inertial frames" as it were. This theme is played out even within geometry itself where a viewpoint or interpretive scheme translates to the notion of a coordinate system on a differentiable manifold.

A physicist has no trouble believing that a vector field is something beyond its representation in any particular coordinate system since the vector field itself is something physical. It is the way that the various coordinate descriptions relate to each other (covariance) that manifests to the understanding the presence of an invariant physical reality. This seems to be very much how human perception works and it is interesting that the language of tensors has shown up in the cognitive science literature. On the other hand, there is a similar idea as to what should count as a geometric reality. According to Felix Klein the task of geometry is

> "given a manifold and a group of transformations of the manifold, to study the manifold configurations with respect to those features which are not altered by the transformations of the group"
> -Felix Klein 1893

The geometric is then that which is invariant under the action of the group. As a simple example we may consider the set of points on a plane. We may impose one of an infinite number of rectangular coordinate systems on the plane. If, in one such coordinate system (x, y), two points P and Q have coordinates $(x(P), y(P))$ and $(x(Q), y(Q))$ respectively, then while the differences $\Delta x = x(P) - x(Q)$ and $\Delta y = y(P) - y(Q)$ are very much dependent on the choice of these rectangular coordinates, the quantity $(\Delta x)^2 + (\Delta y)^2$ is *not* so dependent.

If (X, Y) are any other set of rectangular coordinates then we have $(\Delta x)^2 + (\Delta y)^2 = (\Delta X)^2 + (\Delta Y)^2$. Thus we have the intuition that there is something more real about that later quantity. Similarly, there exists distinguished systems for assigning three spatial coordinates (x, y, z) and a single temporal coordinate t to any simple event in the physical world as conceived of in relativity theory. These are called inertial coordinate systems. Now according to special relativity the invariant relational quantity that exists between any two events is $(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (\Delta t)^2$. We see that there is a similarity between the physical notion of the objective event and the abstract notion of geometric point. And yet the minus sign presents some conceptual challenges.

While the invariance under a group action approach to geometry is powerful it is becoming clear to many researchers that the looser notions of groupoid and pseudogroup has a significant role to play.

Since physical thinking and geometric thinking are so similar, and even at times identical, it should not seem strange that we not only understand the physical through mathematical thinking but conversely we gain better mathematical understanding by a kind of physical thinking. Seeing differential geometry applied to physics actually helps one understand geometric mathematics better. Physics even inspires purely mathematical questions for research. An example of this is the various mathematical topics that center around the notion of quantization. There are interesting mathematical questions that arise when one starts thinking about the connections between a quantum system and its classical analogue. In some sense, the study of the Laplace operator on a differentiable manifold and its spectrum is a "quantized version" of the study of the geodesic flow and the whole Riemannian apparatus; curvature, volume, and so forth. This is not the definitive interpretation of what a quantized geometry should be and there are many areas of mathematical research that seem to be related to the physical notions of quantum verses classical. It comes as a surprise to some that the uncertainty principle is a completely mathematical notion within the purview of harmonic analysis. Given a specific context in harmonic analysis or spectral theory, one may actually prove the uncertainty principle. Physical intuition may help even if one is studying a "toy physical system" that doesn't exist in nature or only exists as an approximation (e.g. a nonrelativistic quantum mechanical system). At the very least, physical thinking inspires good mathematics.

I have purposely allowed some redundancy to occur in the presentation because I believe that important ideas should be repeated.

Finally we mention that for those readers who have not seen any physics for a while we put a short and extremely incomplete overview of physics in an appendix. The only purpose of this appendix is to provide a sort of warm up which might serve to jog the readers memory of a few forgotten bits of undergraduate level physics.

Chapter 1

Preliminaries and Local Theory

I never could make out what those damn dots meant.

Lord Randolph Churchill

Differential geometry is one of the subjects where notation is a continual problem. Notation that is highly precise from the vantage point of set theory and logic tends to be fairly opaque with respect to the underlying geometric intent. On the other hand, notation that is uncluttered and handy for calculations tends to suffer from ambiguities when looked at under the microscope as it were. It is perhaps worth pointing out that the kind of ambiguities we are talking about are accepted by every calculus student without much thought. For instance, we find (x, y, z) being used to refer variously to "indeterminates", "a triple of numbers", or functions of some variable as when we write

$$\vec{x}(t) = (x(t), y(t), z(t)).$$

Also, we often write $y = f(x)$ and then even write $y = y(x)$ and $y'(x)$ or dy/dx which have apparent ambiguities. This does not mean that this notation is bad. In fact, it can be quite useful to use slightly ambiguous notation. In fact, human beings are generally very good at handling ambiguity and it is only when the self conscious desire to avoid logical inconsistency is given priority over everything else do we begin to have problems. The reader should be warned that while we will develop fairly pedantic notation we shall also not hesitate to resort to abbreviation and notational shortcuts as the need arises (and with increasing frequency in later chapters).

The following is short list of notational conventions:

Category	**Sets**	**Elements**	**Maps**
Vector Spaces	V , W	v,w, x, y	A, B, λ, L
Topological vector spaces (TVS)	E,V,W,$\mathbb{R}^n$	v,w,x,y	A, B, λ, L
Open sets in TVS	U, V, O, U_α	p,x,y,v,w	f, g, φ, ψ
Lie Groups	G, H, K	g, h, x, y	h, f, g
The real (resp. complex) numbers	$\mathbb{R}$, (resp.$\mathbb{C}$)	$t, s, x, y, z,$	f, g, h
One of $\mathbb{R}$ or $\mathbb{C}$	$\mathbb{F}$	”‘	‘’

A more complete chart may be found at the end of the book.

The reader is reminded that for two sets A and B the **Cartesian product** $A \times B$ is the set of pairs $A \times B := \{(a, b) : a \in A, b \in B\}$. More generally, $\prod_i A_i := \{(a_i) : a_i \in A_i\}$.

Notation 1.1 *Here and throughout the book the symbol combination ":= " means "equal by definition".*

In particular, $\mathbb{R}^n := \mathbb{R} \times \cdots \times \mathbb{R}$ the product of -copies of the real numbers $\mathbb{R}$. Whenever we represent a linear transformation by a matrix, then the matrix acts on column vectors from the left. This means that in this context elements of $\mathbb{R}^n$ are thought of as column vectors. It is sometimes convenient to represent elements of the dual space $(\mathbb{R}^n)^*$ as row vectors so that if $\alpha \in (\mathbb{R}^n)^*$ is represented by $(a_1, ..., a_n)$ and $v \in \mathbb{R}^n$ is represented by $\left(v^1, ..., v^n\right)^t$ then

$$\alpha(v) = (a_1.....a_n) \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}.$$

Since we do not want to have to write the transpose symbol in every instance and for various other reasons we will sometimes use upper indices (superscripts) to label the component entries of elements of $\mathbb{R}^n$ and lower indices (subscripts) to label the component entries of elements of $(\mathbb{R}^n)^*$. Thus $\left(v^1, ..., v^n\right)$ invites one to think of a column vector (*even when the transpose symbol is not present)* while $(a_1, ..., a_n)$ is a row vector. On the other hand, a list of elements of a vector space such as a basis will be labelled using subscripts while superscripts label lists of elements of the dual space of the initially introduced space.

1.1 Calculus

Let V be a finite dimensional vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. Occasionally the quaternion number algebra $\mathbb{H}$ (a skew field) will be considered. Each of these spaces has a conjugation map which we take to be the identity map for $\mathbb{R}$ while for $\mathbb{C}$ and $\mathbb{H}$ we have

$$x + yi \mapsto x - yi \quad \text{and}$$
$$x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k} \mapsto x - y\mathbf{i} - z\mathbf{j} - w\mathbf{k}$$

respectively. The vector spaces that we are going to be dealing with will serve as local models for the global theory. For the most part the vector spaces that serve as models will be isomorphic to $\mathbb{F}^n$ which is the set of $n-$tuples $(x^1, ..., x^n)$ of elements of $\mathbb{F}$. However, we shall take a slightly unorthodox step of introducing notation that exibits the variety of guises that the space $\mathbb{F}^n$ may appear in. The point is that althought these spaces are isomorphic to $\mathbb{F}^n$ some might have different interpretations as say matrices, bilinear forms, tensors, and so on. We have the following spaces:

1. $\mathbb{F}^n$ which is the set of n-tuples of elements of $\mathbb{F}$ which we choose to think of as column vectors. These are written as $(x^1, ..., x^n)^t$ or more commonly simply as $(x^1, ..., x^n)$ where the fact that we have place the indices as superscripts is enough to remind us that in matrix multiplication these are supposed to be columns. The standard basis for this vector space is $\{e_i\}_{1\leq i\leq n}$ where e_i has all zero entries except for a single 1 in the i-th position. The indices have been lowered on purpose to facilitate the expression like $\mathsf{v} = \sum_i v^i \mathsf{e}_i$. The appearence of a repeated index one of which is a subscript while the other a superscript, signals a summation. Acording to the Einstien summatation convention we may omit the $\sum_i$ and write $v = v^i \mathsf{e}_i$ the summation being implied.

2. $\mathbb{F}_n$ is the set n-tuples of elements of $\mathbb{F}$ thought of as row vectors. The elements are written $(\xi_1, ..., \xi_n)$. This space is often identified with the dual of $\mathbb{F}^n$ where the pairing becomes matrix multiplication $\langle \xi, \mathsf{v}\rangle = \xi \mathsf{w}$. Of course, we may also take $\mathbb{F}^n$ to be its own dual but then we must write $\langle \mathsf{v}, \mathsf{w}\rangle = \mathsf{v}^t\mathsf{w}$.

3. $\mathbb{F}^n_m$ is just the set of $m \times n$ matrices, also written $M_{m\times n}$. The elements are written as (x^i_j). The standard basis is $\{\mathsf{e}^j_i\}$ so that $\mathsf{A} = (A^i_j) = \sum_{i,j} A^i_j \mathsf{e}^j_i$

4. $\mathbb{F}_{m,n}$ is also the set of $m \times n$ matrices but the elements are written as (x_{ij}) and these are thought of as giving maps from $\mathbb{F}^n$ to $\mathbb{F}_n$ as in $(v^i) \mapsto (w_i)$ where $w_i = \sum_j x_{ij} v^j$. A particularly interesting example is when the $\mathbb{F} = \mathbb{R}$. Then if (g_{ij}) is a symmetic positive definite matrix since then we get an isomorphism $g : \mathbb{F}^n \cong \mathbb{F}_n$ given by $v_i = \sum_j g_{ij} v^j$. This provides us with an inner product on $\mathbb{F}_n$ given by $\sum_j g_{ij} w^i v^j$ and the usual choice for g_{ij} is $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Using δ_{ij} makes the standard basis on $\mathbb{F}^n$ an orthonormal basis.

5. $\mathbb{F}^{\mathcal{I}}_{\mathcal{J}}$ is the set of all elements of $\mathbb{F}$ indexed as x^I_J where $I \in \mathcal{I}$ and $I \in \mathcal{J}$ for some indexing sets $\mathcal{I}$ and $\mathcal{J}$. The dimension of $\mathbb{F}^{\mathcal{I}}_{\mathcal{J}}$ is the cardinality of $\mathcal{I} \times \mathcal{J}$. To look ahead a bit, this last notation comes in handy since it allows us to reduce a monster like

$$c^{i_1 i_2 \ldots i_r} = \sum_{k_1,k_2,\ldots,k_m} a^{i_1 i_2 \ldots i_r}_{k_1 k_2 \ldots k_m} b^{k_1 k_2 \ldots k_m}$$

to something like

$$c^I = \sum_K a^I_K b^K$$

which is more of a “cookie monster”.

In every case of interest V has a natural topology that is determined by a norm. For example the space $\mathbb{F}^{\mathcal{I}}$ ($:= \mathbb{F}^{\mathcal{I}}_{\emptyset}$) has an inner product $\langle \mathsf{v}_1, \mathsf{v}_2 \rangle = \sum_{I\in\mathcal{I}} a^I \bar{b}^I$ where $\mathsf{v}_1 = \sum_{I\in\mathcal{I}} a^I e_I$ and $\mathsf{v}_2 = \sum_{I\in\mathcal{I}} b^I e_I$. The inner product gives the norm in the usual way $|\mathsf{v}| := \langle \mathsf{v}, \mathsf{v}\rangle^{1/2}$ which determines a topology on V. Under this topology all the vector space operations are continuous. Futhermore, all norms give the same topology on a finite dimensional vector space.

§§ Interlude §§
Infinite dimensions.

What about infinite dimensional spaces? Are there any “standard” spaces in the infinite dimensional case? Well, there are a few problems that must be addresses if one want to include infinite dimensional spaces. We will not systematically treat infinite dimensioanl manifold theory but calculas on infinite dimensional spaces can be fairly nice if one restricts to complete normed spaces (Banach spaces). As a sort of warm up let us step through a progressively ambitious attemp to generalize the above spaces to infinite dimensions.

1. We could just base our generalization on the observation that an element (x^i) of $\mathbb{F}^n$ isreally just a function $x : \{1, 2, ..., n\} \to \mathbb{F}$ so maybe we should consider instead the index set $\mathbb{N} = \{1, 2, 3, ... \to \infty\}$. This is fine except that we must interpret the sums like $v = v^i \mathsf{e}_i$. The reader will no doubt realize that one possible solution is to restrict to ∞-tuple (sequences) that are in ℓ^2. This the the Hilbert space of square summable sequences.This one works out very nicely although there are some things to be concerned about. We could then also consider spaces of matrices with the rows and columns infinite but square summable these provide operetors $\ell^2 \to \ell^2$. But should we restrict to trace class operators? Eventually we get to tensors where which would have to be indexed “tuples” like (Υ^{rs}_{ijk}) which are square summable in the sense that $\sum (\Upsilon^{rs}_{ijk})^2 < \infty$.

2. Maybe we could just replace the indexing sets by subsets of the plane or even some nice measure space Ω. Then our elements would just be functions and imediately we see that we will need measurable functions. We must also find a topology that will be suitable for defining the limits we will need when we define the derivative below. The first possiblity is to restrict to the square integrable functions. In other words, we could try to do everything with the Hilbert space $L^2(\Omega)$. Now what should the standard basis be? OK, now we are starting to get in trouble it seems. But do we really need a standard basis?

3. It turns out that all one needs to do calculus on the space is for it to be a sufficiently nice topological vector space. The common choice of a Banach space is so that the proof of the inverse function theorem goes through (but see [KM]).

   The reader is encouraged to look at the appendices for a more formal treatment of infinite dimensional vector spaces.

§§

Next we record those parts of calculus that will be most important to our study of differentiable manifolds. In this development of calculus the vector spaces are of one of the finite dimensional examples given above and we shall refer to them generically as Euclidean spaces. For convenience we will restrict our attention to vector spaces over the real numbers $\mathbb{R}$. Each of these "Euclidean" vector spaces has a norm where the norm of x denoted by $\|\mathsf{x}\|$. On the other hand, with only minor changes in the proofs, everything works for Banach spaces. In fact, we have put the proofs in an appendix where the spaces are indeed taken to be general Banach spaces.

Definition 1.1 *Let* V *and* W *be Euclidean vector spaces as above (for example* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*). Let* U *an open subset of* V*. A map* $f : U \to \mathsf{W}$ *is said to be differentiable at* $\mathsf{x} \in U$ *if and only if there is a (necessarily unique) linear map* $Df|_p : \mathsf{V} \to \mathsf{W}$ *such that*

$$\lim_{|\mathsf{x}| \to 0} \frac{\left\| f(\mathsf{x}+\mathsf{v}) - f(\mathsf{x}) - Df|_p \,\mathsf{v} \right\|}{\|\mathsf{x}\|}$$

Notation 1.2 *We will denote the set of all linear maps from* V *to* W *by* $L(\mathsf{V},\mathsf{W})$*. The set of all linear isomorphisms from* V *onto* W *will be denoted by* $\mathrm{GL}(\mathsf{V},\mathsf{W})$*. In case,* $\mathsf{V} = \mathsf{W}$ *the corresponding spaces will be denoted by* $\mathfrak{gl}(\mathsf{V})$ *and* $\mathrm{GL}(\mathsf{W})$*. For linear maps* $T : \mathsf{V} \to \mathsf{W}$ *we sometimes write* $T \cdot \mathsf{v}$ *instead of* $T(\mathsf{v})$ *depending on the notational needs of the moment. In fact, a particularly useful notational device is the following: Suppose we have map* $A : X \to L(\mathsf{V};\mathsf{W})$*. Then we would write* $A(x)\cdot \mathsf{v}$ *or* $A|_x \,\mathsf{v}$*.*

Here $\mathrm{GL}(\mathsf{V})$ is a group under composition and is called the general linear group . In particular, $\mathrm{GL}(\mathsf{V},\mathsf{W})$ is a subset of $L(\mathsf{V},\mathsf{W})$ but not a linear subspace.

Definition 1.2 *Let* V_i*,* $i = 1, ..., k$ *and* W *be finite dimensional* $\mathbb{F}$*-vector spaces. A map* $\mu : V_1 \times \cdots \times V_k \to \mathsf{W}$ *is called* ***multilinear*** *(k-multilinear) if for each* i*,* $1 \le i \le k$ *and each fixed* $(\mathsf{w}_1, ..., \widehat{\mathsf{w}_i}, ..., \mathsf{w}_k) \in V_1 \times \cdots \times \widehat{V_1} \times \cdots \times V_k$ *we have that the map*

$$\mathsf{v} \mapsto \mu(\mathsf{w}_1, ..., \underset{i-th}{\mathsf{v}}, ..., \mathsf{w}_{k-1}),$$

obtained by fixing all but the i-th variable is a linear map. In other words, we require that μ be $\mathbb{F}$- linear in each slot separately.

The set of all multilinear maps $\mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ will be denoted by $L(\mathsf{V}_1, ..., \mathsf{V}_k; \mathsf{W})$. If $\mathsf{V}_1 = \cdots = \mathsf{V}_k = \mathsf{V}$ then we write $L^k(\mathsf{V}; \mathsf{W})$ instead of $L(\mathsf{V}, ..., \mathsf{V}; \mathsf{W})$

Since each vector space has a (usually obvious) inner product then we have the group of linear isometries $O(\mathsf{V})$ from V onto itself. That is, $O(\mathsf{V})$ consists of the bijective linear maps $\Phi : \mathsf{V} \to \mathsf{V}$ such that $\langle \Phi \mathsf{v}, \Phi \mathsf{w} \rangle = \langle \mathsf{v}, \mathsf{w} \rangle$ for all $\mathsf{v}, \mathsf{w} \in \mathsf{V}$. The group $O(\mathsf{V})$ is called the **orthogonal group.**

Definition 1.3 *A (bounded) multilinear map $\mu : \mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ is called **symmetric** (resp. **skew-symmetric** or **alternating**) iff for any $\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k \in \mathsf{V}$ we have that*

$$\mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k) = K(\mathsf{v}_{\sigma 1}, \mathsf{v}_{\sigma 2}, ..., \mathsf{v}_{\sigma k})$$
$$\textit{resp. } \mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k) = sgn(\sigma)\mu(\mathsf{v}_{\sigma 1}, \mathsf{v}_{\sigma 2}, ..., \mathsf{v}_{\sigma k})$$

for all permutations σ on the letters $\{1, 2, ..., k\}$. The set of all bounded symmetric (resp. skew-symmetric) multilinear maps $\mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ is denoted $L^k_{sym}(\mathsf{V}; \mathsf{W})$ (resp. $L^k_{skew}(\mathsf{V}; \mathsf{W})$ or $L^k_{alt}(\mathsf{V}; \mathsf{W})$).

Now the space $L(\mathsf{V}, \mathsf{W})$ is a normed space with the norm

$$\|l\| = \sup_{\mathsf{v} \in \mathsf{V}} \frac{\|l(\mathsf{v})\|_{\mathsf{W}}}{\|\mathsf{v}\|_{\mathsf{V}}} = \sup\{\|l(\mathsf{v})\|_{\mathsf{W}} : \|\mathsf{v}\|_{\mathsf{V}} = 1\}.$$

The spaces $L(\mathsf{V}_1, ..., \mathsf{V}_k; \mathsf{W})$ also have norms given by

$$\|\mu\| := \sup\{\|\mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k)\|_{\mathsf{W}} : \|\mathsf{v}_i\|_{\mathsf{V}_i} = 1 \text{ for } i = 1, .., k\}$$

Notation 1.3 *In the context of $\mathbb{R}^n$, we often use the so called "multiindex notation". Let $\alpha = (\alpha_1, ..., \alpha_n)$ where the α_i are integers and $0 \le \alpha_i \le n$. Such an n-tuple is called a multiindex. Let $|\alpha| := \alpha_1 + ... + \alpha_n$ and*

$$\frac{\partial^\alpha f}{\partial x^\alpha} := \frac{\partial^{|\alpha|} f}{\partial (x^1)^{\alpha_1} \partial (x^1)^{\alpha_2} \cdots \partial (x^1)^{\alpha_n}}.$$

Proposition 1.1 *There is a natural linear isomorphism $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}, \mathsf{W})$ given by*

$$l(v_1)(v_2) \longleftrightarrow l(v_1, v_2)$$

and we identify the two spaces. In fact, $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^3(\mathsf{V}; \mathsf{W})$ and in general $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, ..., L(\mathsf{V}, \mathsf{W})) \cong L^k(\mathsf{V}; \mathsf{W})$ etc.

Proof. It is easily checked that if we just define $(\iota\, T)(v_1)(v_2) = T(v_1, v_2)$ then $\iota\, T \leftrightarrow T$ does the job for the $k = 2$ case. The $k > 2$ case can be done by an inductive construction and is left as an exercise. It is also not hard to show that the isomorphism norm preserving. ∎

Definition 1.4 *If it happens that a function f is differentiable for all p throughout some open set U then we say that f is differentiable on U. We then have a map $Df : U \subset \mathsf{V} \to L(\mathsf{V}, \mathsf{W})$ given by $\mathsf{p} \mapsto Df(\mathsf{p})$. If this map is differentiable at some $\mathsf{p} \in \mathsf{V}$ then its derivative at p is denoted $DDf(\mathsf{p}) = D^2f(\mathsf{p})$ or $D^2f\big|_{\mathsf{p}}$ and is an element of $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}; \mathsf{W})$. Similarly, we may inductively define $D^kf \in L^k(\mathsf{V}; \mathsf{W})$ whenever f is sufficiently nice that the process can continue.*

Definition 1.5 *We say that a map $f : U \subset \mathsf{V} \to \mathsf{W}$ is C^r-differentiable on U if $D^rf|_{\mathsf{p}} \in L^r(\mathsf{V}, \mathsf{W})$ exists for all $\mathsf{p} \in U$ and if D^rf is continuous as map $U \to L^r(\mathsf{V}, \mathsf{W})$. If f is C^r-differentiable on U for all $r > 0$ then we say that f is C^∞ or **smooth** (on U).*

Definition 1.6 *A bijection f between open sets $U_\alpha \subset \mathsf{V}$ and $U_\beta \subset \mathsf{W}$ is called a C^r-**diffeomorphism** iff f and f^{-1} are both C^r-differentiable (on U_α and U_β respectively). If $r = \infty$ then we simply call f a diffeomorphism. Often, we will have $\mathsf{W} = \mathsf{V}$ in this situation.*

*Let U be open in V. A map $f : U \to \mathsf{W}$ is called a **local** C^r **diffeomorphism** iff for every $\mathsf{p} \in U$ there is an open set $U_{\mathsf{p}} \subset U$ with $\mathsf{p} \in U_{\mathsf{p}}$ such that $f|_{U_{\mathsf{p}}} : U_{\mathsf{p}} \to f(U_{\mathsf{p}})$ is a C^r-diffeomorphism.*

In the context of undergraduate calculus courses we are used to thinking of the derivative of a function at some $a \in \mathbb{R}$ as a number $f'(a)$ which is the slope of the tangent line on the graph at $(a, f(a))$. From the current point of view $Df(a) = Df|_a$ just gives the linear transformation $h \mapsto f'(a) \cdot h$ and the equation of the tangent line is given by $y = f(a) + f'(a)(x - a)$. This generalizes to an arbitrary differentiable map as $\mathsf{y} = f(\mathsf{a}) + Df(\mathsf{a}) \cdot (\mathsf{x} - \mathsf{a})$ giving a map which is the linear approximation of f at a.

We will sometimes think of the derivative of a curve[1] $c : I \subset \mathbb{R} \to \mathsf{E}$ at $t_0 \in I$, written $\dot{c}(t_0)$, as a velocity vector and so we are identifying $\dot{c}(t_0) \in L(\mathbb{R}, \mathsf{E})$ with $Dc|_{t_0} \cdot 1 \in \mathsf{E}$. Here the number 1 is playing the role of the unit vector in $\mathbb{R}$.

Let $f : U \subset \mathsf{E} \to \mathsf{F}$ be a map and suppose that we have a splitting $\mathsf{E} = \mathsf{E}_1 \times \mathsf{E}_2 \times \cdots \mathsf{E}_n$ for example . We will write $f(\mathsf{x}^1, ..., \mathsf{x}^n)$ for $(\mathsf{x}^1, ..., \mathsf{x}^n) \in \mathsf{E}_1 \times \mathsf{E}_2 \times \cdots \mathsf{E}_n$. Now for every $\mathsf{a} = (\mathsf{a}^1, ..., \mathsf{a}^n) \in \mathsf{E}_1 \times \cdots \times \mathsf{E}_n$ we have the partial map $f_{\mathsf{a},i} : \mathsf{y} \mapsto f(\mathsf{a}^1, ..., \mathsf{y}, ...\mathsf{a}^n)$ where the variable y is is in the i slot. This defined in some neighborhood of a^i in E_i. We define the partial derivatives when they exist by $D_if(\mathsf{a}) = Df_{\mathsf{a},i}(\mathsf{a}^i)$. These are, of course, linear maps.

$$D_if(\mathsf{a}) : \mathsf{E}_i \to \mathsf{F}$$

The partial derivative can exist even in cases where f might not be differentiable in the sense we have defined. The point is that f might be differentiable only in certain directions.

[1] We will often use the letter I to denote a generic (usually open) interval in the real line.

If f has continuous partial derivatives $D_i f(\mathsf{x}) : \mathsf{E}_i \to \mathsf{F}$ near $\mathsf{x} \in \mathsf{E} = \mathsf{E}_1 \times \mathsf{E}_2 \times \cdots \mathsf{E}_n$ then $Df(\mathsf{x})$ exists and is continuous near p. In this case,

$$Df(\mathsf{x}) \cdot \mathrm{v}$$
$$= \sum_{i=1}^{n} D_i f(\mathsf{x}, \mathsf{y}) \cdot \mathrm{v}^i$$

where $\mathrm{v} = (\mathrm{v}^1, ..., \mathrm{v}^n)$.

§Interlude§
Thinking about derivatives in infinite dimensions

The theory of differentiable manifolds is really just an extension of calculus in a setting where, for topological reasons, we must use several coordinates systems. At any rate, once the coordinate systems are in place many endeavors reduce to advanced calculus type calculations. This is one reason that we review calculus here. However, there is another reason. Namely, we would like to introduce calculus on Banach spaces. This will allow us to give a good formulation of the variational calculus that shows up in the study of finite dimensional manifolds (the usual case). The idea is that the set of all maps of a certain type between finite dimensional manifolds often turns out to be an *infinite dimensional* manifold. We use the calculus on Banach spaces idea to define infinite dimensional differentiable manifolds which look locally like Banach spaces. All this will be explained in detail later.

As a sort of conceptual warm up, let us try to acquire a certain flexibility in the way we think about vectors. A vector as it is understood in some contexts is just an $n-$tuple of numbers which we picture either as a point in $\mathbb{R}^n$ or an arrow emanating from some such point but an $n-$tuple $(x^1, ..., x^n)$ is also a function $x : i \mapsto x(i) = x^i$ whose domain is the finite set $\{1, 2, ..., n\}$. But then why not allow the index set to be infinite, even uncountable? In doing so we replace the $n-$tuple (x^i) be the "continuous" tuple $f(x)$. We are used to the idea that something like $\sum_{i=1}^{n} x^i y^i$ should be replaced by an integral $\int f(x)g(x)dx$ when moving to these continuous tuple (functions). Another example is the replacement of matrix multiplication $\sum a^i_j v^i$ by the continuous analogue $\int a(x, y)v(y)dy$. But what would be the analogue of a vector valued functions of a vector variable? Mathematicians would just consider these to be functions or maps again but it is also traditional, especially in physics literature, to called such things functionals. An example might be an "action functional", say S, defined on a set of curves in $\mathbb{R}^3$ with a fixed interval $[t_0, t_1]$ as domain:

$$S[c] = \int_{t_0}^{t_1} L(c(t), \dot{c}(t), t)dt$$

Here, L is defined on $\mathbb{R}^3 \times \mathbb{R}^3 \times [t_0, t_1]$ but S takes a curve as an argument and this is not just composition of functions. Thus we will not write the all too common expression "$S[c(t)]$". Also, $S[c]$ denotes the value of the functional at the curve c and not the functional itself. Physicists might be annoyed with this but it really does help to avoid conceptual errors when learning the subject of calculus on function spaces (or general Banach spaces).

When one defines a directional derivative in the Euclidean space $\mathbb{R}^n := \{x{=}(x^1, ..., x^n) : x^i \in \mathbb{R}\}$ it is through the use of a difference quotient:

$$D_h f(x) := \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon h) - f(x)}{\varepsilon}$$

which is the same thing as $D_h f(x) := \frac{d}{d\varepsilon}\big|_{\varepsilon=0} f(x + \varepsilon h)$. Limits like this one make sense in any topological vector space[2]. For example, if $C([0,1])$ denotes the space of continuous functions on the interval $[0,1]$ then one may speak of functions whose arguments are elements of $C([0,1])$. Here is a simple example of such a "functional":

$$F[f] := \int_{[0,1]} f^2(x) dx$$

The use of square brackets to contain the argument is a physics tradition that serves to warn the reader that the argument is from a space of functions. Notice that this example is not a linear functional. Now given any such functional, say F, we may define the directional derivative of F at $f \in C([0,1])$ in the direction of the function $h \in C([0,1])$ to be

$$D_h F(f) := \lim_{\varepsilon \to 0} \frac{F(f + \varepsilon h) - F(f)}{\varepsilon}$$

whenever the limit exists.

Example 1.1 *Let $F[f] := \int_{[0,1]} f^2(x)dx$ as above and let $f(x) = x^3$ and $h(x) = \sin(x^4)$. Then*

$$\begin{aligned}
D_h F(f) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} F(f + \varepsilon h) - F(f) \\
&= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} F(f + \varepsilon h) \\
&= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_{[0,1]} (f(x) + \varepsilon h(x))^2 dx \\
&= 2\int_{[0,1]} f(x)h(x)dx = 2\int_0^1 x^3 \sin(\pi x^4) dx \\
&= \frac{1}{\pi}
\end{aligned}$$

Note well that h and f are functions but here they are, more importantly, "points" in a function space! What we are differentiating is F. Again, $F[f]$ is *not* a composition of functions; f is the dependent variable here.

[2] "Topological vector space" will be defined shortly.

Exercise 1.1 *See if you can make sense out of the expressions and analogies in the following chart:*

$\vec{x}$	$\rightsquigarrow$	f
$f(\vec{x})$	$\rightsquigarrow$	$F[f]$
df	$\rightsquigarrow$	δF
$\frac{\partial f}{\partial x^i}$	$\rightsquigarrow$	$\frac{\delta F}{\delta f(x)}$
$\sum \frac{\partial f}{\partial x^i} v^i$	$\rightsquigarrow$	$\int \frac{\delta F}{\delta f(x)} v(x) dx$
$\int \cdots \int f(\vec{x}) dx^1 ... dx^n$	$\rightsquigarrow$	$\int_{\rightarrow}^{\rightarrow} \int F[f](\prod_x df(x))$

Exercise 1.2 *Some of these may seem mysterious-especially the last one which still lacks a general rigorous definition that covers all the cases needed in quantum theory. Don't worry if you are not familiar with this one. The third one in the list is only mysterious because we use δ. Once we are comfortable with calculus in the Banach space setting we will see that δF just mean the same thing as dF whenever F is defined on a function space. In this context dF is a linear functional on a Banach space.*

So it seems that we can do calculus on infinite dimensional spaces. There are several subtle points that arise. For instance, there must be a topology on the space with respect to which addition and scalar multiplication are continuous. This is the meaning of topological vector space. Also, in order for the derivative to be unique the topology must be Hausdorff. But there are more things to worry about.

We are also interested in having a version of the inverse mapping theorem. It turns out that most familiar facts from calculus on $\mathbb{R}^n$ go through if we replace $\mathbb{R}^n$ by a complete normed space (see **26.21**). There are at least two issues that remain even if we restrict ourselves to Banach spaces. First, the existence of *smooth* bump functions and *smooth* partitions of unity (to be defined below) are not guaranteed. The existence of smooth bump functions and smooth partitions of unity for infinite dimensional manifolds is a case by case issue while in the finite dimensional case their existence is guaranteed . Second, there is the fact that a subspace of a Banach space is not a Banach space unless it is a closed subspace. This fact forces us to introduce the notion of a split subspace and the statements of the Banach spaces versions of several familiar theorems, including the implicit function theorem, become complicated by extra conditions concerning the need to use split (complemented) subspaces.

§§

1.2 Chain Rule, Product rule and Taylor's Theorem

Theorem 1.1 (Chain Rule) *Let U_1 and U_2 be open subsets of Euclidean spaces E_1 and E_2 respectively. Suppose we have continuous maps composing as*

$$U_1 \xrightarrow{f} U_2 \xrightarrow{g} \mathsf{E}_3$$

where E_3 is a third Euclidean space. If f is differentiable at p and g is differentiable at $f(p)$ then the composition is differentiable at p and $D(g \circ f) = Dg(f(\mathsf{p})) \circ Dg(\mathsf{p})$. In other words, if $\mathsf{v} \in \mathsf{E}_1$ then

$$D(g \circ f)|_p \cdot \mathsf{v} = Dg|_{f(p)} \cdot (Df|_p \cdot \mathsf{v}).$$

Furthermore, if $f \in C^r(U_1)$ and $g \in C^r(U_2)$ then $g \circ f \in C^r(U_1)$.

We will often use the following lemma without explicit mention when calculating:

Lemma 1.1 *Let $f : U \subset \mathsf{V} \to \mathsf{W}$ be twice differentiable at $\mathsf{x}_0 \in U \subset \mathsf{V}$ then the map $D_\mathsf{v} f : \mathsf{x} \mapsto Df(\mathsf{x}) \cdot \mathsf{v}$ is differentiable at x_0 and its derivative at x_0 is given by*

$$D(D_\mathsf{v} f)|_{\mathsf{x}_0} \cdot \mathsf{h} = D^2 f(\mathsf{x}_0)(\mathsf{h}, \mathsf{v}).$$

Theorem 1.2 *If $f : U \subset \mathsf{V} \to \mathsf{W}$ is twice differentiable on U such that $D^2 f$ is continuous, i.e. if $f \in C^2(U)$ then $D^2 f$ is symmetric:*

$$D^2 f(p)(\mathsf{w}, \mathsf{v}) = D^2 f(p)(\mathsf{v}, \mathsf{w}).$$

More generally, if $D^k f$ exists and is continuous then $D^k f\ (\mathbf{p}) \in \mathbf{L}^k_{sym}(\mathsf{V}; \mathsf{W})$.

Theorem 1.3 *Let $\varrho \in L(\mathsf{F}_1, \mathsf{F}_2; \mathsf{W})$ be a bilinear map and let $f_1 : U \subset \mathsf{E} \to \mathsf{F}_1$ and $f_2 : U \subset \mathsf{E} \to \mathsf{F}_2$ be differentiable (resp. $C^r, r \geq 1$) maps. Then the composition $\varrho(f_1, f_2)$ is differentiable (resp. $C^r, r \geq 1$) on U where $\varrho(f_1, f_2) : \mathsf{x} \mapsto \varrho(f_1(\mathsf{x}), f_2(\mathsf{x}))$. Furthermore,*

$$D\varrho|_\mathsf{x} (f_1, f_2) \cdot \mathsf{v} = \varrho(Df_1|_\mathsf{x} \cdot \mathsf{v}, f_2(\mathsf{x})) + \varrho(f_1(\mathsf{x}), Df_2|_\mathsf{x} \cdot \mathsf{v}).$$

In particular, if F is an algebra with product $\star$ and $f_1 : U \subset \mathsf{E} \to \mathsf{F}$ and $f_2 : U \subset \mathsf{E} \to \mathsf{F}$ then $f_1 \star f_2$ is defined as a function and

$$D(f_1 \star f_2) \cdot \mathsf{v} = (Df_1 \cdot \mathsf{v}) \star (f_2) + (Df_1 \cdot \mathsf{v}) \star (Df_2 \cdot \mathsf{v}).$$

1.3 Local theory of maps

Inverse Mapping Theorem

Definition 1.7 *Let E and F be Euclidean vector spaces. A map will be called a C^r* ***diffeomorphism near*** *p if there is some open set $U \subset \operatorname{dom}(f)$ containing*

p *such that* $f|_U : U \to f(U)$ *is a* C^r *diffeomorphism onto an open set* $f(U)$*. The set of all maps which are diffeomorphisms near* p *will be denoted* $\mathrm{Diff}^r_p(\mathsf{E},\mathsf{F})$*. If* f *is a* C^r *diffeomorphism near* p *for all* $\mathsf{p} \in U = \mathrm{dom}(f)$ *then we say that* f *is a* ***local*** C^r ***diffeomorphism****.*

Theorem 1.4 (Implicit Function Theorem I) *Let* $\mathsf{E}_1, \mathsf{E}_2$ *and* F *Euclidean vector spaces and let* $U \times V \subset \mathsf{E}_1 \times \mathsf{E}_2$ *be open. Let* $f : U \times V \to \mathsf{F}$ *be a* C^r *mapping such that* $f(\mathsf{x}_0, \mathsf{y}_0) = 0$*. If* $D_2 f_{(\mathsf{x}_0,\mathsf{y}_0)} : \mathsf{E}_2 \to \mathsf{F}$ *is a continuous linear isomorphism then there exists a (possibly smaller) open set* $U_0 \subset U$ *with* $\mathsf{x}_0 \in U_0$ *and unique a mapping* $g : U_0 \to V$ *with* $g(x_0) = y_0$ *such that*

$$f(\mathsf{x}, g(\mathsf{x})) = 0$$

for all $x \in U_0$*.*

Proof. Follows from the following theorem. ■

Theorem 1.5 (Implicit Function Theorem II) *Let* $\mathsf{E}_1, \mathsf{E}_2$ *and* F *be as above and* $U \times V \subset \mathsf{E}_1 \times \mathsf{E}_2$ *open. Let* $f : U \times V \to \mathsf{F}$ *be a* C^r *mapping such that* $f(\mathsf{x}_0, \mathsf{y}_0) = \mathsf{w}_0$*. If* $D_2 f(\mathsf{x}_0, \mathsf{y}_0) : \mathsf{E}_2 \to \mathsf{F}$ *is a continuous linear isomorphism then there exists (possibly smaller) open sets* $U_0 \subset U$ *and* $W_0 \subset \mathsf{F}$ *with* $\mathsf{x}_0 \in U_0$ *and* $\mathsf{w}_0 \in W_0$ *together with a unique mapping* $g : U_0 \times W_0 \to V$ *such that*

$$f(\mathsf{x}, g(\mathsf{x}, \mathsf{w})) = \mathsf{w}$$

for all $\mathsf{x} \in U_0$*. Here unique means that any other such function* h *defined on a neighborhood* $U_0' \times W_0'$ *will equal* g *on some neighborhood of* $(\mathsf{x}_0, \mathsf{w}_0)$*.*

Proof. Sketch: Let $\Psi : U \times V \to \mathsf{E}_1 \times \mathsf{F}$ be defined by $\Psi(\mathsf{x}, \mathsf{y}) = (\mathsf{x}, f(\mathsf{x}, \mathsf{y}))$. Then $D\Psi(\mathsf{x}_0, \mathsf{y}_0)$ has the operator matrix

$$\begin{bmatrix} \mathrm{id}_{E_1} & 0 \\ D_1 f(\mathsf{x}_0, \mathsf{y}_0) & D_2 f(\mathsf{x}_0, \mathsf{y}_0) \end{bmatrix}$$

which shows that $D\Psi(\mathsf{x}_0, \mathsf{y}_0)$ is an isomorphism. Thus Ψ has a unique local inverse Ψ^{-1} which we may take to be defined on a product set $U_0 \times W_0$. Now Ψ^{-1} must have the form $(\mathsf{x}, \mathsf{y}) \mapsto (\mathsf{x}, g(\mathsf{x}, \mathsf{y}))$ which means that $(\mathsf{x}, f(\mathsf{x}, g(\mathsf{x}, \mathsf{w}))) = \Psi(\mathsf{x}, g(\mathsf{x}, \mathsf{w})) = (\mathsf{x}, \mathsf{w})$. Thus $f(\mathsf{x}, g(\mathsf{x}, \mathsf{w})) = \mathsf{w}$. The fact that g is unique follows from the local uniqueness of the inverse Ψ^{-1} and is left as an exercise. ■

Let U be an open subset of V and let $I \subset \mathbb{R}$ be an open interval containing 0. A (local) time dependent vector field on U is a C^r-map $F : I \times U \to \mathsf{V}$ (where $r \geq 0$). An integral curve of F with initial value x_0 is a map c defined on an open subinterval $J \subset I$ also containing 0 such that

$$c'(t) = F(t, c(t))$$
$$c(0) = x_0$$

A local flow for F is a map $\alpha : I_0 \times U_0 \to \mathsf{V}$ such that $U_0 \subset U$ and such that the curve $\alpha_x(t) = \alpha(t, x)$ is an integral curve of F with $\alpha_x(0) = x$

If $f : U \to V$ is a map between open subsets of V and W we have the notion of **rank at** $p \in U$ which is just the rank of the linear map $D_p f : \mathsf{V} \to \mathsf{W}$.

Definition 1.8 *Let X, Y be topological spaces. When we write $f :: \mathsf{X} \to \mathsf{Y}$ we imply only that f is defined on some open set in X. If we wish to indicate that f is defined near $p \in \mathsf{X}$ and that $f(\mathsf{p}) = \mathsf{q}$ we will used the pointed category notation together with the symbol "$::$":*

$$f :: (\mathsf{X}, \mathsf{p}) \to (\mathsf{Y}, \mathsf{q})$$

*We will refer to such maps as **local maps at** p. Local maps may be composed with the understanding that the domain of the composite map may become smaller: If $f :: (\mathsf{X}, \mathsf{p}) \to (\mathsf{Y}, \mathsf{q})$ and $g :: (\mathsf{Y}, \mathsf{q}) \to (\mathsf{G}, \mathsf{z})$ then $g \circ f :: (\mathsf{X}, \mathsf{p}) \to (\mathsf{G}, \mathsf{z})$ and the domain of $g \circ f$ will be a non-empty open set.*

Theorem 1.6 (The Rank Theorem) *Let $f : (\mathsf{V}, p) \to (\mathsf{W}, q)$ be a local map such that Df has constant rank r in an open set containing p. Suppose that $\dim(\mathsf{V}) = n$ and $\dim(\mathsf{W}) = m$ Then there are local diffeomorphisms $g_1 :: (\mathsf{V}, p) \to (\mathbb{R}^n, 0)$ and $g_2 :: (\mathsf{W}, q) \to (\mathbb{R}^m, 0)$ such that $g_2 \circ f \circ g_1^{-1}$ is a local diffeomorphism near 0 with the form*

$$(x^1, ...x^n) \mapsto (x^1, ...x^r, 0, ..., 0).$$

Proof. Without loss of generality we may assume that $f : (\mathbb{R}^n, 0) \to (\mathbb{R}^m, 0)$ and that (reindexing) the $r \times r$ matrix

$$\left(\frac{\partial f^j}{\partial x^j}\right)_{1 \le i,j \le r}$$

is nonsingular in an open ball centered at the origin of $\mathbb{R}^n$. Now form a map $g_1(x^1, ...x^n) = (f^1(x), ..., f^r(x), x^{r+1}, ..., x^n)$. The Jacobian matrix of g_1 has the block matrix form

$$\begin{bmatrix} \left(\frac{\partial f^i}{\partial x^j}\right) & \\ 0 & I_{n-r} \end{bmatrix}$$

which clearly has nonzero determinant at 0 and so by the inverse mapping theorem g_1 must be a local diffeomorphism near 0. Restrict the domain of g_1 to this possibly smaller open set. It is not hard to see that the map $f \circ g_1^{-1}$ is of the form $(z^1, ..., z^n) \mapsto (z^1, ..., z^r, \gamma^{r+1}(z), ..., \gamma^m(z))$ and so has Jacobian matrix of the form

$$\begin{bmatrix} I_r & 0 \\ * & \left(\frac{\partial \gamma^i}{\partial x^j}\right) \end{bmatrix}.$$

Now the rank of $\left(\frac{\partial \gamma^i}{\partial x^j}\right)_{r+1 \le i \le m,\ r+1 \le j \le n}$ must be zero near 0 since the $\operatorname{rank}(f) = \operatorname{rank}(f \circ h^{-1}) = r$ near 0. On the said (possibly smaller) neighborhood we now define the map $g_2 : (\mathbb{R}^m, q) \to (\mathbb{R}^m, 0)$ by

$$(y^1, ..., y^m) \mapsto (y^1, ..., y^r, y^{r+1} - \gamma^{r+1}(y_*, 0), ..., y^m - \gamma^m(y_*, 0))$$

where $(y_*, 0) = (y^1, ..., y^r, 0, ..., 0)$. The Jacobian matrix of g_2 has the form

$$\begin{bmatrix} I_r & 0 \\ * & I \end{bmatrix}$$

and so is invertible and the composition $g_2 \circ f \circ g_1^{-1}$ has the form

$$
\begin{aligned}
z &\overset{f\circ g_1^{-1}}{\mapsto} (z_*, \gamma_{r+1}(z), ..., \gamma_m(z)) \\
&\overset{g_2}{\mapsto} (z_*, \gamma_{r+1}(z) - \gamma_{r+1}(z_*, 0), ..., \gamma_m(z) - \gamma_m(z_*, 0))
\end{aligned}
$$

where $(z_*, 0) = (z^1, ..., z^r, 0, ..., 0)$. It is not difficult to check that $g_2 \circ f \circ g_1^{-1}$ has the required form near 0. ■

Starting with a fixed V, say the usual example $\mathbb{F}^n$, there are several standard methods of associating related vector space using multilinear algebra. The simplest example is the dual space $(\mathbb{F}^n)^*$. Now beside $\mathbb{F}^n$ there is also $\mathbb{F}_n$ which is also a space of $n-$tuples but this time thought of as *row vectors*. We shall often identify the dual space $(\mathbb{F}^n)^*$ with $\mathbb{F}_n$ so that for $\mathsf{v} \in \mathbb{F}^n$ and $\xi \in (\mathbb{F}^n)^*$ the duality is just matrix multiplication $\xi(\mathsf{v}) = \xi\mathsf{v}$. The group of nonsingular matrices, the general linear group $Gl(n, \mathbb{F})$ acts on each of these a natural way:

1. The primary action on $\mathbb{F}^n$ is a left action and corresponds to the standard representation and is simply multiplication from the left: $(g, \mathsf{v}) \to g\mathsf{v}$.
2. The primary action on $(\mathbb{F}^n)^* = \mathbb{F}_n$ is also a left action and is $(g, \mathsf{v}) \to \mathsf{v}g^{-1}$ (again matrix multiplication). In a setting where one insists on using only column vectors (even for the dual space) then this action appears as $(g, \mathsf{v}^t) \to (g^{-1})^t\mathsf{v}^t$. The reader may recognize this as giving the contragradient representation.

Differential geometry strives for invariance and so we should try to get away from the special spaces V such as $\mathbb{F}^n$ and $\mathbb{F}_n$ which often have a standard preferred basis. So let V be an abstract $\mathbb{F}-$vector space and V^* its dual. For every choice of basis $e = (e_1, ..., e_n)$ for V there is the natural map $u_e : \mathbb{F}^n \to V$ given by $e : v \mapsto e(v) = v^i e_i$. Identifying e with the row of basis vectors $(e^1, ..., e^n)$ we see that $u_e(v)$ is just formal matrix multiplication

$$
\begin{aligned}
u_e &: v \mapsto ev \\
&= (e_1, ..., e_n) \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}
\end{aligned}
$$

Corresponding to $e = (e_1, ..., e_n)$ there is the dual basis for V^* which we write as $e^* = (e^1, ..., e^n)$. In this case too we have a natural map $u_{e^*} : \mathbb{F}^{n*} := \mathbb{F}_n \to \mathsf{V}^*$ given by

$$
\begin{aligned}
u_{e^*} &: v^* \mapsto v^* e^* \\
&= (v^1, ..., v^n) \begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix}
\end{aligned}
$$

In each, the definition of basis tells us that $e : \mathbb{F}^n \to \mathsf{V}$ and $e^* : \mathbb{F}_n \to \mathsf{V}^*$ are both linear isomorphisms. Thus for a fixed frame e each $\mathsf{v} \in \mathsf{V}$ may be written uniquely $\mathsf{v} = ev$ while each $\mathsf{v}^* \in \mathsf{V}^*$ we have the expansion $\mathsf{v}^* = ve$

Chapter 2

Differentiable Manifolds

An undefined problem has an infinite number of solutions.
-Robert A. Humphrey

2.1 Rough Ideas I

The space of n-tuples $\mathbb{R}^n$ is often called Euclidean space by mathematicians but it might be a bit more appropriate the refer to this a Cartesian space which is what physics people often call it. The point is that Euclidean space (denoted here as E^n) has both more structure and less structure than Cartesian space. More since it has a notion of distance and angle, less because Euclidean space as it is conceived of in pure form has no origin or special choice of coordinates. Of course we almost always give $\mathbb{R}^n$ it usual structure as an inner product space from which we get the angle and distance and we are on our way to having a set theoretic model of Euclidean space.

Let us imagine we have a pure Euclidean space. The reader should think physical of space as it is normally given to intuition. Rene Descartes showed that if this intuition is axiomatized in a certain way then the resulting abstract space may be put into one to one correspondence with the set of n-tuples, the Cartesian space $\mathbb{R}^n$. There is more than one way to do this but if we want the angle and distance to match that given by the inner product structure on $\mathbb{R}^n$ then we get the familiar rectilinear coordinates.

After imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers. In fact, since a Euclidean space in this sense is an object of intuition (at least in 2d and 3d) some may insist that to be sure such a space of points really exists that we should in fact *start* with $\mathbb{R}^n$ and "forget" the origin and all the vector space structure while retaining the notion of point and distance. The coordinatization of Euclidean space is then just a "remembering" of this forgotten structure. Thus our coordinates arise from a map $x : E^n \to \mathbb{R}^n$ which is just the identity map.

The student must learn how differential geometry is actually done. These remarks are meant to encourage the student to stop and seek the simplest most intuitive viewpoint whenever feeling overwhelmed by notation. The student is encouraged to experiment with abbreviated personal notation when checking calculations and to draw diagrams and schematics that encode the geometric ideas whenever possible. The maxim should be "Let the picture write the equations".

Now this approach works fine as long as intuition doesn't mislead us. But on occasion intuition does mislead us and this is where the pedantic notation and the various abstractions can save us from error.

2.2 Topological Manifolds

A **topological manifold** is a paracompact[1] Hausdorff topological space M such that every point $p \in M$ is contained in some open set U_p which is the domain of a homeomorphism $\phi : U_p \to V$ onto an open subset of some Euclidean space $\mathbb{R}^n$. Thus we say that M is "locally Euclidean". Many authors assume that a manifolds is second countable and then show that paracompactness follows. It seems to the author that paracompactness is the really important thing. In fact, it is surprising how far one can go without assume second countability. This has the advantage of making foliations (defined later) manifolds. Nevertheless we will assume also second countability unless otherwise stated.

It might seem that the n in the definition might change from point to point or might not even be a well defined function on M depending essentially on the homeomorphism chosen. However, this in fact not true. It is a consequence of a fairly difficult result of Brower called "invariance of domain" that the "dimension" n must be a locally constant function and therefore constant on connected manifolds. This result is rather trivial if the manifold has a differentiable structure (defined below). We shall simply record Brower's theorem:

Theorem 2.1 (Invariance of Domain) *The image of an open set $U \subset \mathbb{R}^n$ by a 1-1 continuous map $f : U \to \mathbb{R}^n$ is open. It follows that if $U \subset \mathbb{R}^n$ is homeomorphic to $V \subset \mathbb{R}^m$ then $m = n$.*

Each connected component of a manifold could have a different dimension but we will restrict our attention to so called "pure manifolds" for which each component has the same dimension which we may then just refer to as the **dimension** of M. The latter is denoted $\dim(M)$. A **topological manifold with boundary** is a second countable Hausdorff topological space M such that point $p \in M$ is contained in some open set U_p which is the domain of a homeomorphism $\psi : U \to V$ onto an open subset V of some Euclidean half space $\mathbb{R}^n_- =: \{\vec{x} : x^1 \leq 0\}$[2]. A point that is mapped to the hypersurface $\mathbb{R}^n_- =: \{\vec{x} : x^1 = 0\}$ under one of these homeomorphism is called a boundary

[1]Paracompact means every open cover has a locally finite refinement.

[2]Using $\mathbb{R}^n_+ =: \{x : x^1 \geq 0\}$ is equivalent at this point in the development and is actually the more popular choice. Later on when we define orientation on a (smooth) manifold this

point. As a corollary to Brower's invariance of domain theorem this concept is independent of the homeomorphism used. The set of all boundary points of M is called the boundary of M and denoted ∂M. The interior is $\text{int}(M) := M - \partial M$.

Topological manifolds are automatically normal and paracompact. This means that each topological manifold supports C^0-partitions of unity: Given any cover of M by open sets $\{U_\alpha\}$ there is a family of continuous functions $\{\beta_i\}$ whose domains form a cover of M such that

(i) $\text{supp}(\beta_i) \subset U_\alpha$ for some α,

(ii) each $p \in M$ is contained in a neighborhood which intersect the support of only a finite number of the β_i.

(iii) we have $\sum \beta_i = 1$ (notice that the sum $\sum \beta_i(x)$ is finite for each $p \in M$ by (ii)).

Remark 2.1 *For differentiable manifolds we will be much more interested in the existence of smooth partitions of unity.*

2.3 Differentiable Manifolds and Differentiable Maps

The art of doing mathematics consists in finding that special case which contains all the germs of generality. Hilbert, David (1862-1943)

Definition 2.1 *Let M be a topological manifold. A pair $(U, \mathtt{x})$ where U is an open subset of M and $\mathtt{x} : U \to R^n$ is a homeomorphism is called a chart or coordinate system on M.*

If $(U, \mathtt{x})$ chart (with range in R^n) then $\mathtt{x} = (x^1, ..., x^n)$ for some functions x^i $(i = 1, ..., n)$ defined on U called coordinate functions. To be precise, we are saying that if $p_i : R^n \to \mathbb{R}$ is the obvious projection onto the i-th factor of $R^n := \mathbb{R} \times \cdots \times \mathbb{R}$ then $x^i := p_i \circ \mathtt{x}$ and so for $p \in M$ we have $\mathtt{x}(p) = (x^1(p), ..., x^n(p)) \in R^n$.

By the very definition of topological manifold we know that we may find a family of charts $\{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ whose domains cover M; that is $M = \cup_{\alpha \in A} U_\alpha$. Such a cover by charts is called an atlas for M. It is through the notion of change of coordinate maps (also called transition maps or overlap maps etc.) that we define the notion of a differentiable structure on a manifold.

Definition 2.2 *Let $\mathcal{A} = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ be an atlas on a topological manifold M. Whenever the overlap $U_\alpha \cap U_\beta$ between two chart domains is nonempty we have the* ***change of coordinates*** *map $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U_\beta) \to \mathtt{x}_\beta(U_\alpha \cap U_\beta)$. If all such change of coordinates maps are C^r-diffeomorphisms then we call the atlas a* ***C^r-atlas****.*

"negative" half space will be more convenient since we will be faced with less fussing over minus signs.

Now we might has any number of C^r**-atlases** on a topological manifold but we must have some notion of compatibility. A chart $(U, \mathtt{x})$ is **compatible** with some atlas $\mathcal{A} = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ on M if the maps $\mathtt{x} \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U) \to \mathtt{x}(U_\alpha \cap U)$ are C^r-diffeomorphisms defined. More generally, two C^r**-atlases** $\mathcal{A} = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ and $\mathcal{A}' = \{\mathtt{x}_{\alpha'}, U_{\alpha'}\}_{\alpha' \in A'}$ are said to be compatible if the union $\mathcal{A} \cup \mathcal{A}'$ is a C^r-atlas. It should be pretty clear that given a C^r-atlases on M there is a unique maximal atlas that contains $\mathcal{A}$ and is compatible with it. Now we are about to say a C^r-atlas on a topological manifold elevates it too the status C^r-differentiable manifold by giving the manifold a so-called C^r-structure (smooth structure) but there is a slight problem or two. First, if two different atlases are compatible then we don't really want to consider them to be giving different C^r-structures. To avoid this problem we will just use our observation about maximal atlases. The definition is as follows:

Definition 2.3 *A maximal C^r-atlas for a manifold M is called a C^r-differentiable structure. The manifold M together with this structure is called a C^r-differentiable manifold.*

Now note well that any atlas determine a unique C^r-differentiable structure on M since it determine the unique maximal atlas that contains it. So in practice we just have to cover a space with mutually C^r-compatible charts in order to turn it into (or show that it has the structure of) a C^r-differentiable manifold. In practice, we just need some atlas or maybe some small family of atlases that will become familiar to the reader for the most commonly studied smooth manifolds. For example, the space $\mathbb{R}^n$ is itself a C^∞ manifold (and hence a C^r-manifold for any $r \geq 0$) as we can take for an atlas for $\mathbb{R}^n$ the single chart $(id, \mathbb{R}^n)$ where $id : \mathbb{R}^n \to \mathbb{R}^n$ is just the identity map $id(x) = x$. Other atlases may be used in a given case and with experience it becomes more or less obvious which of the common atlases are mutually compatible and so technical idea of a maximal atlas usually fades into the background. For example, once we have the atlas $\{(id, \mathbb{R}^2)\}$ on the plane (consisting of the single chart) we have determined a differentiable structure on the plane. But then the chart given by polar coordinates is compatible with later atlas and so we could though this chart into the atlas and "fatten it up" a bit. In fact, there are many more charts that could be thrown into the mix if we needed then because in this case any local diffeomorphism $U \subset \mathbb{R}^2 \to \mathbb{R}^2$ would be compatible with the "identity" chart $(id, \mathbb{R}^2)$ and so would also be a chart within the same differentiable structure on $\mathbb{R}^2$. By the way, it is certainly possible for there to be two different differentiable structures on the same topological manifold. For example the chart given by the cubing function $(x \mapsto x^3, \mathbb{R}^1)$ is not compatible with the identity chart $(id, \mathbb{R}^1)$ but since the cubing function also has domain all of $\mathbb{R}^1$ it too provides an atlas. But then this atlas cannot be compatible with the usual atlas $\{(id, \mathbb{R}^1)\}$ and so they determine different maximal atlases. Now we have two different differentiable structures on the line $\mathbb{R}^1$. Actually, the two atlases are equivalent in a sense that we will make precise below (they are diffeomorphic). We say that the two differentiable structures are different but equivalent or diffeomorphic. On the other hand, it is a deep result proved fairly recently that there exist infinitely many non-diffeomorphic differentiable structures on $\mathbb{R}^4$. The reader ought to be wondering what is so special about dimension four.

Example 2.1 *Each Euclidean space $\mathbb{R}^n$ is a differentiable manifold in a trivial way. Namely, there is a single chart that forms an atlas*[3] *which is just the identity map $\mathbb{R}^n \to \mathbb{R}^n$. Notice however that the map $\varepsilon : (x^1, x^2, ..., x^n) \mapsto ((x^1)^{1/3}, x^2, ..., x^n)$ is also a chart. Thus we seem to have two manifolds $\mathbb{R}^n, \mathcal{A}_1$ and $\mathbb{R}^n, \mathcal{A}_2$. This is true but they are equivalent in another sense. Namely, they are diffeomorphic via the map ε. See definition 2.13 below. Actually, if V is any vector space with a basis $(f_1, ..., f_n)$ and dual basis $(f_1^*, ..., f_n^*)$ then once again, we have an atlas consisting of just one chart defined on all of V which is the map $\mathtt{x} : \mathrm{v} \mapsto (f_1^*\mathrm{v}, ..., f_n^*\mathrm{v}) \in \mathbb{R}^n$. On the other hand V may as well be modelled (in a sense to be defined below) on itself using the identity map as the sole member of an atlas! The choice is a matter of convenience and taste.*

Example 2.2 *The sphere $S^2 \subset \mathbb{R}^3$. Choose two points as north and south poles. Then off of these two pole points and off of a single half great circle connecting the poles we have the usual spherical coordinates. We actually have many such systems of spherical coordinates since we can re-choose the poles in many different ways. We can also use projection onto the coordinate planes as charts. For instance let U_z^+ be all $(x, y, z) \in S^2$ such that $z > 0$. Then $(x, y, z) \mapsto (x, y)$ provides a chart $U_z^+ \to \mathbb{R}^2$. The various transition functions can be computed explicitly and are clearly smooth. We can also use* ***stereographic projection*** *to give charts. More generally, we have the n-sphere $S^n \subset \mathbb{R}^{n+1}$ with two charts U^+, ψ^+ and U^-, ψ^- where*

$$U^{\pm} = \{p = (x_1, ..., x_{n+1}) \in S^n : x_{n+1} \neq \pm 1\}$$

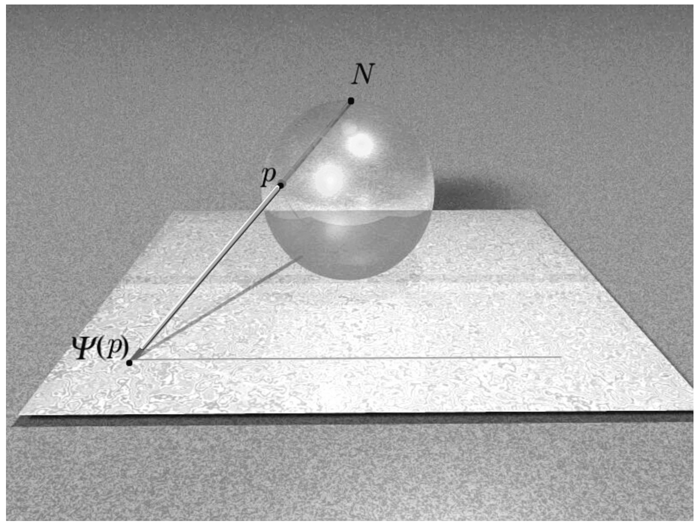
and ψ^+ (resp. ψ_-) is stereographic projection from the north pole $(0, 0....0, 1)$ (resp. south pole $(0, 0, ..., 0, -1)$). Explicitly we have

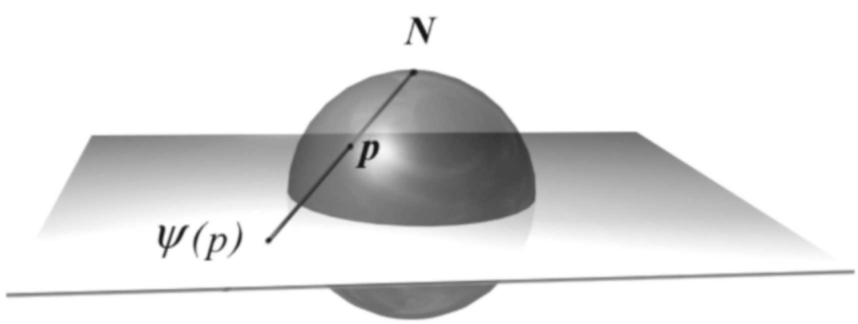
$$\psi_+(p) = \frac{1}{(1 - x_{n+1})}(x_1, ..., x_n) \in \mathbb{R}^n \tag{2.1}$$
$$\psi_-(p) = \frac{1}{(1 + x_{n+1})}(x_1, ..., x_n) \in \mathbb{R}^n$$

Exercise 2.1 *Compute $\psi_+ \circ \psi_-^{-1}$ and $\psi_-^{-1} \circ \psi_+$.*

Example 2.3 *The set of all lines through the origin in $\mathbb{R}^3$ is denoted $P_2(\mathbb{R})$ and is called the real projective plane . Let U_z be the set of all lines $\ell \in P_2(\mathbb{R})$*

[3] Of course there are many other compatible charts so this doesn't form a maximal atlas by a long shot.


N
p
Ψ(p)


N
p
ψ(p)

not contained in the x, y plane. Every line $p \in U_z$ intersects the plane $z = 1$ at exactly one point of the form $(x(\ell), y(\ell), 1)$. We can define a bijection $\psi_z : U_z \to \mathbb{R}^2$ by letting $p \mapsto (x(\ell), y(\ell))$. This is a chart for $P_2(\mathbb{R})$ and there are obviously two other analogous charts ψ_x, U_x and ψ_y, U_y which cover $P_2(\mathbb{R})$. More generally, the set of all lines through the origin in $\mathbb{R}^{n+1}$ is called projective n-space denoted $P_n(\mathbb{R})$ and can be given an atlas consisting of charts of the form ψ_i, U_i where

$$\begin{aligned}
U_i &= \{l \in P_n(\mathbb{R}) : \ell \text{ is not contained in the hyperplane } x^1 = 0 \\
\psi_i(\ell) &= \text{the unique coordinates } (u^1, ..., u^n) \text{ such that } (u^1, ..., 1, ..., u^n) \text{ is} \\
&\quad \text{on the line } \ell.
\end{aligned}$$

Example 2.4 *The graph of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ is the subset of the Cartesian product $\mathbb{R}^n \times \mathbb{R}$ given by $\Gamma_f = \{(x, f(x)) : x \in \mathbb{R}^n\}$. The projection map $\Gamma_f \to \mathbb{R}^n$ is a homeomorphism and provides a global chart on Γ_f making it a smooth manifold. More generally, let $S \subset \mathbb{R}^{n+1}$ be a subset which has the property that for all $x \in S$ there is an open neighborhood $U \subset \mathbb{R}^{n+1}$ and some function $f :: \mathbb{R}^n \to \mathbb{R}$ such that $U \cap S$ consists exactly of the points of in U of the form*

$$(x^1, .., x^{j-1}, f(x^1, ..., \widehat{x^j}, .., x^{n+1}), x^{j+1}, ..., x^n).$$

Then on $U \cap S$ the projection

$$(x^1, .., x^{j-1}, f(x^1, ..., \widehat{x^j}, .., x^{n+1}), x^{j+1}, ..., x^n) \mapsto (x^1, .., x^{j-1}, x^{j+1}, ..., x^n)$$

is a chart for S. In this way, S is a differentiable manifold. Notice that S is a subset of the manifold $\mathbb{R}^{n+1}$ and the manifold topology indu??ced by the atlas just described is the same as the relative topology of S in $\mathbb{R}^{n+1}$. The notion of regular submanifold generalizes this idea to arbitrary smooth manifolds.

Example 2.5 *The set of all $m \times n$ matrices $\mathbb{M}_{m\times n}$ (also written $\mathbb{R}^m_n$) is an mn-manifold modelled on $\mathbb{R}^{mn}$. We only need one chart again since it clear that $\mathbb{M}_{m\times n}$ is in natural 1-1 correspondence with $\mathbb{R}^{mn}$ by the map $[a_{ij}] \mapsto (a_{11}, a_{12}, ..., a_{mn})$. Also, the set of all non-singular matrices $\mathrm{GL}(n, \mathbb{R})$ is an open submanifold of $\mathbb{M}_{n\times n} \cong \mathbb{R}^{n^2}$.*

If we have two manifolds M_1 and M_2 we can form the topological Cartesian product $M_1 \times M_2$. We may give $M_1 \times M_2$ a differentiable structure that induces this same product topology in the following way: Let $\mathcal{A}_{M_1}$ and $\mathcal{A}_{M_2}$ be atlases for M_1 and M_2. Take as charts on $M_1 \times M_2$ the maps of the form

$$\mathtt{x}_\alpha \times \mathtt{y}_\gamma : U_\alpha \times V_\gamma \to \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$$

where $(\mathtt{x}_\alpha, U_\alpha)$ is a chart form $\mathcal{A}_{M_1}$ and $\mathtt{y}_\gamma, V_\gamma$ a chart from $\mathcal{A}_{M_2}$. This gives $M_1 \times M_2$ an atlas called the product atlas which induces a maximal atlas and hence a differentiable structure.

Example 2.6 *The circle is clearly a manifold and hence so is the product* $T = S^1 \times S^1$ *which is a torus.*

Example 2.7 *For any manifold* M *we can construct the "cylinder"* $M \times I$ *where* I *is some open interval in* $\mathbb{R}$.

2.4 Pseudo-Groups and Models Spaces

Without much work we can reformulate our definition of a manifold to include as special cases some very common generalizations which include complex manifolds and manifolds with boundary. If fact, we would also like to include infinite dimensional manifolds where the manifolds are modelled on infinite dimensional Banach spaces rather than $\mathbb{R}^n$. It is quite important for our purposes to realize that the spaces (so far just $\mathbb{R}^n$) which will be the model spaces on which we locally model our manifolds should have a distinguished family of local homeomorphisms. For example, C^r-differentiable manifolds are modelled on $\mathbb{R}^n$ where on the latter space we single out the local C^r-diffeomorphisms between open sets. But we also study complex manifolds, foliated manifolds, manifolds with boundary, Hilbert manifolds and so on. Thus we need appropriate model space but also, significantly, we need a distinguished family on maps on the model space. Next we are going to define the notion of a transformation pseudogroup which will be a family of maps with certain properties. The definition will seem horribly complex without first having something concrete in mind so we first single out a couple of examples that fit the abstract pattern we are after. The first one is just the set of all diffeomorphisms between open subsets of $\mathbb{R}^n$ (or any manifold). The second one, based on an example in the article [We4], is a bit more fanciful-a sort of "toy pseudogroup". Consider the object labelled "The model M" in figure 2.1. Consider this set as made of tiles and their edges (grout between the tiles plus the outer boundary). Let Γ be the set of all maps from open sets of the plane to open sets of the plane that are restrictions of rigid motions of the plane. The we take as our example $\Gamma_{toy} := \Gamma|_{\mathsf{M}}$ which is the homeomorphisms of (relatively) open sets of M to open sets of M which are restrictions of the maps in Γ (to the intersections of their domains with M).

Definition 2.4 *A* ***pseudogroup of transformations****, say* Γ*, of a topological space* X *is a family* $\{\Phi_\gamma\}_{\gamma \in I}$ *of homeomorphisms with domain* U_γ *and range* V_γ *both open subsets of* X*, which satisfies the following properties:*

1) $\mathrm{id}_X \in \Gamma$.

2) $\Phi_\gamma \in \Gamma$ *implies* $\Phi_\gamma^{-1} \in \Gamma$.

3) For any open set $U \subset X$*, the restrictions* $\Phi_\gamma|_U$ *are in* Γ *for all* $\Phi_\gamma \in \Gamma$.

4) The composition of any two elements Φ_γ*,* $\Phi_\nu \in \Gamma$ *are elements of* Γ *whenever the composition is defined:*

$$\Phi_\gamma \circ \Phi_\nu^{-1} : \Phi_\nu(U_\gamma \cap U_\nu) \to \Phi_\gamma(U_\gamma \cap U_\nu)$$

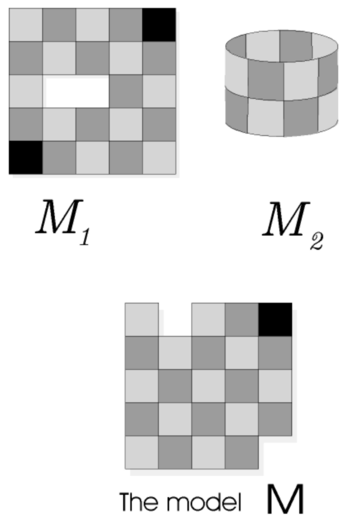


Figure 2.1: Tile and grout spaces

5) For any subfamily $\{\Phi_\gamma\}_{\gamma\in G_1} \subset \Gamma$ such that $\Phi_\gamma|_{U_\gamma\cap U_\nu} = \Phi_\nu|_{U_\gamma\cap U_\nu}$ whenever $U_\gamma \cap U_\nu \neq \emptyset$ then the mapping defined by $\Phi : \bigcup_{\gamma\in G_1} U_\gamma \to \bigcup_{\gamma\in G_1} V_\gamma$ is an element of Γ if it is a homeomorphism.

Exercise 2.2 *Check that each of these axioms is satisfied by our "fanciful example" Γ_{toy}.*

Definition 2.5 *A **sub-pseudogroup** Σ of a pseudogroup is a subset of Γ that is also a pseudogroup (and so closed under composition and inverses).*

We will be mainly interested in C^r-pseudogroups and the spaces which support them. Our main example will be the set $\Gamma^r_{\mathbb{R}^n}$ of all C^r diffeomorphisms between open subsets of $\mathbb{R}^n$. More generally, for a Banach space B we have the C^r-pseudo-group Γ^r_{B} consisting of all C^r diffeomorphisms between open subsets of a Banach space B. Since this is our prototype the reader should spend some time thinking about this example.

Definition 2.6 *A C^r- pseudogroup Γ of transformations of a subset M of Banach space B is a pseudogroup arising as the restriction to M of some sub-pseudogroup of Γ^r_{M}. The pair (M,Γ) is called a **model space** . As is usual in cases like this we sometimes just refer to M as the model space.*

Example 2.8 *Recall that a map $U \subset \mathbb{C}^n \to \mathbb{C}^n$ is holomorphic if the derivative (from the point of view of the underlying real space $\mathbb{R}^{2n}$) is in fact complex linear. A map holomorphic map with holomorphic inverse is called biholomorphic. The set of all biholomorphic maps between open subsets of $\mathbb{C}^n$ is a pseudogroup.*

This is a C^r-pseudogroup for all r including $r = \omega$. In this case the subset M *we restrict to is just $\mathbb{C}^n$ itself.*

Let us begin again redefine a few notions in greater generality. Let M be a topological space. An M-**chart** on M is a homeomorphism $\mathtt{x}$ whose domain is some subset $U \subset M$ and such that $\mathtt{x}(U)$ is an open subset (in the relative topology) of a fixed model space $\mathsf{M} \subset \mathsf{B}$.

Definition 2.7 *Let Γ be a C^r-pseudogroup of transformations on a model space* M*. A Γ-atlas, for a topological space M is a family of charts $\mathcal{A}_\Gamma = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ (where A is just an indexing set) which cover M in the sense that $M = \bigcup_{\alpha \in A} U_\alpha$ and such that whenever $U_\alpha \cap U_\beta$ is not empty then the map*

$$\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U_\beta) \to \mathtt{x}_\beta(U_\alpha \cap U_\beta)$$

is a member of Γ. The maps $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ are called various things by various authors including "transition maps", "coordinate change maps", and "overlap maps".

Now the way we set up the definition the model space M is a subset of a Banach space. If M the whole Banach space (the most common situation) and if $\Gamma = \Gamma^r_{\mathsf{M}}$ (the whole pseudogroup of local C^rdiffeomorphisms) then we have what was before called a C^r atlas M.

Exercise 2.3 *Show that this definition of C^r atlas is the same as our original definition in the case where* M *is the finite dimensional Banach space $\mathbb{R}^n$.*

In practice, a $\Gamma-$manifold is just a space M (soon to be a topological manifold) together with an $\Gamma-$atlas $\mathcal{A}$ but as before we should tidy things up a bit for our formal definitions. First, let us say that a bijection onto an open set in a model space, say $\mathtt{x} : U \to \mathtt{x}(U) \subset \mathsf{M}$, is **compatible** with the atlas $\mathcal{A}$ if for every chart $(\mathtt{x}_\alpha, U_\alpha)$ from the atlas $\mathcal{A}$ we have that the composite map

$$\mathtt{x} \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U) \to \mathtt{x}_\beta(U_\alpha \cap U)$$

is in Γ. The point is that we can then add this map in to form a larger equivalent atlas: $\mathcal{A}' = \mathcal{A} \cup \{\mathtt{x}, U\}$. To make this precise let us say that two different Γ atlases, say $\mathcal{A}$ and $\mathcal{B}$, are equivalent if every map from the first is compatible (in the above sense) with the second and visa-versa. In this case $\mathcal{A}' = \mathcal{A} \cup \mathcal{B}$ is also an atlas. The resulting equivalence class is called a $\Gamma-$**structure** on M.

Now it is clear that every equivalence class of atlases contains a unique **maximal atlas** which is just the union of all the atlases in the equivalence class. Of course every member of the equivalence class determines the maximal atlas also–just toss in every possible compatible chart and we end up with the maximal atlas again. Since the atlas was born out of the pseudogroup we might denote it by $\mathcal{A}_\Gamma$ and say that it gives M a

Definition 2.8 *A topological manifold M is called a C^r-**differentiable manifold** (or just C^r manifold) if it comes equipped with a differentiable structure. Whenever we speak of a differentiable manifold we will have a fixed differentiable structure and therefore a maximal C^r-atlas $\mathcal{A}_M$ in mind. A chart from $\mathcal{A}_M$ will be called an **admissible chart**.*

We started out with a topological manifold but if we had just started with a set M and then defined a chart to be a bijection $\mathtt{x} : U \to \mathtt{x}(U)$, only assuming $\mathtt{x}(U)$ to be open then a maximal atlas $\mathcal{A}_M$ would generate a topology on M . Then the set U would be open. Of course we have to check that the result is a paracompact space but once that is thrown into our list of demand we have ended with the same notion of differentiable manifold. To see how this approach would go the reader should consult the excellent book [A,B,R].

In the great majority of examples the subset $\mathsf{M} \subset \mathsf{V}$ is in fact equal to V itself. One important exception to this will lead us to a convenient formulation of manifold with boundary. First we need a definition:

Definition 2.9 *Let $\lambda \in \mathsf{M}^*$. In the case of $\mathbb{R}^n$ it will be enough to consider projection onto the first coordinate x^1. Now let $\mathsf{M}^+_\lambda = \{\mathsf{x} \in \mathsf{M}\colon \lambda(\mathsf{x}) \geq 0\}$ and $\mathsf{M}^-_\lambda = \{\mathsf{x} \in \mathsf{M}\colon \lambda(\mathsf{x}) \leq 0\}$ and $\partial\mathsf{M}^+_\lambda = \partial\mathsf{M}^-_\lambda = \{\mathsf{x} \in \mathsf{M}\colon \lambda(\mathsf{x}) = 0\}$ is the kernel of λ. Clearly M^+_λ and M^-_λ are homeomorphic and $\partial\mathsf{M}^-_\lambda$ is a closed subspace.* [4]

Example 2.9 *Let $\Gamma^r_{\mathsf{M}^-_\lambda}$ be the restriction to M^-_λ of the set of C^r-diffeomorphisms ϕ from open subset of M to open subsets of M which have the following property*

**) If the domain U of $\phi \in \Gamma^r_{\mathsf{M}}$ has nonempty intersection with $\mathsf{M}_0 := \{\mathsf{x} \in \mathsf{M}\colon \lambda(x) = 0\}$ then $\phi|_{\mathsf{M}_0 \cap U} : \mathsf{M}_0 \cap U \to \mathsf{M}_0 \cap U$.*

The model spaces together with an associated C^r-pseudogroup will be the basis of many of our geometric construction even if we do not explicitly mention it

Notation 2.1 *Most of the time we will denote the model space for a manifold M (resp. N etc.) by M (resp. N etc.) That is, we use the same letter but use the sans serif font (this requires the reader to be tuned into font differences). There will be exceptions. One exception will be the case where we want to explicitly indicate that the manifold is finite dimensional and thus modelled on $\mathbb{R}^n$ for some n. Another exception will be when E is the total space of a vector bundle over M. In this case E will be modelled on a space of the form $\mathsf{M} \times \mathsf{E}$. This will be explained in detail when study vector bundles.*

Now from the vantage point of this general notion of model space and the spaces modelled on them we get a slick definition of manifold with boundary. A topological manifold M is called a C^r-**differentiable manifold with boundary** (or just C^r manifold with boundary) if it comes equipped with a $\Gamma^r_{\mathsf{M}^-_\lambda}-$structure.

[4] The reason we will use both E^+ and E^- in the following definition for a technical reason having to do with the consistency of our definition of induced orientation of the boundary.

Remark 2.2 *It may be the case that there are two or more different differentiable structures on the same topological manifold. But see remark 2.4 below.*

Notice that the model spaces used in the definition of the charts were assumed to be a fixed space from chart to chart. We might have allowed for different model spaces but for topological reasons the model spaces must have constant dimension ($\leq \infty$) over charts with connected domain in a given connected component of M. In this more general setting if all charts of the manifold have range in a fixed M (as we have assumed) then the manifold is said to be a **pure manifold** and is said to be **modelled on** M. If in this case $\mathsf{M} = \mathbb{R}^n$ for some (fixed) $n < \infty$ then n is the **dimension** of M and we say that M is an n**-dimensional manifold** or n**-manifold** for short.

Convention Because of the way we have defined things all differentiable manifolds referred to in this book are assumed to be pure. We will denote the dimension of a (pure) manifold by $\dim(M)$.

Definition 2.10 *A chart* $(\mathtt{x}, U)$ *on* M *is said to be* ***centered at*** p *if* $\mathtt{x}(p) = 0 \in \mathsf{M}$.

If U is some open subset of a differentiable manifold M with atlas $\mathcal{A}_M$, then U is itself a differentiable manifold with an atlas of charts being given by all the restrictions $(\,\mathtt{x}_\alpha|_{U_\alpha \cap U}\,, U_\alpha \cap U\,)$ where $(\mathtt{x}_\alpha, U_\alpha) \in \mathcal{A}_M$. We call refer to such an open subset $U \subset M$ with this differentiable structure as an **open submanifold** of M.

Example 2.10 *Each Banach space* M *is a differentiable manifold in a trivial way. Namely, there is a single chart that forms an atlas*[5] *which is just the identity map* $\mathsf{M} \to \mathsf{M}$. *In particular* $\mathbb{R}^n$ *with the usual coordinates is a smooth manifold. Notice however that the map* $\varepsilon : (x^1, x^2, ..., x^n) \mapsto ((x^1)^{1/3}, x^2, ..., x^n)$ *is also a chart. It induces the usual topology again but the resulting maximal atlas is different! Thus we seem to have two manifolds* $\mathbb{R}^n, \mathcal{A}_1$ *and* $\mathbb{R}^n, \mathcal{A}_2$. *This is true but they are equivalent in another sense. Namely, they are diffeomorphic via the map* ε. *See definition 2.13 below. Actually, if* V *is any vector space with a basis* $(f_1, ..., f_n)$ *and dual basis* $(f_1^*, ..., f_n^*)$ *then one again, we have an atlas consisting of just one chart define on all of* V *defined by* $\mathtt{x} : \mathrm{v} \mapsto (f_1^*\mathrm{v}, ..., f_n^*\mathrm{v}) \in \mathbb{R}^n$. *On the other hand* V *may as well be modelled on itself using the identity map! The choice is a matter of convenience and taste.*

If we have two manifolds M_1 and M_2 we can form the topological Cartesian product $M_1 \times M_2$. We may give $M_1 \times M_2$ a differentiable structure that induces this same product topology in the following way: Let $\mathcal{A}_{M_1}$ and $\mathcal{A}_{M_2}$ be atlases for M_1 and M_2. Take as charts on $M_1 \times M_2$ the maps of the form

$$\mathtt{x}_\alpha \times \mathtt{y}_\gamma : U_\alpha \times V_\gamma \to \mathsf{M}_1 \times \mathsf{M}_2$$

[5] Of course there are many other compatible charts so this doesn't form a maximal atlases by a long shot.

where $(\mathtt{x}_\alpha, U_\alpha)$ is a chart form $\mathcal{A}_{M_1}$ and $\mathtt{y}_\gamma, V_\gamma$ a chart from $\mathcal{A}_{M_2}$. This gives $M_1 \times M_2$ an atlas called the product atlas which induces a maximal atlas and hence a differentiable structure. Thus we have formed the **product manifold** $M_1 \times M_2$ where it is understood that the differentiable structure is as described above.

It should be clear from the context that M_1 and M_2 are modelled on M_1 and M_2 respectively. Having to spell out what is obvious from context in this way would be tiring to both the reader and the author. Therefore, let us forgo such explanations to a greater degree as we proceed and depend rather on the common sense of the reader.

2.5 Smooth Maps and Diffeomorphisms

A function defined on a manifold or on some open subset is differentiable by definition if it appears differentiable in every coordinate system which intersects the domain of the function. The definition will be independent of which coordinate system we use because that is exactly what the mutual compatibility of the charts in an atlas guarantees. To be precise we have

Definition 2.11 *Let $f : O \subset M \to \mathbb{R}$ be a function on M with open domain O. We say that f is C^r-differentiable iff for every admissible chart $U, \mathtt{x}$ with $U \cap O \neq \emptyset$ the function*

$$f \circ \mathtt{x}^{-1} : \mathtt{x}(U \cap O) \to \mathbb{R}$$

is C^r-differentiable.

The reason that this definition works is because if $U, \mathtt{x}$, $\acute{U}, \acute{\mathtt{x}}$ are any two charts with domains intersecting O then

$$f \circ \mathtt{x}^{-1} = (f \circ \acute{\mathtt{x}}^{-1}) \circ (\acute{\mathtt{x}} \circ \mathtt{x}^{-1})$$

we have whenever both sides are defined and since $\acute{\mathtt{x}} \circ \mathtt{x}^{-1}$ is a C^r-diffeomorphism, we see that $f \circ \mathtt{x}^{-1}$ is C^r if and only if $f \circ \acute{\mathtt{x}}^{-1}$ is C^r The chain rule is at work here of course.

Remark 2.3 *We have seen that when we compose various maps as above the domain of the result will in general be an open set which is the largest open set so that the composition makes sense. If we do not wish to write out explicitly what the domain is then will just refer to the **natural domain** of the composite map.*

Definition 2.12 *Let M and N be C^r manifolds with corresponding maximal atlases $\mathcal{A}_M$ and $\mathcal{A}_N$ and modelled on $\mathbb{R}^n$ and $\mathbb{R}^d$ respectively. A continuous map $f : M \to N$ is said to be k **times continuously differentiable** or C^r if for every choice of charts $(\mathtt{x}, U)$ from $\mathcal{A}_M$ and $(\mathtt{y}, V)$ from $\mathcal{A}_N$ the composite map*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: \mathbb{R}^n {\to} \mathbb{R}^d$$

is C^r on its natural domain (see convention 26.41). The set of all C^r maps $M \to N$ is denoted $C^r(M, N)$ or sometimes $C^r(M \to N)$.

Exercise 2.4 *Explain why this is a well defined notion. Hint: Think about the chart overlap maps.*

Sometimes we may wish to speak of a map being C^r at a point and for that we have a modified version of the last definition: Let M and N be C^r manifolds with corresponding maximal atlases $\mathcal{A}_M$ and $\mathcal{A}_N$ and modelled on $\mathbb{R}^n$ and $\mathbb{R}^d$ respectively. A (pointed) map $f : (M, p) \to (N, q)$ is said to be k **times continuously differentiable** or C^r at p if for every choice of charts $(\mathtt{x}, U)$ from $\mathcal{A}_M$ and $(\mathtt{y}, V)$ from $\mathcal{A}_N$ containing p and $q = f(p)$ respectively, the composite map

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: (\mathbb{R}^n, \mathtt{x}(p)) \to (\mathbb{R}^d, \mathtt{y}(q))$$

is C^r on some open set containing $\psi(p)$.

Just as for maps between open sets of Euclidean spaces we have

Definition 2.13 *A bijective map $f : M \to N$ such that f and f^{-1} are C^r with $r \geq 1$ is called a C^r-**diffeomorphism**. In case $r = \infty$ we shorten C^∞-diffeomorphism to just **diffeomorphism**. The group of all diffeomorphisms of a C^∞ onto itself is denoted* $\mathrm{Diff}(M)$.

Example 2.11 *The map $r_\theta : S^2 \to S^2$ given by $r_\theta(x, y, z) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, z)$ for $x^2 + y^2 + z^2 = 1$ is a diffeomorphism (and also an isometry).*

Example 2.12 *The map $f : S^2 \to S^2$ given by $f(x, y, z) = (x\cos((1 - z^2)\theta) - y\sin((1 - z^2)\theta), x\sin((1 - z^2)\theta) + y\cos((1 - z^2)\theta), z)$ is also a diffeomorphism (but not an isometry). Try to picture this map.*

Definition 2.14 *C^r differentiable manifolds M and N will be called C^r diffeomorphic and then said to be in the same C^r diffeomorphism class iff there is a C^r diffeomorphism $f : M \to N$.*

Remark 2.4 *It may be that the same underlying topological space M carries two different differentiable structures and so we really have two differentiable manifolds. Nevertheless it may still be the case that they are diffeomorphic. The more interesting question is whether a topological manifold can carry differentiable structures that are not diffeomorphic. It turns out that $\mathbb{R}^4$ carries infinitely many pair-wise non-diffeomorphic structures (a very deep and difficult result) but $\mathbb{R}^k$ for $k \geq 5$ has only one diffeomorphism class.*

Definition 2.15 *A map $f : M \to N$ is called a local diffeomorphism iff every point $p \in M$ is in an open subset $U_p \subset M$ such that $f|_{U_p} : U_p \to f(U)$ is a diffeomorphism. (note?to?self: Must we assume $f(U)$ is open?)*

Example 2.13 *The map $\pi : S^2 \to \mathbb{RP}^2$ given by taking the point (x, y, z) to the line through this point and the origin is a local diffeomorphism but is not a diffeomorphism since it is 2-1 rather than 1-1.*

Example 2.14 *If we integrate the first order system of differential equations with initial conditions*

$$\begin{aligned} y &= x' \\ y' &= x \\ x(0) &= \xi \\ y(0) &= \theta \end{aligned}$$

we get solutions

$$\begin{aligned} x\,(t;\xi,\theta) &= \left(\tfrac{1}{2}\theta + \tfrac{1}{2}\xi\right)e^{t} - \left(\tfrac{1}{2}\theta - \tfrac{1}{2}\xi\right)e^{-t} \\ y\,(t;\xi,\theta) &= \left(\tfrac{1}{2}\theta + \tfrac{1}{2}\xi\right)e^{t} + \left(\tfrac{1}{2}\theta - \tfrac{1}{2}\xi\right)e^{-t} \end{aligned}$$

that depend on the initial conditions (ξ,θ)*. Now for any* t *the map* $\Phi_t : (\xi,\theta) \mapsto (x(t,\xi,\theta), y(t,\xi,\theta))$ *is a diffeomorphism* $\mathbb{R}^2 \to \mathbb{R}^2$*. This is a special case of a moderately hard theorem.*

Example 2.15 *The map* $(x,y) \mapsto (\frac{1}{1-z(x,y)}x, \frac{1}{1-z(x,y)}y)$ *where* $z(x,y) = \sqrt{1-x^2-y^2}$ *is a diffeomorphism from the open disk* $B(0,1) = \{(x,y) : x^2+y^2 < 1\}$ *onto the whole plane. Thus* $B(0,1)$ *and* $\mathbb{R}^2$ *are diffeomorphic and in this sense the "same" differentiable manifold.*

We shall often need to consider maps which are defined on subsets $S \subset M$ that are not necessarily open. We shall call such a map f smooth (resp. C^r) if at each point it locally as restriction of a smooth map $\widetilde{f}$ defined on an open set O containing S. In particular a curve defined on a closed interval $[a,b]$ is called smooth if it has a smooth extension to an open interval containing $[a,b]$. We will occasionally need the following simple concept:

Definition 2.16 *A continuous curve* $c : [a,b] \to M$ *into a smooth manifold is called* ***piecewise smooth*** *is there exists a partition* $a = t_0 < t_1 < \cdots < t_k = b$ *such that* c *restricted to each* $[t_i, t_{i+1}]$ *is smooth*
for $0 \le i \le k-1$.

Before going to the next section let us compare local expressions in index notation with the index free notation (the latter being amenable to infinite dimensions). Consider an arbitrary pair of charts $\mathtt{y}$and $\mathtt{x}^{-1}$ and the transition maps $\mathtt{y} \circ \mathtt{x}^{-1} : \mathtt{x}(U \cap V) \to \mathtt{y}(U \cap V)$. We write

$$\mathtt{y}(p) = \mathtt{y} \circ \mathtt{x}^{-1}(\mathtt{x}(p))$$

for $p \in U \cap V$. For finite dimensional manifolds we see this written as

$$y^i(p) = y^i(x^1(p), ..., x^n(p)) \tag{2.2}$$

which make sense but we also see

$$y^i = y^i(x^1, ..., x^n). \tag{2.3}$$

In this last expression one might wonder if the x^i are functions or numbers. But this ambiguity is sort of purposeful for if 2.5 is true for all $p \in U \cap V$ then

2.5 is true for all $(x^1, ..., x^n) \in \mathtt{x}(U \cap V)$ and so we are unlikely to be led into error. This common and purposely notational ambiguity is harder to pull of in the case of index free notation. We will instead, write two different expressions in which the lettering and fonts are intended to be at least reminiscent of the classical index notation:

$$\mathtt{y}(p) = \mathtt{y} \circ \mathtt{x}^{-1}(\mathtt{x}(p))$$
and
$$\mathsf{y} = \mathtt{y} \circ \mathtt{x}^{-1}(\mathsf{x}).$$

In the first case, $\mathtt{x}$ and $\mathtt{y}$ are functions on $U \cap V$ while in the second x and y are elements of $\mathtt{x}(U \cap V)$ and $\mathtt{y}(U \cap V)$ respectively[6]. In order not to interfere with our subsequent development let us anticipate the fact that this notational principle will be manifest later when we compare and make sense out of the following familiar looking expressions:

$$d\mathtt{y}(\xi) = \left.\frac{\partial \mathsf{y}}{\partial \mathsf{x}}\right|_{\mathtt{x}(p)} \circ d\mathtt{x}(\xi)$$
and
$$\mathsf{w} = \left.\frac{\partial \mathsf{y}}{\partial \mathsf{x}}\right|_{\mathtt{x}(p)} \mathsf{v}$$

which should be compared with the classical expressions

$$dy^i(\xi) = \frac{\partial y^i}{\partial x^k} dx^k(\xi)$$
and
$$w^i = \frac{\partial y^i}{\partial x^k} v^k.$$

2.6 Coverings and Discrete groups

2.6.1 Covering spaces and the fundamental group

In this section and later when we study fiber bundles many of the results are interesting and true in either the purely topological category or in the differentiable category. In order to not have to do things twice let us agree to mean by C^r-manifold if $r \geq 1$ and if $r = 0$ simply a paracompact Hausdorff topological space in case $r = 0$. All relevant maps are to be C^r where if $r = 0$ we just mean continuous.

"C^0-diffeomorphism"	$=$	C^0-isomorphism	$=$	homeomorphism
"C^0-manifold"	$=$	Hausdorff space		
C^0-group	$=$	topological group		

[6]Notice the font differences.

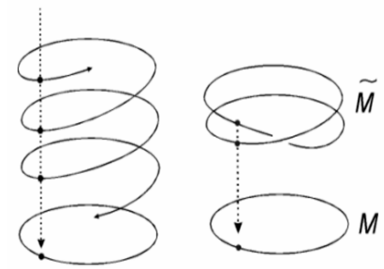


Figure 2.2: The line covers the circle.

In this section we recall a few facts about the fundamental group and covering spaces. In order to unify the presentation let us agree that "C^r diffeomorphism" just means homeomorphism in case $r = 0$. Of course a C^0 map is just a continuous map. Also, much of what we do for C^0 maps works for more general topological spaces and so the word "manifolds" could be replaced by topological space although the technical condition of being "locally simply con?nected" (LSC) is sometimes needed. All manifolds are LSC.

We may define a simple equivalence relation on a topological space by declaring

$$p \sim q \Leftrightarrow \text{ there is a continuous curve connecting } p \text{ to } q.$$

The equivalence classes are called path components and if there is only one such class then we say that M is path connected. The following exercise will be used whenever needed without explicit mention:

Exercise 2.5 *The path components of a manifold are exactly the connected components of M. Thus, a manifold is connected if and only if it is path connected.*

Definition 2.17 *Let $\widetilde{M}$ and M be C^r-spaces. A surjective C^r map $\wp : \widetilde{M} \to M$ is called a C^r* ***covering map*** *if every point $p \in M$ has an open connected neighborhood U such that each connected component $\widetilde{U}_i$ of $\wp^{-1}(U)$ is C^r diffeomorphic to U via the restriction $\wp|_{\widetilde{U}_i} : \widetilde{U}_i \to U$. We say that U is* ***evenly covered****. The triple $(\widetilde{M}, \wp, M)$ is called a covering space. We also refer to the space $\widetilde{M}$ (somewhat informally) as a* ***covering space*** *for M.*

We are mainly interested in the case where the spaces and maps are smooth. In this case we call $\widetilde{M}$ (informally) a **covering manifold**.

Example 2.16 *The map $\mathbb{R} \to S^1$ given by $t \mapsto e^{it}$ is a covering. If $e^{i\theta} \in S^1$ $(0 \leq \theta < 2\pi)$ then the points of the form $\{e^{it} : \theta - \pi < t < \theta + \pi\}$ is an open set evenly covered by the intervals I_n in the real line given by $I_n := (\theta - \pi + n2\pi, \theta + \pi + n2\pi)$.*

Exercise 2.6 *Explain why the map* $(-2\pi, 2\pi) \to S^1$ *given by* $t \mapsto e^{it}$ *is* not *a covering map.*

The set of all C^r covering spaces are the objects of a category. A morphism between covering spaces, say $(\widetilde{M}_1, \wp_1, M_1)$ and $(\widetilde{M}_2, \wp_2, M_2)$ is a pair of maps $(\widetilde{f}, f)$ which give a commutative diagram

$$\begin{array}{ccc} \widetilde{M}_1 & \xrightarrow{\widetilde{f}} & \widetilde{M}_2 \\ \downarrow & & \downarrow \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

which means that $f \circ \wp_1 = \wp_2 \circ \widetilde{f}$. Similarly the set of coverings of a fixed space M are the objects of a category where the morphisms are maps $\Phi : \widetilde{M}_1 \to \widetilde{M}_2$ required to make the following diagram commute:

$$\begin{array}{ccc} \widetilde{M}_1 & \xrightarrow{\Phi} & \widetilde{M}_2 \\ \downarrow & & \downarrow \\ M & \xrightarrow{id_M} & M \end{array}$$

so that $\wp_1 = \wp_2 \circ \Phi$. Now let $(\widetilde{M}, \wp, M)$ be a C^r covering space. The set of all C^r-diffeomorphisms Φ which are automorphisms in the above category; that is, diffeomorphisms for which $\wp = \wp \circ \Phi$, are called **deck transformations**. A deck transformation permutes the elements of each fiber $\wp^{-1}(p)$. In fact, it is easy to see that if $U \subset M$ is evenly covered then Φ permutes the connected components of $\wp^{-1}(U)$.

Proposition 2.1 *If* $\wp : \widetilde{M} \to M$ *is a* C^r *covering map with* M *being connected then the cardinality of* $\wp^{-1}(p)$ *is either infinite or is independent of* p*. In the latter case the cardinality of* $\wp^{-1}(p)$ *is the* ***multiplicity*** *of the covering.*

Proof. Fix $k \leq \infty$. Let U_k be the set of all points such that $\wp^{-1}(p)$ has cardinality k. It is easy to show that U_k is both open and closed and so, since M is connected, U_k is either empty or all of M. ■

Definition 2.18 *Let* $\alpha : [0,1] \to M$ *and* $\beta : [0,1] \to M$ *be two continuous (or* C^r*) maps (paths) both starting at* $p \in M$ *and both ending at* q*. A* ***fixed end point (***C^r***) homotopy*** *from* α *to* β *is a family of maps* $H_s : [0,1] \to M$ *parameterized by* $s \in [0,1]$ *such that*

1) $H : [0,1] \times [0,1] \to M$ *is continuous (or* C^r*) where* $H(t,s) := H_s(t)$*,*

2) $H_0 = \alpha$ *and* $H_1 = \beta$*,*

3) $H_s(0) = p$ *and* $H_s(1) = q$ *for all* $s \in [0,1]$*.*

Definition 2.19 *If there is a (*C^r*) homotopy from* α *to* β *we say that* α *is* ***homotopic*** *to* β *and write* $\alpha \simeq \beta$ *(*C^r*).*

It turns out that every continuous path on a C^r manifold may be uniformly approximated by a C^r path. Furthermore, if two C^r paths are continuously homotopic then they are C^r homotopic. Thus we may use smooth paths and smooth homotopies whenever convenient.

It is easily checked that homotopy is an equivalence relation. Let $P(p,q)$ denote the set of all continuous paths from p to q defined on $[0,1]$. Every $\alpha \in P(p,q)$ has a unique inverse path $\alpha^{\leftarrow}$ defined by

$$\alpha^{\leftarrow}(t) := \alpha(1-t).$$

If p_1, p_2 and p_3 are three points in M then for $\alpha \in P(p_1,p_2)$ and $\beta \in P(p_2,p_3)$ we can "multiply" the paths to get a path $\alpha * \beta \in P(p_1,p_3)$ defined by

$$\alpha * \beta(t) := \begin{cases} \alpha(2t) & \text{for } 0 \le t < 1/2 \\ \\ \beta(2t-1) & \text{for } 1/2 \le t < 1 \end{cases}.$$

An important observation is that if $\alpha_1 \simeq \alpha_2$ and $\beta_1 \simeq \beta_2$ then

$\alpha_1 * \beta_1 \simeq \alpha_2 * \beta_2$ where the homotopy between $\alpha_1 * \beta_1$ and $\alpha_2 * \beta_2$ is given in terms of the homotopy $H_\alpha : \alpha_1 \simeq \alpha_2$ and $H_\beta : \beta_1 \simeq \beta_2$ by

$$H(t,s) := \begin{cases} H_\alpha(2t,s) & \text{for } 0 \le t < 1/2 \\ \\ H_\beta(2t-1,s) & \text{for } 1/2 \le t < 1 \end{cases} \quad \text{and } 0 \le s < 1$$

Similarly, if $\alpha_1 \simeq \alpha_2$ then $\alpha_1^{\leftarrow} \simeq \alpha_2^{\leftarrow}$. Using this information we can define a group structure on the set of homotopy equivalence classes of loops, that is, of paths in $P(p,p)$ for some fixed $p \in M$. First of all, we can always form $\alpha * \beta$ for any $\alpha, \beta \in P(p,p)$ since we are always starting and stopping at the same point p. Secondly we have the following

Proposition 2.2 *Let $\pi_1(M,p)$ denote the set of fixed end point homotopy classes of paths from $P(p,p)$. For $[\alpha],[\beta] \in \pi_1(M,p)$ define $[\alpha] \cdot [\beta] := [\alpha * \beta]$. This is a well define multiplication and with this multiplication $\pi_1(M,p)$ is a group. The identity element of the group is the homotopy class 1 of the constant map $1_p : t \to p$, the inverse of a class $[\alpha]$ is $[\alpha^{\leftarrow}]$.*

Proof. We have already shown that $[\alpha] \cdot [\beta] := [\alpha * \beta]$ is well defined. One must also show that

1) For any α, the paths $\alpha \circ \alpha^{\leftarrow}$ and $\alpha^{\leftarrow} \circ \alpha$ are both homotopic to the constant map 1_p.

2) For any $\alpha \in P(p,p)$ we have $1_p * \alpha \simeq \alpha$ and $\alpha * 1_p \simeq \alpha$.

3) For any $\alpha, \beta, \gamma \in P(p,p)$ we have $(\alpha * \beta) * \gamma \simeq \alpha * (\beta * \gamma)$.

The first two of these are straight forward and left as exercises. L??L ■

The group $\pi_1(M,p)$ is called the **fundamental group** of M at p. If $\gamma : [0,1] \to M$ is a path from p to q then we have a group isomorphism $\pi_1(M,q) \to \pi_1(M,p)$ given by

$$[\alpha] \mapsto [\gamma * \alpha * \gamma^{\leftarrow}].$$

(One must check that this is well defined.) As a result we have

Proposition 2.3 *For any two points p, q in the same path component of M, the groups $\pi_1(M,p)$ and $\pi_1(M,q)$ are isomorphic (by the map described above).*

Corollary 2.1 *If M is connected then the fundamental groups based at different points are all isomorphic.*

Because of this last proposition, if M is connected we may refer to the fundamental group of M.

Definition 2.20 *A path connected topological space is called **simply connected** if $\pi_1(M) = \{1\}$.*

The fundamental group is actually the result of applying a functor. To every pointed space (M,p) we assign the fundamental group $\pi_1(M,p)$ and to every base point preserving map (pointed map) $f : (M,p) \to (N, f(p))$ we may assign a group homomorphism $\pi_1(f) : \pi_1(M,p) \to \pi_1(N, f(p))$ by

$$\pi_1(f)([\alpha]) = [f \circ \alpha].$$

It is easy to check that this is a covariant functor and so for composable pointed maps $(M,x) \xrightarrow{f} (N,y) \xrightarrow{g} (P,z)$ we have $\pi_1(g \circ f) = \pi_1(g)\pi_1(f)$.

Notation 2.2 *To avoid notational clutter we will denote $\pi_1(f)$ by $f_\#$.*

Theorem 2.2 *Every connected manifold M has a simply connected covering space. Furthermore, if H is any subgroup of $\pi_1(M,p)$, then there is a connected covering $\wp : \widetilde{M} \to M$ and a point $\widetilde{p} \in \widetilde{M}$ such that $\wp_\#(\pi_1(\widetilde{M}, \widetilde{p})) = H$.*

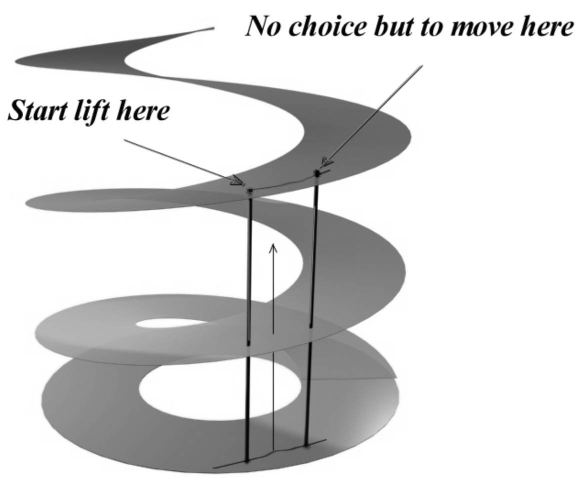
Definition 2.21 *Let $f : P \to M$ be a C^r−map. A map $\widetilde{f} : P \to \widetilde{M}$ is said to be a **lift** of the map f if $\wp \circ \widetilde{f} = f$.*

Theorem 2.3 *Let $\wp : \widetilde{M} \to M$ be a covering of C^r manifolds and $\gamma : [a,b] \to M$ a C^r−curve and pick a point y in $\wp^{-1}(\gamma(a))$. Then there exist a unique C^r lift $\widetilde{\gamma} : [a,b] \to \widetilde{M}$ of γ such that $\widetilde{\gamma}(a) = y$. Thus the following diagram commutes.*

$$\begin{tikzcd} & \widetilde{M} \arrow[d, "\wp"] \\ {[a,b]} \arrow[ur, "\widetilde{\gamma}"] \arrow[r] & M \end{tikzcd}$$

Similarly, if $h : [a,b] \times [c,d] \to M$ is a C^r−map then it has a unique lift $\widetilde{h} : [a,b] \times [c,d] \to \widetilde{M}$.

Proof. Figure **??** shows the way. Decompose the curve γ into segments which lie in evenly covered open sets. Lift inductively starting by using the inverse of $\wp$ in the first evenly covered open set. It is clear that in order to connect up continuously, each step is forced and so the lifted curve is unique.


No choice but to move here
Start lift here

The proof of the second half is just slightly more complicated but the idea is the same and the proof is left to the curious reader. A tiny technicality in either case it the fact that for $r > 0$ a C^r-map on a closed set is defined to mean that there is a C^r-map on a slightly larger open set. For instance, for the curve γ we must lift an extension $\gamma_{ext} : (a-\varepsilon, b+\varepsilon) \to M$ but considering how the proof went we see that the procedure is the same and gives a C^r- extension $\widetilde{\gamma}_{ext}$ of the lift $\widetilde{\gamma}$. ■

There are two important corollaries to this result. The first is that if $\alpha, \beta : [0,1] \to M$ are fixed end point homotopic paths in M and $\widetilde{\alpha} \simeq \widetilde{\beta}$ are lifts with $\widetilde{\alpha}(0) = \widetilde{\beta}(0)$ then any homotopy $h_t : \alpha \simeq \beta$ lifts to a homotopy $\widetilde{h}_t : \widetilde{\alpha} \simeq \widetilde{\beta}$. The theorem then implies that $\widetilde{\alpha}(1) = \widetilde{\beta}(1)$. From this one easily prove the following

Corollary 2.2 *For every $[\alpha] \in \pi_1(M,p)$ there is a well defined map $[\alpha]_\sharp : \wp^{-1}(p) \to \wp^{-1}(p)$ given by letting $[\alpha]_\sharp y$ be $\widetilde{\alpha}(1)$ for the lift of α with $\widetilde{\alpha}(0) = y$. (Well defined means that any representative $\alpha' \in [\alpha]$ gives the same answer.)*

Now recall the notion of a deck transformation. Since we now have two ways to permute the elements of a fiber, on might wonder about their relationship. For instance, given a deck transformation Φ and a chosen fiber $\wp^{-1}(p)$ when do we have $\Phi|_{\wp^{-1}(p)} = [\alpha]_\sharp$ for some $[\alpha] \in \pi_1(M,p)$?

2.6.2 Discrete Group Actions

Let G be a group and endow G with the discrete topology so in particular every point is an open set. In this case we call G a discrete group. If M is a topological space then so is $G \times M$ with the product topology. What does it mean for a map $\alpha : G \times M \to M$ to be continuous? The topology of $G \times M$ is clearly generated by sets of the form $S \times U$ where S is an arbitrary subset of G and U is open in M. The map $\alpha : G \times M \to M$ will be continuous if for any point $(g_0, x_0) \in G \times M$ and any open set $U \subset M$ containing $\alpha(g_0, x_0)$ we can find an open set $S \times V$ containing (g_0, x_0) such that $\alpha(S \times V) \subset U$. Since the topology of G is discrete, it is necessary and sufficient that there is an open V such that $\alpha(g_0 \times V) \subset U$. It is easy to see that a necessary and sufficient condition for α to be continuous on all of $G \times M$ is that the partial maps $\alpha_g(.) := \alpha(g, .)$ are continuous for every $g \in G$.

Definition 2.22 *Let G and M be as above. A (left) discrete group action is a map $\alpha : G \times M \to M$ such that for every $g \in G$ the partial map $\alpha_g(.) := \alpha(g, .)$ is continuous and such that the following hold:*

1) $\alpha(g_2, \alpha(g_1, x)) = \alpha(g_2 g_1, x)$ for all $g_1, g_2 \in G$ and all $x \in M$.

2) $\alpha(e, x) = x$ for all $x \in M$.

It follows that if $\alpha : G \times M \to M$ is a discrete action then each partial map $\alpha_g(.)$ is a homeomorphism with $\alpha_g^{-1}(.) = \alpha_{g^{-1}}(.)$. It is traditional to write $g \cdot x$ or just gx in place of the more accurate $\alpha(g, x)$. Using this notation we have $g_2(g_1 x) = (g_2 g_1)x$ and $ex = x$.

Definition 2.23 *A discrete group action is C^r if M is a C^r manifold and each $\alpha_g(.)$ is a C^r map.*

If we have a discrete action $\alpha : G \times M \to M$ then for a fixed x, the set $G \cdot x := \{g \cdot x : g \in G\}$ is called the orbit of x. It is easy to see that two orbits $G \cdot x$ and $G \cdot y$ are either disjoint or identical. In fact, we have equivalence relation on M where $x \sim y$ iff there exists a $g \in G$ such that $gx = y$. The equivalence classes are none other than the orbits. The natural projection onto set of orbits $p : M \to M/G$ given by

$$x \mapsto G \cdot x.$$

If we give M/G the quotient topology then of course p is continuous but more is true: The map $p : M \to M/G$ is an open map. To see this notice that if $U \subset M$ and we let $\widetilde{U} := p(U)$ then $p^{-1}(\widetilde{U})$ is open since

$$p^{-1}(\widetilde{U}) = \bigcup \{gU : g \in G\}$$

which is a union of open sets. Now since $p^{-1}(\widetilde{U})$ is open, $\widetilde{U}$ is open by definition of the quotient topology.

Example 2.17 *Let $\phi : M \to M$ be a diffeomorphism and let $\mathbb{Z}$ act on M by $n \cdot x := \phi^n(x)$ where*

$$\begin{aligned} \phi^0 &:= \mathrm{id}_M, \\ \phi^n &:= \phi \circ \cdots \circ \phi \text{ for } n > 0 \\ \phi^{-n} &:= (\phi^{-1})^n \text{ for } n > 0. \end{aligned}$$

This gives a discrete action of $\mathbb{Z}$ on M.

Definition 2.24 *A discrete group action $\alpha : G \times M \to M$ is said to act* ***properly*** *if for every for every $x \in M$ there is an open set $U \subset M$ containing x such that unless $g = e$ we have $gU \cap U = \emptyset$ for all $g \neq e$. We shall call such an open set* ***self avoiding****.*

It is easy to see that if $U \subset M$ is self avoiding then any open subset $V \subset U$ is also self avoiding. Thus every point $x \in M$ has a self avoiding neighborhood that is connected.

Proposition 2.4 *If $\alpha : G \times M \to M$ is a proper action and $U \subset M$ is self avoiding then p maps gU onto $p(U)$ for all $g \in G$ and the restrictions $p|_{gU} : gU \to p(U)$ are homeomorphisms. In fact, $p : M \to M/G$ is a covering map.*

Proof. Since $U \cong gU$ via $x \mapsto gx$ and since x and gx are in the same orbit, we see that gU and U both project to same set $p(U)$. Now if $x, y \in gU$ and $p|_{gU}(x) = p|_{gU}(y)$ then $y = hx$ for some $h \in G$. But also $x = ga$ (and $y = gb$) for some $a, b \in U$. Thus $h^{-1}gb = x$ so $x \in h^{-1}gU$. On the other hand we also know that $x \in gU$ so $h^{-1}gU \cap gU \neq \emptyset$ which implies that $g^{-1}h^{-1}gU \cap U \neq \emptyset$. Since U is self avoiding this means that $g^{-1}h^{-1}g = e$ and so $h = e$ from which we get $y = x$. Thus $p|_{gU} : gU \to p(U)$ is 1-1. Now since $p|_{gU}$ is clearly onto and since we also know that $p|_{gU}$ is an open continuous map the result follows. ∎

Example 2.18 *Fix a basis $(\mathbf{f}_1, \mathbf{f}_2)$ of $\mathbb{R}^2$. Let $\mathbb{Z}^2$ act on $\mathbb{R}^2$ by $(m,n) \cdot (x,y) := (x,y) + m\mathbf{f}_1 + n\mathbf{f}_2$. This action is easily seen to be proper.*

Example 2.19 *Let $\mathbb{Z}_2 := \{1, -1\}$ act on the sphere by $(\pm 1) \cdot \vec{x} := \pm\vec{x}$. Thus the action is generated by letting -1 send a point on the sphere to its antipode. This action is also easily seen to be proper.*

Exercise 2.7 (!) *Let $\alpha : G \times M \to M$ act properly on M. Show that if U_1 and U_2 are self avoiding and $p(U_1) \cap p(U_2) \neq \emptyset$ then there is a $g \in G$ such that $\alpha_g(U_1) \cap U_2 \neq \emptyset$. Show also that if α_g is C^r then α_g maps $U_1 \cap \alpha_{g^{-1}}(U_2) := O_1$ diffeomorphically onto $\alpha_g(U_1) \cap U_2 := O_2$ (homeomorphically if $r = 0$) and in this case*

$$\alpha_g|_{O_1} = p|_{O_2}^{-1} \circ p|_{O_1}.$$

Hint: If U_1 and U_2 are self avoiding then so are O_1 and O_2 and $p(O_1) = p(O_2)$.

Proposition 2.5 *Let $\alpha : G \times M \to M$ act properly by C^r-diffeomorphisms on a C^r- manifold M. Then the quotient space M/G has a natural C^r structure such that the quotient map is C^r local diffeomorphism. The quotient map is a covering map*

Proof. We exhibit the atlas on M/G and then let the reader finish the (easy) proof. Let $\mathcal{A}_M$ be an atlas for M. Let $\bar{x} = Gx$ be a point in M/G and pick an open $U \subset M$ which contains a point, say x, in the orbit Gx which (as a set) is the preimage of $\bar{x}$ and such that unless $g = e$ we have $gU \cap U = \emptyset$. Now let $U_\alpha, \mathtt{x}_\alpha$ be a chart on M containing x. By replacing U and U_α by $U \cap U_\alpha$ we may assume that $\mathtt{x}_\alpha$ is defined on $U = U_\alpha$. In this situation, if we let $U^* := p(U)$ then each restriction $p|_U : U \to U^*$ is a homeomorphism. We define a chart map $\mathtt{x}_\alpha^*$ with domain U_α^* by

$$\mathtt{x}_\alpha^* := \mathtt{x}_\alpha \circ p|_{U_\alpha}^{-1} : U_\alpha^* \to \mathbb{R}^n.$$

Let $\mathtt{x}_\alpha^*$ and $\mathtt{x}_\beta^*$ be two such chart maps with domains U_α^* and U_β^*. If $U_\alpha^* \cap U_\beta^* \neq \emptyset$ then we have to show that $\mathtt{x}_\beta^* \circ (\mathtt{x}_\alpha^*)^{-1}$ is a C^r diffeomorphism. Let $\bar{x} \in U_\alpha^* \cap U_\beta^*$ and abbreviate $U_{\alpha\beta}^* = U_\alpha^* \cap U_\beta^*$. Since $U_\alpha^* \cap U_\beta^* \neq \emptyset$ there must be a $g \in G$ such that $\alpha_g(U_\alpha) \cap U_\beta \neq \emptyset$. Using exercise 2.7and letting $\alpha_g(U_\alpha) \cap U_\beta := O_2$ and $O_1 := \alpha_{g^{-1}} O_2$ we have

$$\begin{aligned}
&\mathtt{x}_\beta^* \circ \mathtt{x}_\alpha^*|^{-1} \\
&= \mathtt{x}_\beta \circ p|_{O_2}^{-1} \circ (\mathtt{x}_\alpha \circ p|_{O_1}^{-1})^{-1} \\
&= \mathtt{x}_\beta \circ p|_{O_2}^{-1} \circ p|_{O_1} \circ \mathtt{x}_\alpha^{-1} \\
&= \mathtt{x}_\beta \circ \alpha_g \circ \mathtt{x}_\alpha^{-1}
\end{aligned}$$

which is C^r. The rest of the proof is straight forward and is left as an exercise.

Remark 2.5 *In the above, we have suppressed some of information about domains. For example, what is the domain and range of* $\mathtt{x}_\beta \circ p|_{O_2}^{-1} \circ p|_{O_1} \circ \mathtt{x}_\alpha^{-1}$ *? For completeness and in order to help the reader interpret the composition we write out the sequence of domains:*

$$\mathtt{x}_\alpha(U_\alpha \cap \alpha_{g^{-1}} U_\beta) \to U_\alpha \cap \alpha_{g^{-1}} U_\beta \\ \to U_\alpha^* \cap U_\beta^* \to \alpha_g(U_\alpha) \cap U_\beta \to \mathtt{x}_\beta(\alpha_g(U_\alpha) \cap U_\beta).$$

The clutter hides a simple idea that would be better expressed using the idea of a chart germ:

Let $\bar{x}_0 \in U_\alpha^* \cap U_\beta^*$ and consider the composition of germs *of diffeomorphisms* $\mathtt{x}_\beta^* \circ (\mathtt{x}_\alpha^*)^{-1} = \mathtt{x}_\beta \circ p^{-1} \circ (\mathtt{x}_\alpha \circ p^{-1})^{-1} = \mathtt{x}_\beta \circ \alpha_g \circ \mathtt{x}_\alpha^{-1}$.

■

Example 2.20 *We have seen the torus as a differentiable manifold previously presented as $T^2 = S^1 \times S^1$. Another presentation that emphasizes the symmetries is given as follows: Let the group $\mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$ act on $\mathbb{R}^2$ by*

$$(m, n) \times (x, y) \mapsto (x + m, y + n).$$

It is easy to check that proposition 2.5 applies to give a manifold $\mathbb{R}^2/\mathbb{Z}^2$. This is actually the torus and we have a diffeomorphism $\phi : \mathbb{R}^2/\mathbb{Z}^2 \to S^1 \times S^1 = T^2$ given by $[(x, y)] \mapsto (e^{ix}, e^{iy})$. The following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^2 & \to & S^1 \times S^1 \\ \downarrow & \nearrow & \\ \mathbb{R}^2/\mathbb{Z}^2 & & \end{array}$$

Covering spaces $\wp : \widetilde{M} \to M$ that arise from a properly discontinuous group action are special in that if M is connected then the covering is a normal cover?ing.

2.7 Grassmannian manifolds

A very useful generalization of the projective spaces is the Grassmannian manifolds. Let $G_{n,k}$ denote the set of k-dimensional subspaces of $\mathbb{R}^n$. We will exhibit a natural differentiable structure on this set. The idea here is the following. An alternative way of defining the points of projective space which is as equivalence classes of $n-$tuples $(v^1, ..., v^n) \in \mathbb{R}^n - \{0\}$ where $(v^1, ..., v^n) \sim (\lambda v^1, ..., \lambda v^n)$ for any nonzero. This is clearly just a way of specifying a line through the origin. Generalizing, we shall represent a $k-$plane as a matrix whose column vectors span the $k-$ plane. Thus we are putting an equivalence relation on the set of $n \times k$ matrices where $A \sim Ag$ for any nonsingular $k \times k$ matrix g.

To describe this important example we start with the set $\mathbb{M}_{n\times k}$ of $n \times k$ matrices with rank $k < n$ (maximal rank). The columns of each matrix from $\mathbb{M}_{n\times k}$ span a k-dimensional subspace of $\mathbb{R}^n$. Define two matrices from $\mathbb{M}_{n\times k}$ to be equivalent if they span the same k-dimensional subspace. Thus the set $G(k, n)$ of equivalence classes is in one to one correspondence with the set of real k dimensional subspaces of $\mathbb{R}^n$.

Now let U be the set of all $[A] \in G(k, n)$ such that A has its first k rows linearly independent. This property is independent of the representative A of the equivalence class $[A]$ and so U is a well defined set. This last fact is easily proven by a Gaussian column reduction argument. Now every element $[A] \in U \subset G(k, n)$ is an equivalence class that has a unique member A_0 of the form

$$\begin{pmatrix} I_{k\times k} \\ Z \end{pmatrix}.$$

Thus we have a map on U defined by $\Psi : [A] \mapsto Z \in \mathbb{M}_{(n-k)\times k} \cong \mathbb{R}^{k(n-k)}$. We wish to cover $G(k, n)$ with sets U_σ similar to U and defined similar maps. Let $\sigma_{i_1\ldots i_k}$ be the shuffle permutation that puts the k columns indexed by $i_1, ..., i_k$ into the positions $1, ..., k$ without changing the relative order of the remaining columns. Now consider the set $U_{i_1\ldots i_k}$ of all $[A] \in G(k, n)$ such that any representative A has its k rows indexed by $i_1, ..., i_k$ linearly independent. The the permutation induces an obvious 1-1 onto map $\widetilde{\sigma_{i_1\ldots i_k}}$ from $U_{i_1\ldots i_k}$ onto $U = U_{1\ldots k}$. We now have maps $\Psi_{i_1\ldots i_k} : U_{i_1\ldots i_k} \to \mathbb{M}_{(n-k)\times k} \cong \mathbb{R}^{k(n-k)}$ given by composition $\Psi_{i_1\ldots i_k} := \Psi\circ\widetilde{\sigma_{i_1\ldots i_k}}$. These maps form an atlas $\{\Psi_{i_1\ldots i_k}, U_{i_1\ldots i_k}\}$ for $G(k, n)$ which turns out to be a holomorphic atlas (biholomorphic transition maps) and so gives $G(k, n)$ the structure of a smooth manifold called the **Grassmannian manifold** of real k-planes in $\mathbb{R}^n$. Try?graph coordinates!

2.8 Partitions of Unity

A partition of unity is a technical tool that is used quite often in connection with constructing tensor fields, connections, metrics and other objects out of local data. We will not meet tensor fields for a while and the reader may wish to postpone a detailed reading of the proofs in this section until we come to our first use of partitions of unity and/or so called “bump functions”. Partitions of unity are also used in proving the existence of immersions and embeddings; topics we will also touch on later.

It is often the case that we are able to define some object or operation locally and we wish to somehow “glue together” the local data to form a globally defined object. The main and possible only tool for doing this is the partition of unity. For differential geometry it is a smooth partition of unity that we need.

Definition 2.25 *The support of a smooth function is the closure of the set in its domain where it takes on nonzero values. The support of a function f is denoted* $\mathrm{supp}(f)$.

One of the basic ingredients we will need is the so called "bump function" a special case of which was defined in section 7. A bump function is basically a smooth function with support inside some prescribed open set. Notice that this would not in general be possible for a complex analytic function.

Lemma 2.1 (Existence of bump functions) *Let K be a compact subset of $\mathbb{R}^n$ and U an open set containing K. There exists a smooth function β on $\mathbb{R}^n$ which is identically equal to 1 on K, has compact support in U and such that $0 \leq \beta \leq 1$.*

Proof. Special case: Assume that $U = B(0, R)$ and $K = \overline{B}(0, r)$. In this case we may take

$$\phi(x) = \frac{\int_{|x|}^{R} g(t)dt}{\int_{r}^{R} g(t)dt}$$

where

$$g(t) = \begin{cases} e^{-(t-r)^{-1}} e^{-(t-R)^{-1}} & \text{if } 0 < t < R \\ 0 & \text{otherwise.} \end{cases}$$

General case: Let $K \subset U$ be as in the hypotheses. Let $K_i \subset U_i$ be concentric balls as in the special case above but with various choices of radii and such that $K \subset \cup K_i$ and the U_i chosen small enough that $U_i \subset U$. Let ϕ_i be the corresponding functions provided in the proof of the special case. By compactness there are only a finite number of pairs $K_i \subset U_i$ needed so assume that this reduction to a finite cover has been made. Examination of the following function will convince the reader that it is well defined and provides the needed bump function;

$$\beta(x) = 1 - \prod_i (1 - \phi_i(x)).$$

■

A **refinement** of an open cover $\{U_\beta\}_{\beta \in B}$ of a topological space is another $\{V_i\}_{i \in I}$ open cover such that every open set from the second cover is contain in at least one open set from the original cover. This means that means that if $\{U_\beta\}_{\beta \in B}$ is the given cover of X, then a refinement is a cover $\{V_i\}_{i \in I}$ and a set map $I \to B$ of the index sets $i \mapsto \beta(i)$ such that $V_i \subset U_{\beta(i)}$. We say that $\{V_i\}_{i \in I}$ is a **locally finite** cover if in every point of X has a neighborhood that intersects only a finite number of the sets from $\{V_i\}_{i \in I}$. In other words, the every V_i is contained in some U_β and the new cover has only a finite number of member sets that are "near" any given point.

Definition 2.26 *A topological space X is called* ***paracompact*** *if it is Hausdorff and if every open cover of X has a refinement to a locally finite cover.*

Definition 2.27 *A* ***base*** *(or basis) for the topology of a topological space X is a collection of open $\mathfrak{B}$ sets such that all open sets from the topology $\mathfrak{T}$ are unions of open sets from the family$\mathfrak{B}$. A topological space is called* ***second countable*** *if its topology has a countable base.*

Definition 2.28 *A topological space is called* ***locally convex*** *if every point has a neighborhood with compact closure.*

Note that a finite dimensional differentiable manifold is always locally compact and we have agreed that a finite dimensional manifold should by assumed Hausdorff unless otherwise stated. The following lemma is sometimes helpful. It shows that we can arrange to have the open sets of a cover and a locally refinement of the cover to be indexed by the same set in a consistent way:

Lemma 2.2 *If X is a paracompact space and $\{U_i\}_{i\in I}$ is an open cover, then there exists a locally finite refinement $\{O_i\}_{i\in I}$ of $\{U_i\}_{i\in I}$ with $O_i \subset U_i$.*

Proof. Let $\{V_k\}_{i\in K}$ be a locally finite refinement of $\{U_i\}_{i\in I}$ with the index map $k \mapsto i(k)$. Let O_i be the union of all V_k such that $i(k) = k$. Notice that if an open set U intersects an infinite number of the O_i then it will meet an infinite number of the V_k. It follows that $\{O_i\}_{i\in I}$ is locally finite. ■

Theorem 2.4 *A second countable, locally compact Hausdorff space X is paracompact.*

Sketch of proof. If follows from the hypotheses that there exists a sequence of open sets $U_1, U_2, ...$which cover X and such that each U_i has compact closure $\overline{U_i}$. We start an inductive construction: Set $V_n = U_1 \cup U_2 \cup ... \cup U_n$ for each positive integer n. Notice that $\{V_n\}$ is a new cover of X and each V_n has compact closure. Now let $O_1 = V_1$. Since $\{V_n\}$ is an open cover and $\overline{O_1}$ is compact we have

$$\overline{O_1} \subset V_{i_1} \cup V_{i_2} \cup ... \cup V_{i_k}.$$

Next put $O_2 = V_{i_1} \cup V_{i_2} \cup ... \cup V_{i_k}$ and continue the process. Now we have the X is the countable union of these open sets $\{O_i\}$ and each O_{i-1} has compact closure in O_i. Now we define a sequence of compact sets; $K_i = \overline{O_i} \setminus O_{i-1}$.
Now if $\{W_\beta\}_{\beta\in B}$ is *any* open cover of X we can use those W_β which meet K_i to cover K_i and then reduce to a finite subcover since K_i is compact. We can arrange that this cover of K_i consists only of sets each of which is contained in one of the sets $W_\beta \cap O_{i+1}$ and disjoint from O_{i-1}. Do this for all K_i and collect all the resulting open sets into a countable cover for X. This is the desired locally finite refinement. ■

Definition 2.29 *A C^r partition of unity on a C^r manifold M is a collection $\{V_i, \rho_i\}$ where*

1. $\{V_i\}$ is a locally finite cover of M;
2. each ρ_i is a C^r function with $\rho_i \geq 0$ and compact support contained in V_i;
3. for each $x \in M$ we have $\sum \rho_i(x) = 1$ (This sum is finite since $\{V_i\}$ is locally finite.)

If the cover of M by chart map domains $\{U_\alpha\}$ of some atlas $\mathcal{A} = \{U_\alpha, \mathtt{x}_\alpha\}$ of M has a partition of unity $\{V_i, \rho_i\}$ such that each V_i is contained in one of the chart domains $U_{\alpha(i)}$ (locally finite refinement), then we say that $\{V_i, \rho_i\}$ is **subordinate to** $\mathcal{A}$. We will say that a manifold **admits a smooth partition of unity** if every atlas has a subordinate smooth partition of unity.

Notice that in theorem 2.4 we have proven a bit more than is part of the definition of paracompactness. Namely, the open sets of the refinement $V_i \subset U_{\beta(i)}$ have compact closure in $U_{\beta(i)}$.

Theorem 2.5 *Every second countable finite dimensional C^r manifold admits a C^r partition of unity.*

Let M be the manifold in question. We have seen that the hypotheses imply paracompactness and that we may choose our locally finite refinements to have the compact closure property mentioned above. Let $\mathcal{A} = \{U_i, \mathtt{x}_i\}$ be an atlas for M and let $\{W_i\}$ be a locally finite refinement of the cover $\{U_i\}$ with $\overline{W}_i \subset U_i$. By lemma 2.1 above there is a smooth bump function β_i with $\text{supp}(\beta_i) = \overline{W}_i$. For any $x \in M$ the following sum is finite and defines a smooth function:

$$\beta(x) = \sum_i \beta_i(x).$$

Now we normalize to get the needed functions that from the partition of unity:

$$\rho_i = \frac{\beta_i}{\beta}.$$

It is easy to see that $\rho_i \geq 0$, and $\sum \rho_i = 1$.

2.9 Manifolds with boundary.

For the general Stokes theorem where the notion of flux has its natural setting we will need to have a concept of a manifold with boundary . A basic example to keep in mind the closed hemisphere S^2_+ which is the set of all $(x, y, z) \in S^2$ with $z \geq 0$.

Let $\lambda \in \mathbb{R}^{n*}$ be a continuous from on aEuclidean space $\mathbb{R}^n$. In the case of $\mathbb{R}^n$ it will be enough to consider projection onto the first coordinate x^1. Now let $\mathbb{R}^{n+}_\lambda = \{\mathsf{x} \in \mathbb{R}^n \colon \lambda(x) \geq 0\}$ and $\mathbb{R}^{n-}_\lambda = \{\mathsf{x} \in \mathbb{R}^n \colon \lambda(x) \leq 0\}$ and $\partial\, \mathbb{R}^{n+}_\lambda = \partial\, \mathbb{R}^{n-}_\lambda = \{\mathsf{x} \in \mathbb{R}^n \colon \lambda(x) = 0\}$ is the kernel of λ. Clearly $\mathbb{R}^{n+}_\lambda$ and $\mathbb{R}^{n-}_\lambda$ are homeomorphic and $\partial\, \mathbb{R}^{n+}_\lambda$ is a closed subspace. The space $\mathbb{R}^{n-}_\lambda$ is the model space for a manifold with boundary and is called a (negative) **half space**.

Remark 2.6 *We have chosen the space $\mathbb{R}^{n-}_\lambda$ rather than $\mathbb{R}^{n-}_\lambda$ on purpose. The point is that later we will wish to have simple system whereby one of the coordinate vectors $\frac{\partial}{\partial x^i}$ will always be outward pointing at $\partial\, \mathbb{R}^{n-}_\lambda$ while the remaining coordinate vectors in their given order are positively oriented on $\partial\, \mathbb{R}^{n-}_\lambda$ in a sense we will define later. Now, $\frac{\partial}{\partial x^1}$ is outward pointing for $\mathbb{R}^n_{x^1 \leq 0}$ but not for*

$\mathbb{R}^n_{x^1\geq 0}$. *One might be inclined to think that we should look at* $\mathbb{R}^n_{x^j\geq 0}$ *for some other choice of* j *- the most popular being the last coordinate* x^n *but although this could be done it would actually only make things more complicated. The problem is that if we declare* $\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, ..., \frac{\partial}{\partial x^{n-1}}$ *to be positively oriented on* $\mathbb{R}^{n-1}\times 0$ *whenever* $\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, ..., \frac{\partial}{\partial x^n}$ *is positively oriented on* $\mathbb{R}^n$ *we will introduce a minus sign into Stokes' theorem in every other dimension!*

We let $\mathbb{R}^{n-}_\lambda$ (and $\mathbb{R}^{n+}_\lambda$) have the relative topology. Since $\mathbb{R}^{n-}_\lambda \subset \mathbb{R}^n$ we already have a notion of differentiability on $\mathbb{R}^{n-}_\lambda$ (and hence $\mathbb{R}^{n+}_\lambda$) via definition **??**. The notions of C^r maps of open subsets of half space and diffeomorphisms etc. is now defined in the obvious way. For convenience let us define for an open set $U \subset \mathbb{R}^{n-}_\lambda$ (always relatively open) the following slightly inaccurate notations let ∂U denote $\mathbb{R}^{n-}_\lambda \cap U$ and $int(U)$ denote $U \setminus \partial U$.

We have the following three facts:

1. First, it is an easy exercise to show that if $f : U \subset \mathbb{R}^n \to \mathbb{R}^d$ is C^r differentiable (with $r \geq 1$) and g is another such map with the same domain, then if $f = g$ on $\mathbb{R}^{n-}_\lambda \cap U$ then $D_{\mathsf{x}} f = D_{\mathsf{x}} g$ for all $\mathsf{x} \in \mathbb{R}^{n-}_\lambda \cap U$.

2. Let $\mathbb{R}^{d+}_\ell$ be a half space in aEuclidean space $\mathbb{R}^d$. If $f : U \subset \mathbb{R}^n \to \mathbb{R}^{d-}_\alpha$ is C^r differentiable (with $r \geq 1$) and $f(\mathsf{x}) \in \partial\, \mathbb{R}^{d-}_\ell$ then $D_{\mathsf{x}} f : \mathbb{R}^n \to \mathbb{R}^d$ must have its image in $\partial\, \mathbb{R}^{d-}_\ell$.

3. Let $f : U_1 \subset \mathbb{R}^{n-}_\lambda \to U_2 \subset \mathbb{R}^{d-}_\ell$ be a diffeomorphism (in our new extended sense). Assume that $\mathbb{R}^{n-}_\lambda \cap U_1$ and $\mathbb{R}^{d-}_\ell \cap U_2$ are not empty. Then f induces diffeomorphisms $\partial U_1 \to \partial U_2$ and $int(U_1) \to int(U_2)$.

These three claims are not exactly obvious but there are very intuitive. On the other hand, none of them are difficult to prove and we will leave these for exercises (actually the proof of 3 is more or less obvious from the proof of theorem). These facts show that the notion of a boundary defined here and in general below is a well defined concept and is a natural notion in the context of differentiable manifolds; it is a "differentiable invariant".

We can now form a definition of manifold with boundary in a fashion completely analogous to the definition of a manifold without boundary. A **half space chart** $\mathtt{x}_\alpha$ for a set M is a bijection of some subset U_α of M onto an open subset of $\mathbb{R}^{n-}_\lambda$ (or $\mathbb{R}^{n+}_\lambda$ for many authors). A C^r **half space atlas** is a collection $(\mathtt{x}_\alpha, U_\alpha)$ of such half space charts such that for any two, say $(\mathtt{x}_\alpha, U_\alpha)$ and $(\mathtt{x}_\beta, U_\beta)$, the map $\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}$ is a C^r diffeomorphism on its natural domain (if non-empty). Note: "Diffeomorphism" means in the extended the sense of a being homeomorphism and such that both $\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1} :: \mathbb{R}^{n-}_\lambda \to \mathbb{R}^n$ and its inverse are C^r in the sense of definition 2.5.

Example 2.21 *Review the section on pseudogroups were we defined manifold with boundary in that context. Is the definition there the same as that given here? 2.4*

Overlap of boundary charts.

Definition 2.30 *A C^r-**manifold with boundary** $M, \mathcal{A}$ is a set M together with a maximal atlas of half space charts $\mathcal{A}$. The manifold topology is that generated by the domains of all such charts. The boundary of M is denoted by ∂M and is the set of point which have images in the boundary $\mathbb{R}_0^n$ of $\mathbb{R}_\lambda^{n-}$ under some and hence every chart.*

Definition 2.31 *The interior of a manifold with boundary is a manifold without boundary and is denoted $\overset{\circ}{M}$. The manifold $\overset{\circ}{M}$ is never compact and is referred to as an open manifold.*

Definition 2.32 *In our present context, a manifold without boundary which is compact (and hence closed in the usual topological sense if M is Hausdorff) is called a **closed manifold**. If no component of a manifold without boundary is compact it is called an **open manifold**.*

Remark 2.7 *The phrase "closed manifold" is a bit problematic since the word closed is acting as an adjective and so conflicts with the notion of closed in the ordinary topological sense. For this reason we will try to avoid this terminology and use instead the phrase "**compact manifold without boundary**".*

Exercise 2.8 *Show that $M \cup \partial M$ is closed and that $M - \partial M$ is open.*

Remark 2.8 (Warning) *Some authors let M denote the interior, so that $M \cup \partial M$ is the closure and is the manifold with boundary in our sense.*

Theorem 2.6 *∂M is a C^r manifold (without boundary) with an atlas being given by all maps of the form $\mathtt{x}_\alpha|, U_\alpha \cap \partial M$. The manifold ∂M is called the* ***boundary of*** *M.*

Idea of Proof. The truth of this theorem becomes obvious once we recall what it means for a chart overlap map $\mathtt{y} \circ \mathtt{x}^{-1} : U \to V$ to be a diffeomorphism in a neighborhood a point $x \in U \cap \mathbb{R}^{n+}_{\lambda}$. First of all there must be a set U' containing U which is open in $\mathbb{R}^n$ and an extension of $\mathtt{y} \circ \mathtt{x}^{-1}$ to a differentiable map on U'. But the same is true for $(\mathtt{y} \circ \mathtt{x}^{-1})^{-1} = \mathtt{x} \circ \mathtt{y}^{-1}$. The extensions are inverses of each other on U and V. But we must also have that the derivatives maps of the transition maps are isomorphisms at all points up to and including ∂U and ∂V. But then the inverse function theorem says that there are neighborhoods of points in ∂U in $\mathbb{R}^n$ and ∂V in $\mathbb{R}^n$ such that these extensions are actually diffeomorphisms and inverses of each other. Now it follows that the restrictions $\mathtt{y} \circ \mathtt{x}^{-1}\big|_{\partial U} : \partial U \to \partial V$ are diffeomorphisms. In fact, this is the main point of the comment (3) above and we have now seen the idea of its proof also. ■

Example 2.22 *The closed ball $\overline{B}(p, R)$ in $\mathbb{R}^n$ is a manifold with boundary $\partial \overline{B}(p, R) = S^{n-1}$.*

Example 2.23 *The hemisphere $S^n_+ = \{x \in \mathbb{R}^{n+1} : x^{n+1} \geq 0\}$ is a manifold with boundary.*

Exercise 2.9 *Is the Cartesian product of two manifolds with boundary a manifold with boundary?*

Chapter 3

The Tangent Structure

3.1 Rough Ideas II

Let us suppose that we have two coordinate systems $\mathtt{x} = (x^1, x^2, ...x^n)$ and $\mathtt{y} = (y^1, y^2, ...y^n)$ defined on some common open set of a differentiable manifold M as defined above in 26.88. Let us also suppose that we have two lists of numbers $v^1, v^2, ..., v^n$ and $\bar{v}^1, \bar{v}^2, ...\bar{v}^n$ somehow coming from the respective coordinate systems and associated to a point p in the common domain of the two coordinate systems. Suppose that the lists are related to each other by

$$v^i = \sum_{k=1}^{n} \frac{\partial x^i}{\partial y^k} \bar{v}^k$$

where the derivatives $\frac{\partial x^i}{\partial y^k}$ are evaluated at the coordinates $y^1(p), y^2(p), ..., y^n(p)$. Now if f is a function also defined in a neighborhood of p then the representative functions for f in the respective systems are related by

$$\frac{\partial f}{\partial x^i} = \sum_{k=1}^{n} \frac{\partial \bar{x}^k}{\partial x^i} \frac{\partial f}{\partial \bar{x}^k}.$$

The chain rule then implies that

$$\frac{\partial f}{\partial x^i} v^i = \frac{\partial f}{\partial \bar{x}^i} \bar{v}^i.$$

Thus if we had a list $v^1, v^2, ..., v^n$ for every coordinate chart on the manifold whose domains contain the point p and related to each other as above then we say that we have a tangent vector v at $p \in M$. It then follows that if we define the **directional derivative** of a function f at p in the direction of v by

$$vf := \frac{\partial f}{\partial x^i} v^i$$

then we are in business since it doesn't matter which coordinate system we use. Because of this we think of (v^i) and $(\bar{v}^i)$ as representing the same geometric object (a tangent vector at p). Where do we get such vectors in a natural way? Well one good way is from the notion of the velocity of a curve. A differentiable curve though $p \in M$ is a map $c : (-a, a) \to M$ with $c(0) = p$ such that the coordinate expressions for the curve $x^i(t) = (x^i \circ c)(t)$ are all differentiable. We then take

$$v^i := \frac{dx^i}{dt}(0)$$

for each coordinate system $\mathtt{x} = (x^1, x^2, ...x^n)$ with p in its domain. This gives a well defined tangent vector v at p called the velocity of c at $t = 0$. We denote this by $c'(0)$ or by $\frac{dc}{dt}(0)$.Of course we could have done this for each $t \in (-a, a)$ by defining $v^i(t) := \frac{dx^i}{dt}(t)$ and we would get a smoothly varying family of velocity vectors $c'(t)$ defined at the points $c(t) \in M$.

If we look at the set of all tangent vectors at a point $p \in M$ we get a vector space since we can clearly choose a coordinate system in which to calculate and then the vectors just appear as $n-$tuples; that is, elements of R^n. The vector space operations (scaling and vector addition) remain consistently related when viewed in another coordinate system since the relations are linear. The set of all tangent vectors at $p \in M$ is called the tangent space at p. We will denote these by T_pM for the various points p in M. The tangent spaces combine to give another differentiable manifold of twice the dimension. The coordinates come from those that exist on M already by adding in the "point" coordinates $(x^1, ..., x^n)$ the components of the vectors. Thus the coordinates of a tangent vector v at p are simply $(x^1, ..., x^n; v^1, ..., v^n)$.

3.2 Tangent Vectors

For a submanifold S of $\mathbb{R}^n$ we have a good idea what a tangent vector ought to be. Let $t \mapsto c(t) = (x^1(t), ..., x^n(t))$ be a curve with image contained in S and passing through the point $p \in S$ at time $t = 0$. Then the vector $v = \dot{c}(t) = \frac{d}{dt}\big|_{t=0} c(t)$ is tangent to S at p. So to be **tangent** to S at p is to be the velocity at p of some curve in S through p. Of course, we must consider v to be *based* at $p \in S$ in order to distinguish it from parallel vectors of the same length that may be velocity vectors of curves going through other points. One way to do this is to write the tangent vector as a pair (p, v) where p is the base point. In this way we can construct the space TS of all vectors tangent to S as a subset of $\mathbb{R}^n \times \mathbb{R}^n$

$$TS = \{(p, v) \in \mathbb{R}^n \times \mathbb{R}^n : p \in S \text{ and } v \text{ tangent to } S \text{ at } p\}$$

This method will not work well for manifolds that are not given as submanifolds of $\mathbb{R}^n$. We will now give three methods of defining tangent vectors at a point of a differentiable manifold.

Definition 3.1 (Tangent vector via charts) *Consider the set of all admissible charts* $(\mathtt{x}_\alpha, U_\alpha)_{\alpha\in A}$ *on* M *indexed by some set* A *for convenience. Next consider the set* T *of all triples* (p, v, α) *such that* $p \in U_\alpha$. *Define an equivalence relation so that* $(p, \mathsf{v}, \alpha) \sim (q, \mathsf{w}, \beta)$ *iff* $p = q$ *and*

$$D(\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}(p)} \cdot \mathsf{v} = \mathsf{w}$$

In other words, the derivative at $\mathtt{x}(p)$ *of the coordinate change* $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ *"identifies"* v *with* w. *Tangent vectors are then equivalence classes with the tangent vectors at a point* p *being those equivalence classes represented by triples with first slot occupied by* p. *The set of all tangent vectors at* p *is written as* T_pM. *The* ***tangent bundle*** TM *is the disjoint union of all the tangent spaces for all points in* M.

$$TM := \bigsqcup_{p\in M} T_pM$$

This viewpoint takes on a more familiar appearance in finite dimensions if we use a more classical notation; Let $(\mathtt{x}, U)$ and $(\mathtt{y}, V)$ two charts containing p in there domains. If an $n-$tuple $(v^1, ..., v^n)$ represents a tangent vector at p from the point of view of $(\mathtt{x}, U)$ and if the $n-$tuple $(w^1, ..., w^n)$ represents the same vector from the point of view of $(\mathtt{y}, V)$ then

$$w^i = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j}\bigg|_{\mathtt{x}(p)} v^j$$

where we write the change of coordinates as $y^i = y^i(x^1, ..., x^n)$ with $1 \leq i \leq n$.

We can get a similar expression in the infinite dimensional case by just letting $D(\mathtt{y} \circ \mathtt{x}^{-1})\big|_{\mathtt{x}(p)}$ be denoted by $\frac{\partial \mathtt{y}}{\partial \mathtt{x}}\big|_{\mathtt{x}(p)}$ then we write

$$\mathsf{w} = \frac{\partial \mathtt{y}}{\partial \mathtt{x}}\bigg|_{\mathtt{x}(p)} \mathsf{v}.$$

Definition 3.2 (Tangent vectors via curves) *Let* p *be a point in a* C^r *manifold with* $k > 1$. *Suppose that we have* C^r *curves* c_1 *and* c_2 *mapping into manifold* M, *each with open domains containing* $0 \in \mathbb{R}$ *and with* $c_1(0) = c_2(0) = p$. *We say that* c_1 *is tangent to* c_2 *at* p *if for all* C^r *functions* $f : M \to \mathbb{R}$ *we have* $\frac{d}{dt}\big|_{t=0} f \circ c_1 = \frac{d}{dt}\big|_{t=0} f \circ c_2$. *This is an equivalence relation on the set of all such curves. Define a* ***tangent vector at*** p *to be an equivalence class* $X_p = [c]$ *under this relation. In this case we will also write* $\dot{c}(0) = X_p$. *The* ***tangent space*** T_pM *is defined to be the set of all tangents vectors at* $p \in M$. *The* ***tangent bundle*** TM *is the disjoint union of all the tangent spaces for all points in* M.

$$TM := \bigsqcup_{p\in M} T_pM$$

The tangent bundle is actually a differentiable manifold itself as we shall soon see.

If $X_p \in T_pM$ for p in the domain of an admissible chart $(U_\alpha, \mathtt{x}_\alpha)$. In this chart X_p is represented by a triple (p, v, α). We denote by $[X_p]_\alpha$ the principle part v of the representative of X_p. Equivalently, $[X_p]_\alpha = D(\mathtt{x}_\alpha \circ c)|_0$ for any c with $c'(0) = X_p$ i.e. $X_p = [c]$ as in definition 3.2.

For the next definition of tangent vector we need to think about the set of functions defined near a point. We want a formal way of considering two functions that agree on some open set containing a point as being locally the same at that point. To this end we take the set F_p of all smooth functions with open domains of definition containing $p \in M$. Define two such functions to be equivalent if they agree on some small open set containing p. The equivalence classes are called **germs of smooth functions at** p and the set of all such is denoted $\mathcal{F}_p = F_p/\sim$. It is easily seen that $\mathcal{F}_p$ is naturally a vector space and we can even multiply germs in the obvious way. This makes $\mathcal{F}_p$ a ring (and an algebra over $\mathbb{R}$). Furthermore, if f is a representative for the equivalence class $\breve{f} \in \mathcal{F}_p$ then $\breve{f}(p) = f(p)$ is well defined and so we have an **evaluation map** $ev_p : F_p \to R$. We are really just thinking about functions defined near a point and the germ formalism is convenient whenever we do something where it only matters what is happening near p. We will thus sometimes abuse notation and write f instead of $\breve{f}$ to denote the germ represented by a function f. In fact, we don't really absolutely need the germ idea for the following kind of definition to work so we put the word "germs" in parentheses.

We have defined $\mathcal{F}_p$ using smooth functions but we can also define in an obvious way $\mathcal{F}_p^r$ using C^r functions.

Definition 3.3 *Let $\breve{f}$ be the germ of a function $f :: M \to \mathbb{R}$. Let us define the* ***differential of*** f ***at*** p *to be a map $df(p) : T_pM \to \mathbb{R}$ by simply composing a curve c representing a given vector $X_p = [c]$ with f to get $f \circ c :: \mathbb{R} \to \mathbb{R}$. Then define $df(p) \cdot X_p = \frac{d}{dt}\big|_{t=0} f \circ c \in \mathbb{R}$. Clearly we get the same answer if we use another function with the same germ at p. The differential at p is also often written as $df|_p$. More generally, if $f :: M \to \mathsf{E}$ for someEuclidean space E then $df(p) : T_pM \to \mathsf{E}$ is defined by the same formula.*

It is easy to check that $df(p) : T_pM \to \mathsf{E}$ the composition of the tangent map T_pf defined below and the canonical map $T_\mathsf{y}\mathsf{E} \cong \mathsf{E}$ where $\mathsf{y} = f(\mathsf{p})$. Diagrammatically we have

$$df(p) : T_pM \stackrel{Tf}{\to} T\mathsf{E} = \mathsf{E} \times \mathsf{E} \stackrel{pr_1}{\to} \mathsf{E}.$$

Remark 3.1 (Very useful notation) *This use of the "differential" notation for maps into vector spaces is useful for coordinates expressions. Let $p \in U$ where $(\mathtt{x}, U)$ is a chart and consider again a tangent vector v at p. Then the local representative of v in this chart is exactly $d\mathtt{x}(v)$.*

Definition 3.4 *A derivation of the algebra $\mathcal{F}_p$ is a map $\mathcal{D} : \mathcal{F}_p \to \mathbb{R}$ such that $\mathcal{D}(\breve{f}\breve{g}) = \breve{f}(p)\mathcal{D}\breve{g} + \breve{g}(p)\mathcal{D}\breve{f}$ for all $\breve{f}, \breve{g} \in \mathcal{F}_p$.*

Notation 3.1 *The set of all derivations on $\mathcal{F}_p$ is easily seen to be a real vector space and we will denote this by $\mathsf{Der}(\mathcal{F}_p)$.*

We will now define the operation of a tangent vector on a function or more precisely, on germs of functions at a point.

Definition 3.5 *Let $\mathcal{D}_{X_p} : \mathcal{F}_p \to \mathbb{R}$ be given by the rule $\mathcal{D}_{X_p}\breve{f} = df(p) \cdot X_p$.*

Lemma 3.1 *$\mathcal{D}_{X_p}$ is a derivation of the algebra $\mathcal{F}_p$. That is we have $\mathcal{D}_{X_p}(\breve{f}\breve{g}) = \breve{f}(p)\mathcal{D}_{X_p}\breve{g} + \breve{g}(p)\mathcal{D}_{X_p}\breve{f}$.*

A basic example of a derivation is the partial derivative operator $\frac{\partial}{\partial x^i}\big|_{x_0} : f \mapsto \frac{\partial f}{\partial x^i}(x_0)$. We shall show that for a smooth n-manifold these form a basis for the space of all derivations at a point $x_0 \in M$. This vector space of all derivations is naturally isomorphic to the tangent space at x_0. Since this is certainly a local problem it will suffice to show this for $x_0 \in \mathbb{R}^n$.

Notation 3.2 *In the literature $\mathcal{D}_{X_p}\breve{f}$ is written $X_p f$ and we will also use this notation. As indicated above, if M is finite dimensional and C^∞ then all derivations of $\mathcal{F}_p$ are given in this way by tangent vectors. Thus in this case (and not for Banach manifolds of lower differentiability) we could abbreviate $\mathcal{D}_{X_p} f = X_p f$ and* define *tangent vector to be derivations. For this we need a couple of lemmas:*

Lemma 3.2 *If c is (the germ of) a constant function then $\mathcal{D}c = 0$.*

Proof. Since $\mathcal{D}$ is linear this is certainly true if $c = 0$. Also, linearity shows that we need only prove the result for $c = 1$. Then

$$\begin{aligned}\mathcal{D}1 &= \mathcal{D}(1^2) \\ &= (\mathcal{D}1)c + 1\mathcal{D}1 = 2\mathcal{D}1\end{aligned}$$

and so $\mathcal{D}1 = 0$. ∎

Lemma 3.3 *Let $f :: \mathbb{R}^n, x_0 \to \mathbb{R}, f(x_0)$ be defined and C^∞ in a neighborhood of x_0. Then near x_0 we have*

$$f(x) = f(x_0) + \sum_{1 \le i \le n} (x^i - x_0^i)\left[\frac{\partial f}{\partial x^i}(x_0) + a^i(x)\right]$$

for some smooth functions $a^i(x)$ with $a^i(x_0) = 0$.

Proof. Write $f(x) - f(x_0) = \int_0^1 \frac{\partial}{\partial t}\left[f(x_0 + t(x - x_0)\right] dt = \sum_{i=1}^n (x^i - x_0^i) \int_0^1 \frac{\partial f}{\partial x^i}(x_0 + t(x - x_0))dt$. Integrate the last integral by parts to get

$$\begin{aligned}
&\int_0^1 \frac{\partial f}{\partial x^i}\left[(x_0 + t(x - x_0))\right] dt \\
&= t\frac{\partial f}{\partial x^i}\left[(x_0 + t(x - x_0))\right]\Big|_0^1 - \int_0^1 t\sum_{i=1}^n (x^i - x_0^i)\frac{\partial^2}{\partial x^i \partial x^j}(x_0 + t(x - x_0))dt \\
&= \frac{\partial f}{\partial x^i}(x_0) + a^i(x)
\end{aligned}$$

where the term $a^i(x)$ clearly satisfies the requirements. ∎

Proposition 3.1 *Let $\mathcal{D}_{x_0}$ be a derivation on $\mathcal{F}_{x_0}$ where $x_0 \in \mathbb{R}^n$. Then*

$$\mathcal{D}_{x_0} = \sum_{i=1}^n \mathcal{D}_{x_0}(x^i) \left.\frac{\partial}{\partial x^i}\right|_{x_0}.$$

In particular, $\mathcal{D}$ corresponds to a unique vector at x_0 and by the association $(\mathcal{D}_{x_0}(x^1), ..., \mathcal{D}_{x_0}(x^n)) \mapsto \mathcal{D}_{x_0}$ we get an isomorphism of $\mathbb{R}^n$ with $\mathsf{Der}(\mathcal{F}_p)$.

Proof. Apply $\mathcal{D}$ to both sides of

$$f(x) = f(x_0) + \sum_{1 \le i \le n} (x^i - x_0^i)\left[\frac{\partial f}{\partial x^i}(x_0) + a^i(x)\right].$$

and use 3.2 to get the formula of the proposition. The rest is easy. ∎

An important point is that the above construction carries over via charts to a similar statement on a manifold. The reason for this is that if $(\mathtt{x}, U)$ is a chart containing a point p in a smooth n manifold then we can define a an isomorphism between $\mathsf{Der}(\mathcal{F}_p)$ and $\mathsf{Der}(\mathcal{F}_{\mathtt{x}(p)})$ by the following simple rule:

$$\begin{aligned}
&\mathcal{D}_{\mathtt{x}(p)} \mapsto \mathcal{D}_p \\
&\mathcal{D}_p f = \mathcal{D}_{\mathtt{x}(p)}(f \circ \mathtt{x}^{-1}).
\end{aligned}$$

The one thing that must be noticed is that the vector $(\mathcal{D}_{\mathtt{x}(p)}(x^1), ..., \mathcal{D}_{\mathtt{x}(p)}(x^n))$ transforms in the proper way under change of coordinates so that the correspondence induces a well defined 1-1 linear map between T_pM and $\mathsf{Der}(\mathcal{F}_p)$. So using this we have one more possible definition of tangent vectors that works on smooth finite dimensional manifolds:

Definition 3.6 (Tangent vectors as derivations) *Let M be a smooth manifold of dimension $n < \infty$. Consider the set of all (germs of) smooth functions $\mathcal{F}_p$ at $p \in M$. A* ***tangent vector*** *at p is a linear map $X_p : \mathcal{F}_p \to \mathbb{R}$ which is also a derivation in the sense that for $f, g \in \mathcal{F}_p$*

$$X_p(fg) = g(p)X_pf + f(p)X_pg.$$

Once again the tangent space at p is the set of all tangent vectors at p and the tangent bundle is define by disjoint union as before.

In any event, even in the general case of a C^r Banach manifold with $r \geq 1$ a tangent vector determines a unique derivation written $X_p : f \mapsto X_p f$. However, in this case the derivation maps $\mathcal{F}_p^r$ to $\mathcal{F}_p^{r-1}$. Also, on infinite dimensional manifolds, even if we consider only the C^∞ case, there may be derivations not coming from tangent vectors as given in definition 3.2 or in definition 3.1.

3.3 Interpretations

We will now show how to move from one definition of tangent vector to the next. For simplicity let us assume that M is a smooth (C^∞) n-manifold.

1. Suppose that we think of a tangent vector X_p as an equivalence class of curves represented by $c : I \to M$ with $c(0) = p$. We obtain a derivation by defining
$$X_p f := \left.\frac{d}{dt}\right|_{t=0} f \circ c$$
We can define a derivation in this way even if M is infinite dimensional but the space of derivations and the space of tangent vectors may not match up.

2. If X_p is a derivation at p and $U_\alpha, \mathtt{x}_\alpha = (x^1, ..., x^n)$ an admissible chart with domain containing p, then X_p,as a tangent vector as in definition 3.1, is represented by the triple (p, v, α) where $\mathsf{v} = (v^1, ...v^n)$ is given by
$$v^i = X_p x^i \text{ (acting as a derivation)}$$

3. Suppose that, a la definition 3.1, a vector $X_p$at $p \in M$ is represented by (p, v, α) where $\mathsf{v} \in \mathbb{R}^n$ and α names the chart $(\mathtt{x}_\alpha, U_\alpha)$. We obtain a derivation by defining
$$X_p f = D(f \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}_\alpha(p)} \cdot \mathsf{v}$$
In case the manifold if modelled on $\mathbb{R}^n$ then we have the more traditional notation
$$X_p f = \sum v^i \left.\frac{\partial}{\partial x^i}\right|_p f.$$
for $\mathsf{v} = (v^1, ...v^n)$.

The notation $\left.\frac{\partial}{\partial x^i}\right|_p$ is made precise by the following:

Definition 3.7 *For a chart* $\mathtt{x} = (x^1, ..., x^n)$ *with domain* U *containing a point* p *we define a tangent vector* $\left.\frac{\partial}{\partial x^i}\right|_p \in T_p M$ *by*

$$\left.\frac{\partial}{\partial x^i}\right|_p f = D_i(f \circ \mathtt{x}^{-1})(\mathtt{x}(p))$$

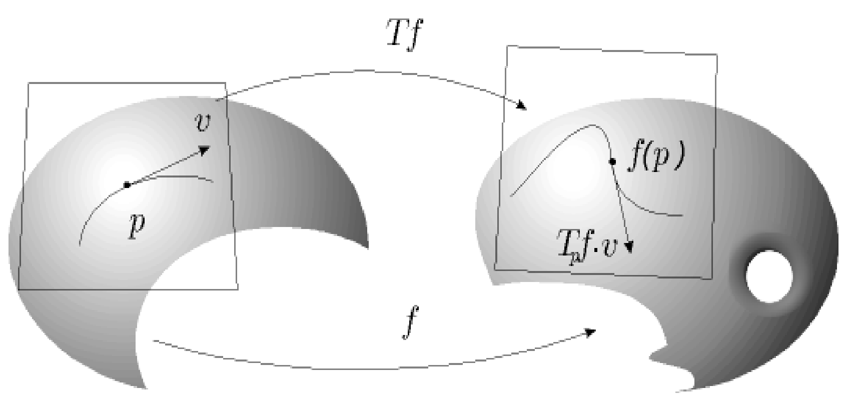
Alternatively, we may take $\left.\frac{\partial}{\partial x^i}\right|_p$ *to be the equivalence class of a coordinate curve. In other words,* $\left.\frac{\partial}{\partial x^i}\right|_p$ *is the velocity at* $\mathtt{x}(p)$ *of the curve* $t \mapsto \mathtt{x}^{-1}(x^1(p), ..., x^i(p)+t, ..., x^n(p))$ *defined for sufficiently small* t*.*

We may also identify $\left.\frac{\partial}{\partial x^i}\right|_p$ *as the vector represented by the triple* $(p, \mathtt{e}_i, \alpha)$ *where* $\mathtt{e}_i$ *is the* i*-th member of the standard basis for* $\mathbb{R}^n$ *and* α *refers to the current chart* $\mathtt{x} = \mathtt{x}_\alpha$*.*

Exercise 3.1 *For a finite dimensional* $C^\infty-$*manifold* M *and* $p \in M$*, let* $\mathtt{x}_\alpha = (x^1, ..., x^n), U_\alpha$ *be a chart whose domain contains* p*. Show that the vectors* $\left.\frac{\partial}{\partial x^1}\right|_p, ..., \left.\frac{\partial}{\partial x^n}\right|_p$ *are a basis for the tangent space* T_pM*.*

3.4 The Tangent Map

The first definition of the tangent map of a map $f : M, p \to N, f(p)$ will be considered our main definition but the others are actually equivalent at least for finite dimensional manifolds. Given f and p as above wish to define a linear map $T_pf : T_pM \to T_{f(p)}N$



Tangent map as understood via curve transfer.

For the next version recall remark 3.1.

Definition 3.8 (Tangent map I) *If we have a smooth function between manifolds*

$$f : M \to N$$

and we consider a point $p \in M$ *and its image* $q = f(p) \in N$ *then we define the* ***tangent map*** *at* p *by choosing any chart* $(\mathtt{x}, U)$ *containing* p *and a chart* $(\mathtt{y}, V)$ *containing* $q = f(p)$ *and then for any* $v \in T_pM$ *we have* *the representative* $d\mathtt{x}(v)$ *with respect to* $(\mathtt{x}, U)$. *Then the representative of* $T_p f \cdot v$ *is given by*

$$d\mathtt{y}(T_p f \cdot v) = D(\mathtt{y} \circ f \circ \mathtt{x}^{-1}) \cdot d\mathtt{x}(v).$$

This uniquely determines $T_p f \cdot v$ *and the chain rule guarantees that this is well defined (independent of the choice of charts).*

Definition 3.9 (Tangent map II) *If we have a smooth function between manifolds*

$$f : M \to N$$

and we consider a point $p \in M$ *and its image* $q = f(p) \in N$ *then we define the* ***tangent map*** *at* p

$$T_p f : T_p M \to T_q N$$

in the following way: Suppose that $v \in T_p M$ *and we pick a curve* c *with* $c(0) = p$ *so that* $v = [c]$, *then by definition*

$$T_p f \cdot v = [f \circ c] \in T_q N$$

where $[f \circ c] \in T_q N$ *is the vector represented by the curve* $f \circ c$.

An alternative definition for finite dimensional smooth manifolds in terms of derivations is the following.

Definition 3.10 (Tangent Map III) *Let* M *be a smooth* n*-manifold. View tangent vectors as derivation as explained above. Then continuing our set up above and letting* g *be a smooth germ at* $q = f(p) \in N$ *we define the derivation* $T_p f \cdot v$ *by*

$$(T_p f \cdot v)g = v(f \circ g)$$

It is easy to check that this defines a derivation on the (germs) of smooth functions at q *and so is also a tangent vector in* $T_q M$. *Thus we get a map* $T_p f$ *called the* ***tangent map*** *(at* p*).*

3.5 The Tangent and Cotangent Bundles

3.5.1 Tangent Bundle

We have defined the **tangent bundle** of a manifold as the disjoint union of the tangent spaces $TM = \bigsqcup_{p \in M} T_p M$. We show in proposition 3.2 below that TM is itself a differentiable manifold but first we record the following two definitions.

Definition 3.11 *Give a smooth map* $f : M \to N$ *as above then the tangent maps on the individual tangent spaces combine to give a map* $Tf : TM \to TN$ *on the tangent bundles that is linear on each fiber called the tangent lift.*

Definition 3.12 *The map $\tau_M : TM \to M$ defined by $\tau_M(v) = p$ for every $p \in T_pM$ is called the (tangent bundle) projection map. The TM together with the map $\tau_M : TM \to M$ is an example of a vector bundle.*

Proposition 3.2 *TM is a differentiable manifold and $\tau_M : TM \to M$ is a smooth map. Furthermore, for a smooth map $f : M \to N$ the tangent map is smooth and the following diagram commutes.*

$$\begin{array}{ccccc} & TM & \stackrel{Tf}{\to} & TN & \\ \tau_M & \downarrow & & \downarrow & \tau_N \\ & M & \stackrel{f}{\to} & N & \end{array}$$

Now for every chart $(\mathtt{x}, U)$ let $TU = \tau_M^{-1}(U)$. The charts on TM are defined using charts from M are as follows

$$\begin{aligned} &T\mathtt{x} : TU \to T\mathtt{x}(TU) \cong \mathtt{x}(U) \times \mathbb{R}^n \\ &T\mathtt{x} : \xi \mapsto (\mathtt{x} \circ \tau_M(\xi), \mathsf{v}) \end{aligned}$$

where $\mathsf{v} = d\mathtt{x}(\xi)$ is the principal part of ξ in the $\mathtt{x}$ chart. The chart $T\mathtt{x}, TU$ is then described by the composition

$$\xi \mapsto (\tau_M(\xi), \xi) \mapsto (\mathtt{x} \circ \tau_M(\xi), d\mathtt{x}(\xi))$$

but $\mathtt{x} \circ \tau_M(\xi)$ it is usually abbreviated to just $\mathtt{x}$ so we may write the chart in the handy form $(\mathtt{x}, d\mathtt{x})$.

$$\begin{array}{ccc} TU & \to & \mathtt{x}(U) \times \mathbb{R}^n \\ \downarrow & & \downarrow \\ U & \to & \mathtt{x}(U) \end{array}$$

For a finite dimensional manifold and with and a chart $\mathtt{x} = (x^1, ..., x^n)$, any vector $\xi \in \tau_M^{-1}(U)$ can be written

$$\xi = \sum v^i(\xi) \left.\frac{\partial}{\partial x^i}\right|_{\tau_M(\xi)}$$

for some $v^i(\xi) \in \mathbb{R}$ depending on ξ. So in the finite dimensional case the chart is just written $(x^1, ..., x^n, v^1, ..., v^n)$.

Exercise 3.2 *Test your ability to interpret the notation by checking that each of these statements makes sense and is true:*

1) If $\xi = \xi^i \left.\frac{\partial}{\partial x^i}\right|_p$ and $\mathtt{x}_\alpha(p) = (a^1, ..., a^n) \in \mathtt{x}_\alpha(U_\alpha)$ then $T\mathtt{x}_\alpha(\xi) = (a^1, ..., a^n, \xi^1, ..., \xi^n) \in U_\alpha \times \mathbb{R}^n$.

2) If $v = [c]$ for some curve with $c(0) = p$ then

$$T\mathtt{x}_\alpha(v) = (\mathtt{x}_\alpha \circ c(0), \left.\frac{d}{dt}\right|_{t=0} \mathtt{x}_\alpha \circ c) \in U_\alpha \times \mathbb{R}^n$$

Suppose that $(\mathtt{x}, d\mathtt{x})$ and $(\mathtt{y}, d\mathtt{y})$ are two such charts constructed as above from two charts $U, \mathtt{x}$ and $V, \mathtt{y}$ and that $U \cap V \neq \emptyset$. Then $TU \cap TV \neq \emptyset$ and on the overlap we have the coordinate transitions $T\mathtt{y} \circ T\mathtt{x}^{-1}$:$(\mathsf{x}, \mathsf{v}) \mapsto (\mathsf{y}, \mathsf{w})$ where

$$\begin{aligned} \mathsf{y} &= \mathtt{y} \circ \mathtt{x}^{-1}(\mathsf{x}) \\ \mathsf{w} &= \sum_{k=1}^{n} D(\mathtt{y} \circ \mathtt{x}^{-1})\big|_{\mathtt{x}(p)} \mathsf{v} \end{aligned}$$

and so the overlaps will be C^{r-1} whenever the $\mathtt{y} \circ \mathtt{x}^{-1}$ are C^r. Notice that for all $p \in \mathtt{x}(U \cap V)$ we have

$$\begin{aligned} \mathtt{y}(p) &= \mathtt{y} \circ \mathtt{x}^{-1}(\mathtt{x}(p)) \\ d\mathtt{y}(\xi) &= D(\mathtt{y} \circ \mathtt{x}^{-1})\big|_{\mathtt{x}(p)} d\mathtt{x}(\xi) \end{aligned}$$

or with our alternate notation

$$d\mathtt{y}(\xi) = \left.\frac{\partial \mathsf{y}}{\partial \mathsf{x}}\right|_{\mathtt{x}(p)} \circ d\mathtt{x}(\xi)$$

and in finite dimensions the classical notation

$$\begin{aligned} y^i &= y^i(x^1, ..., x^n) \\ dy^i(\xi) &= \frac{\partial y^i}{\partial x^k} dx^k(\xi) \\ &\mathit{or} \\ w^i &= \frac{\partial y^i}{\partial x^k} v^k \end{aligned}$$

$$\begin{aligned} \mathsf{y} &= \mathtt{y} \circ \mathtt{x}^{-1}(\mathsf{x}) \\ \mathsf{w} &= \left.\frac{\partial \mathsf{y}}{\partial \mathsf{x}}\right|_{\mathtt{x}(p)} \mathsf{v} \end{aligned}$$

This classical notation may not be logically precise but it is easy to read and understand. Recall our notational principle **??**.

Exercise 3.3 *If M is actually equal to an open subset U of a model space M then it seems that we have (at least) two definitions of TU. How should this difference be reconciled?*

3.5.2 The Cotangent Bundle

Each T_pM has a dual space T_p^*M. In case M is modelled on aEuclidean space $\mathbb{R}^n$ we have $T_pM \approx \mathbb{R}^n$ and so we want to assume that $T_p^*M \approx \mathbb{R}^{n*}$.

Definition 3.13 *Let us define the* ***cotangent bundle*** *of a manifold* M *to be the set*

$$T^*M := \bigsqcup_{p \in M} T_p^*M$$

and define the map $\pi := \pi_M : \bigsqcup_{p \in M} T_p^*M \to M$ *to be the obvious projection taking elements in each space* T_p^*M *to the corresponding point* p. *Let* $\{U, \mathtt{x}\}_{\alpha \in A}$ *be an atlas of admissible charts on* M. *Now endow* T^*M *with the smooth structure given by the charts*

$$: T^*U = \pi_M^{-1}(U) \to T^*\mathtt{x}(T^*U) \cong \mathtt{x}(U) \times (\mathbb{R}^n)^*$$

where the map $(T\mathtt{x}^{-1})^t$ *the* ***contragradient*** *of* $T\mathtt{x}$.

If M is a smooth n dimensional manifold and $x^1, ..., x^n$ are coordinate functions coming from some chart on M then the "differentials" $dx^1\big|_p, ..., dx^n\big|_p$ are a basis of T_p^*M basis dual to $\frac{\partial}{\partial x^1}\big|_p, ..., \frac{\partial}{\partial x^n}\big|_p$. Let $\alpha \in T^*U$. Then we can write

$$\alpha = \sum a_i(\alpha)\, dx^i\big|_{\pi_M(\alpha)}$$

for some numbers $a_i(\alpha)$ depending on α. In fact, $a_i(\alpha) = \alpha(\frac{\partial}{\partial x^i}\big|_{\pi(\alpha)})$. So if $U, \mathtt{x} = (x^1, ..., x^n)$ is a chart on an n-manifold M, then the natural chart $(TU, T^*\mathtt{x})$ defined above is given by

$$\alpha \mapsto (x^1 \circ \pi(\alpha), ..., x^n \circ \pi(\alpha), a_1(\alpha), ..., a_n(\alpha))$$

and abbreviated to $(x^1, ..., x^n, a_1, ..., a_n)$.

Suppose that $(x^1, ..., x^n, a_1, ..., a_n)$ and $(\overline{x}^1, ..., \overline{x}^n, \overline{a}_1, ..., \overline{a}_n)$ are two such charts constructed in this way from two charts on U and U respectively with $U \cap \grave{U} \neq \emptyset$. Then $T^*U \cap T^*U \neq \emptyset$ and on the overlap we have the coordinate transitions

$$(T\tilde{\mathtt{x}}^{-1})^* \circ (T\mathtt{x})^* : \mathtt{x}(U \cap \grave{U}) \times \mathbb{R}^{n*} \to \tilde{\mathtt{x}}(U \cap \grave{U}) \times (\mathbb{R}^n)^*$$

or

$$(T\mathtt{x} \circ T\tilde{\mathtt{x}}^{-1})^* : \mathtt{x}(U \cap \grave{U}) \times \mathbb{R}^{n*} \to \tilde{\mathtt{x}}(U \cap \grave{U}) \times (\mathbb{R}^n)^*.$$

Notation 3.3 *The contragradient of* $D(\tilde{\mathtt{x}} \circ \mathtt{x}^{-1})$ *at* $\mathsf{x} \in \mathtt{x}(U \cap \grave{U})$ *is the map*

$$\frac{\partial^* \mathtt{x}}{\partial \tilde{\mathtt{x}}}(\mathsf{x}) : (\mathbb{R}^n)^* \to (\mathbb{R}^n)^*$$

defined by

$$\frac{\partial^* \mathtt{x}}{\partial \tilde{\mathtt{x}}}(\mathsf{x}) \cdot \mathsf{a} = \left(D(\mathtt{x} \circ \tilde{\mathtt{x}}^{-1})(\mathsf{x})\right)^* \cdot \mathsf{a}$$

When convenient we also write $\frac{\partial^* \mathtt{x}}{\partial \tilde{\mathtt{x}}}\Big|_{\mathtt{x}(p)} \mathsf{a}$.

With this notation we can write coordinate change maps as

$$(\mathsf{x},\mathsf{a}) \mapsto \left((\tilde{\mathtt{x}} \circ \mathtt{x}^{-1})(\mathsf{x}), \frac{\partial^* \mathtt{x}}{\partial \tilde{\mathtt{x}}}(\mathsf{x}) \cdot \mathsf{a} \right).$$

Write $(\tilde{\mathtt{x}} \circ \mathtt{x}^{-1})^i := pr_i \circ (\tilde{\mathtt{x}} \circ \mathtt{x}^{-1})$ and then

$$\begin{aligned}
\grave{x}^i &= (\tilde{\mathtt{x}} \circ \mathtt{x}^{-1})^i (x^1 \circ \pi, ..., x^n \circ \pi) \\
\grave{a}_i &= \sum_{k=1}^{n} (D(\mathtt{x} \circ \tilde{\mathtt{x}}^{-1}))_i^k a_k.
\end{aligned}$$

and classically abbreviated even further to

$$\begin{aligned}
\grave{x}^i &= \grave{x}^i(x^1, ..., x^n) \\
\grave{p}_i &= p_k \frac{\partial x^k}{\partial \grave{x}^i}.
\end{aligned}$$

This is the socalled "index notation" and does not generalize well to infinite dimensions. The following version is index free and makes sense even in the infinite dimensional case:

$$\begin{aligned}
\grave{\mathsf{x}} &= \tilde{\mathtt{x}} \circ \mathtt{x}^{-1}(\mathsf{x}) \\
\grave{\mathsf{a}} &= \left. \frac{\partial^* \mathtt{x}}{\partial \tilde{\mathtt{x}}} \right|_{\mathtt{x}(p)} \mathsf{a}
\end{aligned}$$

This last expression is very nice if inaccurate and again is in line with our notational principle **??**.

Exercise 3.4 *Show that the notion of a "cotangent lift" only works if the map is a diffeomorphism.*

3.6 Important Special Situations.

If the manifold in question is an open subset U of a vector space V then the tangent space at any $x \in \mathsf{V}$ is canonically isomorphic with V itself. This was clear when we defined the tangent space at x as $\{\mathsf{x}\} \times \mathsf{V}$. Then the identifying map is just $\mathsf{v} \mapsto (\mathsf{x}, \mathsf{v})$. Now one may convince oneself that the new more abstract definition of T_xU is essentially the same thing but we will describe the canonical map in another way: Let $\mathsf{v} \in \mathsf{V}$ and define a curve $c_{\mathsf{v}} : \mathbb{R} \to U \subset \mathsf{V}$ by $c_{\mathsf{v}}(t) = \mathsf{x} + t\mathsf{v}$. Then $T_0 c_{\mathsf{v}} \cdot 1 = \dot{c}_{\mathsf{v}}(0) \in T_xU$. The map $\mathsf{v} \mapsto \dot{c}_{\mathsf{v}}(0)$ is then our identifying map. The fact that there are these various identifications and that some things have several "equivalent" definitions is somewhat of a nuisance to the novice (occasionally to the expert also). The important thing is to think things through carefully, draw a few pictures, and most of all, try to think geometrically. One thing to notice is that for a vector spaces the derivative rather than the tangent map is all one needs in most cases. For example, if one wants to study a map $f : U \subset \mathsf{V} \to \mathsf{W}$ then if $v_{\mathsf{p}} = (\mathsf{p}, \mathsf{v}) \in T_{\mathsf{p}}U$ then $T_{\mathsf{p}}f \cdot v_{\mathsf{p}} = T_{\mathsf{p}}f \cdot (\mathsf{p}, \mathsf{v}) = (\mathsf{p}, Df|_{\mathsf{p}} \cdot \mathsf{v})$. In other words the tangent map is $(\mathsf{p}, \mathsf{v}) \mapsto (\mathsf{p}, Df|_{\mathsf{p}} \cdot \mathsf{v})$ and so one might as well just think about the ordinary derivative $D_{\mathsf{p}}f$. In fact, in the case of a vector space some authors actually identify $T_{\mathsf{p}}f$ with $Df|_{\mathsf{p}}$ as they also identify $T_{\mathsf{p}}U$ with V. There is no harm in this and actually streamlines the calculations a bit.

Another related situation is the case of a manifold of matrices such as $\mathrm{GL}(n, \mathbb{R})$. Here $\mathrm{GL}(n, \mathbb{R})$ is actually an open subset of the set of all $n \times n$-matrices $\mathbb{M}_{n \times n}(\mathbb{R})$. The latter is a vector space so all our comments above apply so that we can think of $\mathbb{M}_{n \times n}(\mathbb{R})$ as any of the tangent spaces $T_x\mathrm{GL}(n, \mathbb{R})$. Another interesting fact is that many important maps such as $c_g : x \mapsto g^t x g$ are actually linear so with the identifications $T_x\mathrm{GL}(n, \mathbb{R}) = \mathbb{M}_{n \times n}(\mathbb{R})$ and we have

$$T_x c_g \text{ `` = '' } Dc_x|_g = c_x : \mathbb{M}_{n \times n}(\mathbb{R}) \to \mathbb{M}_{n \times n}(\mathbb{R}).$$

Definition 3.14 (Partial Tangential) *Suppose that $f : M_1 \times M_2 \to N$ is a smooth map. We can define the partial maps as before and thus define **partial tangent maps**:*

$$\begin{aligned}
(\partial_1 f)(x, y) &: T_x M_1 \to T_{f(x,y)}N \\
(\partial_2 f)(x, y) &: T_y M_2 \to T_{f(x,y)}N
\end{aligned}$$

Next we introduce a natural identification . It is obvious that a curve $c : I \to M_1 \times M_2$ is equivalent to a pair of curves

$$\begin{aligned}
c_1 &: I \to M_1 \\
c_2 &: I \to M_2
\end{aligned}$$

The infinitesimal version of this fact gives rise to a natural identification

$$T_{(x,y)}(M_1 \times M_2) \cong T_x M_1 \times T_y M_2$$

This is perhaps easiest to see if we view tangent vectors as equivalence classes of curves (tangency classes). Then if we choose $c = (c_1, c_2)$ so that $c(0) = (x, y)$ then we identify $\xi = [c] \in T_{(x,y)}(M_1 \times M_2)$ with $([c_1], [c_2]) \in T_x M_1 \times T_y M_2$. For another view, consider the insertion maps $\iota_x : y \mapsto (x, y)$ and $\iota^y : x \mapsto (x, y)$. We have linear monomorphisms $T\iota^y(x) : T_x M_1 \to T_{(x,y)}(M_1 \times M_2)$ and $T\iota_x(y) : T_y M_2 \to T_{(x,y)}(M_1 \times M_2)$. Let us denote the images of $T_x M_1$ and $T_y M_2$ in $T_{(x,y)}(M_1 \times M_2)$ under these two maps by the same symbols $(T_x M)_1$ and $(T_y M)_2$. We then have the internal direct sum $(T_x M)_1 \oplus (T_y M)_2 = T_{(x,y)}(M_1 \times M_2)$ and the map $T\iota^y \times T\iota_x : T_x M_1 \times T_y M_2 \to (T_x M)_1 \oplus (T_y M)_2 \subset T_{(x,y)}(M_1 \times M_2)$. The inverse of this map is $T_{(x,y)}pr_1 \times T_{(x,y)}pr_2$ which is then also taken as an identification. One way to see the naturalness of this identification is to see the tangent functor as taking the commutative diagram in the category of pairs

$$
\begin{tikzcd}
 & M_1 \times M_2, (x,y) \arrow[dl, "pr_1"'] \arrow[dr, "pr_2"] \arrow[dd, leftrightarrow, "\mathrm{id}"'] & \\
M_1, x \arrow[dr, "\iota_y"'] & & M_2, y \arrow[dl, "\iota_x"] \\
 & M_1 \times M_2, (x,y) &
\end{tikzcd}
$$

to the new commutative diagram

$$
\begin{tikzcd}
 & T_{(x,y)}(M_1 \times M_2) \arrow[dl, "Tpr_1"'] \arrow[dr, "Tpr_2"] \arrow[d, leftrightarrow] & \\
T_x M_1 \arrow[dr, "T_x \iota_y"'] & T_x M_1 \times T_y M_2 \arrow[d, leftrightarrow] & T_y M_2 \arrow[dl, "T_x \iota_x"] \\
 & (T_x M)_1 \oplus (T_y M)_2 &
\end{tikzcd}
$$

Notice that we have $f \circ \iota_y = f_{,y}$ and $f \circ \iota_x = f_x$. Looking again at the definition of partial tangential one arrives at

Lemma 3.4 (partials lemma) *For a map $f : M_1 \times M_2 \to N$ we have*

$$T_{(x,y)} f \cdot (v, w) = (\partial_1 f)(x, y) \cdot v + (\partial_2 f)(x, y) \cdot w.$$

where we have used the aforementioned identification $T_{(x,y)}(M_1 \times M_2) = T_{(x,y)}(M_1 \times M_2)$.

Proving this last lemma is much easier and more instructive than reading the proof. Besides, it easy.

The following diagram commutes:

$$
\begin{tikzcd}
T_{(x,y)}(M_1 \times M_2) \arrow[dr] \arrow[dd, leftrightarrow, "T_{(x,y)} pr_1 \times T_{(x,y)} pr_2"'] & \\
 & T_{f(x,y)} N \\
T_x M_1 \times T_y M_2 \arrow[ur] &
\end{tikzcd}
$$

Essentially, both diagonal maps refer to $T_{(x,y)} f$ because of our identification.

Chapter 4

Submanifold, Immersion and Submersion.

4.1 Submanifolds

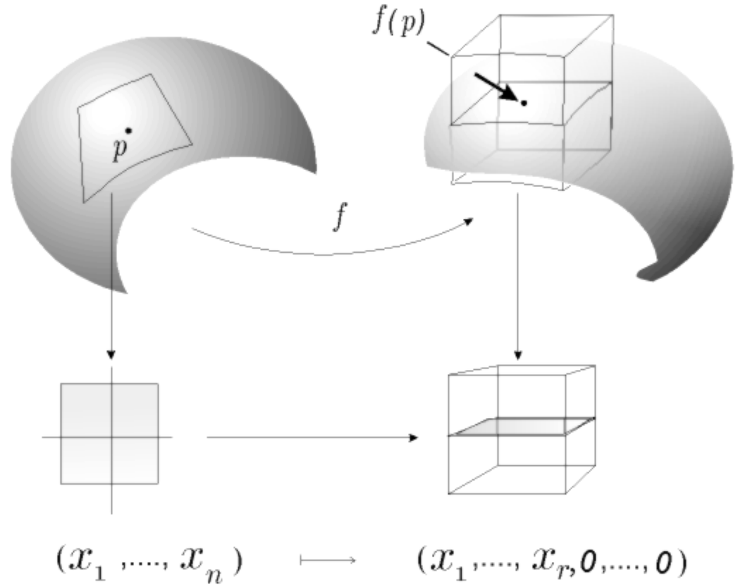
Recall the simple situation from calculus where we have a continuously differentiable function $F(x, y)$ on the x, y plane. We know from the implicit function theorem that if $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$ then *near* (x_0, y_0) the level set $F(x, y) = F(x_0, y_0)$ is the graph of some function $y = g(x)$. The map $(x, y) \mapsto x$ is a homeomorphism onto an open subset of $\mathbb{R}$ and provides a chart. Hence, near this point, the level set is a 1-dimensional differentiable manifold. Now if either $\frac{\partial F}{\partial y}(x, y) \neq 0$ or $\frac{\partial F}{\partial x}(x, y) \neq 0$ at every (x, y) on the level set, then we could cover the level set by these kind of charts (induced by coordinate projection) and so we would have a smooth 1-manifold. This idea generalizes nicely not only to higher dimensions but to manifolds in general and all we need is local theory of maps as described by the inverse and implicit mapping theorems.

There is another description of a level set. Locally these are graphs of functions. But then we can also parameterize portions of the level sets by using this local graph structure. For example, in the simple situation just described we have the map $t \mapsto (t + x_0, g(t + y_0))$ which parameterizes a portion of the level set near x_0, y_0. The inverse of this parameterization is just the chart.

First we define the notion of a submanifold and study some related generalities concerning maps. We then see how this dual idea of level sets and parameterizations generalizes to manifolds. The reader should keep in mind this dual notion of level sets and parameterizations.

A subset S of a C^r-differentiable manifold M (modelled on $\mathbb{R}^n$) is called a (**regular**) **submanifold** (of M) if there exists a decomposition of the model space $\mathbb{R}^n = \mathbb{R}^{n-k} \times \mathbb{R}^k$ such that every point $p \in S$ is in the domain of an admissible chart $(\mathtt{x}, U)$ which has the following **submanifold property**:

$$\mathtt{x}(U \cap S) = \mathtt{x}(U) \cap (\mathbb{R}^{n-k} \times \{0\})$$



Equivalently, we require that $\mathtt{x} : U \to V_1 \times V_2 \subset \mathbb{R}^{n-k} \times \mathbb{R}^k$ is a diffeomorphism such that

$$\mathtt{x}(U \cap S) = V_1 \times \{0\}$$

for open V_1, V_2. We will call such charts **adapted** to S. The restrictions $\mathtt{x}|_{U \cap S}$ of adapted charts provide an atlas for S (called an induced submanifold atlas) making it a differentiable manifold in its own right. The k above is called the **codimension** of S (in M).

Exercise 4.1 *Show that S really is differentiable manifold and that a continuous map $f : N \to M$ which has its image contained in S is differentiable with respect to the submanifold atlas iff it is differentiable as a map in to M.*

When S is a submanifold of M then the tangent space T_pS at $p \in S \subset M$ is intuitively a subspace of T_pM. In fact, this is true as long as one is not bent on distinguishing a curve in S through p from the "same" curve thought of as a map into M. If one wants to be pedantic then we have the inclusion map $\iota : S \hookrightarrow M$ and if $c : I \to S$ curve into S then $\iota \circ c : I \to M$ is a map into M as such. At the tangent level this means that $c'(0) \in T_pS$ while $(\iota \circ c)'(0) \in T_pM$. Thus from this more pedantic point of view we have to explicitly declare $T_p\iota : T_pS \to T_p\iota(T_pS) \subset T_pM$ to be an identifying map. We will avoid the use of inclusion maps when possible and simply write $T_pS \subset T_pM$ and trust the intuitive notion that T_pS is indeed a subspace of T_pM.

4.2 Submanifolds of $\mathbb{R}^n$

If $M \subset \mathbb{R}^n$ is a regular k-dimensional submanifold then, by definition, for every $p \in M$ there is an open subset U of $\mathbb{R}^n$ containing p on which we have new coordinates $\phi : U \to V \subset \mathbb{R}^n$ abbreviated by $(y_1, ..., y_n)$ such that $.M \cap U$ is exactly given by $y_{k+1} = ... = y_n = 0$. On the other hand we have the identity coordinates restricted to U which we denote by $x_1, ..., x_n$. We see that $id \circ \phi^{-1}$ is a diffeomorphism given by

$$\begin{gathered} x_1 = x_1(y_1, ..., y_n) \\ x_2 = x_2(y_1, ..., y_n) \\ \vdots \\ x_n = x_n(y_1, ..., y_n) \end{gathered}$$

which in turn implies that the determinant det $\frac{\partial(x_1,....,x_n)}{\partial(y_1,....,x_n)}$ must be nonzero throughout V. From a little linear algebra we conclude that for some renumbering of the coordinates the $x_1, ..., x_n$ the determinant det $\frac{\partial(x_1,....,x_r)}{\partial(y_1,....,x_r)}$ must be nonzero at and therefore near $\phi(p) \in V \subset \mathbb{R}^n$. On this possibly smaller neighborhood V' we define a map F by

$$\begin{gathered} x_1 = x_1(y_1, ..., y_n) \\ x_2 = x_2(y_1, ..., y_n) \\ \vdots \\ x_r = x_r(y_1, ..., y_n) \\ x_{r+1} = y_{r+1} \\ \vdots \\ x_n = y_n \end{gathered}$$

then we have that F is a local diffeomorphism $V' \subset \mathbb{R}^n \to \mathbb{R}^n$. Now let $\phi^{-1}U' = V'$ and form the composition $\psi := F \circ \phi$ which is defined on U' and must have the form

$$\begin{gathered} z_1 = x_1 \\ z_2 = x_2 \\ \vdots \\ z_r = x_r \\ z_{r+1} = \psi_{r+1}(x_1, ..., x_n) \\ \vdots \\ z_n = \psi_n(x_1, ..., x_n) \end{gathered}$$

From here is it is not hard to show that $z_{r+1} = \cdots = z_n = 0$ is exactly the set $\psi(M \cap U')$ and since ϕ restricted to a $M \cap U'$ is a coordinate system so ψ restricted to $M \cap U'$ is a coordinate system for M. Now notice that in fact ψ maps a point with standard (identity) coordinates $(a_1, ..., a_n)$ onto $(a_1, ..., a_r, 0, ..., 0)$. Now remembering that we renumbered coordinates in the middle of the discussion we have proved the following theorem.

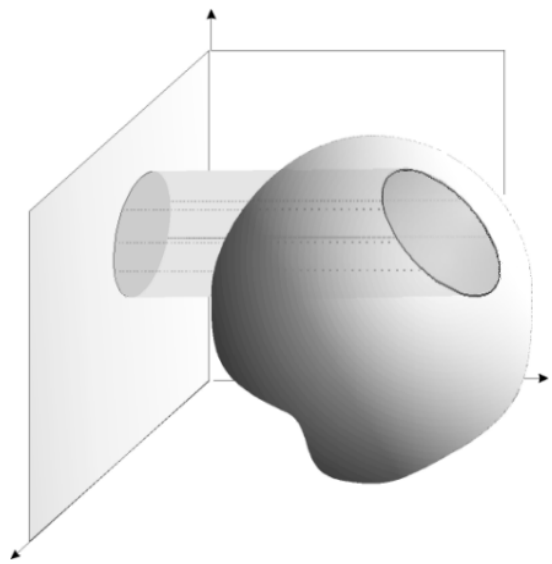


Figure 4.1: Projection onto some plane gives a chart.

Theorem 4.1 *If M is an r-dimensional regular submanifold of $\mathbb{R}^n$ then for every $p \in M$ there exists at least one r-dimensional coordinate plane P such that linear projection $P \to \mathbb{R}^n$ restricts to a coordinate system for M defined in a neighborhood of p.*

4.3 Regular and Critical Points and Values

Proposition 4.1 *If $f : M \to N$ is a smooth map such that $T_p f : T_p M \to T_q N$ is an isomorphism for all $p \in M$ then $f : M \to N$ is a local diffeomorphism.*

Definition 4.1 *Let $f : M \to N$ be C^r-map and $p \in M$ we say that p is a **regular point** for the map f if $T_p f$ is a splitting surjection (see 26.44) and is called a **singular point** otherwise. For finite dimensional manifolds this amounts to the requirement that $T_p f$ have full rank. A point q in N is called a **regular value** of f if every point in the inverse image $f^{-1}\{q\}$ is a regular point for f. A point of N which is not regular is called a **critical value**. The set of regular values is denoted $\mathcal{R}_f$.*

It is a very useful fact that regular values are easy to come by in that most values are regular. In order to make this precise we will introduce the notion of measure zero on a manifold. It is actually no problem to define a Lebesgue measure on a manifold but for now the notion of measure zero is all we need.

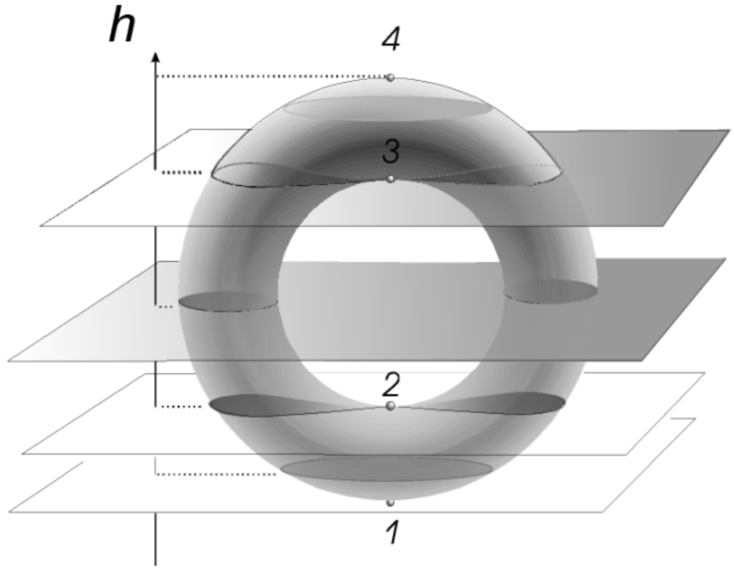


Figure 4.2: Four critical points of the height function.

Definition 4.2 *A set A in a smooth finite dimensional manifold M is said to be of **measure zero** if for every admissible chart U, ϕ the set $\phi(A \cap U)$ has Lebesgue measure zero in $\mathbb{R}^n$ where $\dim M = n$.*

In order for this to be a reasonable definition the manifold must be second countable so that every atlas has a countable subatlas. This way we may be assured that every set which we have defined to be measure zero is the countable union of sets which are measure zero as view in a some chart. We also need to know that the local notion of measure zero is independent of the chart. This follows from

Lemma 4.1 *Let M be a n-manifold. The image of a measure zero set under a differentiable map is of measure zero.*

Proof. We are assuming, of course that M is Hausdorff and second countable. Thus any set is contained in the countable union of coordinate charts we may assume that $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ and A is some measure zero subset of U. In fact, since A is certainly contained in the countable union of compact balls (all of which are translates of a ball at the origin) we may as well assume that $U = B(0, r)$ and that A is contained in a slightly smaller ball $B(0, r - \delta) \subset B(0, r)$. By the mean value theorem, there is a constant c depending only on f and its domain such that for $x, y \in B(0, r)$ we have $|f(y) - f(x)| \leq c\,|x - y|$. Let $\epsilon > 0$ be given. Since A has measure zero there is a sequence of balls $B(x_i, \epsilon_i)$ such that $A \subset \bigcup B(x_i, \epsilon_i)$ and

$$\sum vol(B(x_i, \epsilon_i)) < \frac{\epsilon}{2^n c^n}$$

Thus $f(B(x_i, \epsilon_i)) \subset B(f(x_i), 2c\epsilon_i)$ and while $f(A) \subset \bigcup B(f(x_i), 2c\epsilon_i)$ we also have

$$\begin{aligned} \operatorname{vol}\left(\bigcup B(f(x_i), 2c\epsilon_i)\right) &\leq \\ \sum \operatorname{vol}(B(f(x_i), 2c\epsilon_i)) &\leq \sum \operatorname{vol}(B_1)(2c\epsilon_i)^n \\ &\leq 2^n c^n \sum \operatorname{vol}(B(x_i, \epsilon_i)) \\ &\leq \epsilon. \end{aligned}$$

Thus the measure of A is less that or equal to ϵ. Since ϵ was arbitrary it follows that A has measure zero. ∎

Corollary 4.1 *Given a fixed $\mathcal{A} = \{U_\alpha, \mathtt{x}_\alpha\}$ atlas for M, if $\mathtt{x}_\alpha(A \cap U_\alpha)$ has measure zero for all α then A has measure zero.*

Theorem 4.2 (Sard) *Let M be an n-manifold and N an m-manifold (Hausdorff and second countable). For a smooth map $f : M \to N$ the set of regular values $\mathcal{R}_f$ has Lebesgue measure zero.*

Proof. Through the use of a countable cover of the manifolds in question by charts we may immediately reduce to the problem of showing that for a smooth map $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ the set of critical values $C \subset U$ has image $f(C)$ of measure zero. We will use induction on the dimension n. For $n = 0$, the set $f(C)$ is just a point (or empty) and so has measure zero. Now assume the theorem is true for all dimensions $j \leq n - 1$. We seek to show that the truth of the theorem follows for $j = n$ also.

Using multiindex notation (26.3) let

$$C_i := \{x \in U : \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x) = 0 \text{ for all } |\alpha| \leq i\}.$$

Then

$$C = (C \backslash C_1) \cup (C_1 \backslash C_2) \cup \cdots \cup (C_{k-1} \backslash C_k) \cup C_k$$

so we will be done if we can show that

a) $f(C \backslash C_1)$ has measure zero,

b) $f(C_{j-1} \backslash C_j)$ has measure zero and

c) $f(C_k)$ has measure zero for some sufficiently large k.

Proof of a): We may assume that $m \geq 2$ since if $m = 1$ we have $C = C_1$. Now let $x \in C \backslash C_1$ so that some first partial derivative is not zero at $x = a$. By reordering we may assume that this partial is $\frac{\partial f}{\partial x^1}$ and so the

$$(x^1, ..., x^n) \mapsto (f(x), x^2, ..., x^n)$$

map restricts to a diffeomorphism ϕ on some open neighborhood containing x. Since we may always replace f by the equivalent map $f \circ \phi^{-1}$ we may go ahead and assume without loss of generality that f has the form

$$f : x \mapsto (x^1, f^2(x), ..., f^m(x)) := (x^1, h(x))$$

on some perhaps smaller neighborhood V containing a. The Jacobian matrix for f in V is of the form

$$\begin{bmatrix} 1 & 0 \\ * & Dh \end{bmatrix}$$

and so $x \in V$ is critical for f if and only if it is critical for h. Now $h(C \cap V) \subset \mathbb{R}^{m-1}$ and so by the induction hypothesis $h(C \cap V)$ has measure zero in $\mathbb{R}^{m-1}$. Now $f(C \cap V) \cap (\{x\} \times \mathbb{R}^{m-1}) \subset \{x\} \times h(C \cap V)$ which has measure zero in $\{x\} \times \mathbb{R}^{m-1} \cong \mathbb{R}^{m-1}$ and so by Fubini's theorem $f(C \cap V)$ has measure zero. Since we may cover C by a countable number of sets of the form $C \cap V$ we conclude that $f(C)$ itself has measure zero.

Proof of (b): The proof of this part is quite similar to the proof of (a). Let $a \in C_{j-1} \backslash C_j$. It follows that some k-th partial derivative is 0 and after some permutation of the coordinate functions we may assume that

$$\frac{\partial}{\partial x^1} \frac{\partial^{|\beta|} f^1}{\partial x^\beta}(a) \neq 0$$

for some $j-1$- tuple $\beta = (i_1, ..., i_{j-1})$ where the function $g := \frac{\partial^{|\beta|} f^1}{\partial x^\beta}$ is zero at a since a is in C_{k-1}. Thus as before we have a map

$$x \mapsto (g(x), x^2, ..., x^n)$$

which restricts to a diffeomorphism ϕ on some open set V. We use ϕ, V as a chart about a. Notice that $\phi(C_{j-1} \cap V) \subset 0 \times \mathbb{R}^{n-1}$. We may use this chart ϕ to replace f by $g = f \circ \phi^{-1}$ which has the form

$$x \mapsto (x^1, h(x))$$

for some map $h : V \to \mathbb{R}^{m-1}$. Now by the induction hypothesis the restriction of g to

$$g_0 : \{0\} \times \mathbb{R}^{n-1} \cap V \to \mathbb{R}^m$$

has a set of critical values of measure zero. But each point from $\phi(C_{j-1} \cap V) \subset 0 \times \mathbb{R}^{n-1}$ is critical for g_0 since diffeomorphisms preserve criticality. Thus $g \circ \phi(C_{j-1} \cap V) = f(C_{j-1} \cap V)$ has measure zero.

Proof of (c): Let $I^n(r) \subset U$ be a cube of side r. We will show that if $k > (n/m) - 1$ then $f(I^n(r) \cap C_k)$ has measure zero. Since we may cover by a countable collection of such V the result follows. Now Taylor's theorem give that if $a \in I^n(r) \cap C_k$ and $a + h \in I^n(r)$ then

$$|f(a+h) - f(a)| \leq c\,|h|^{k+1} \tag{4.1}$$

for some constant c which depends only on f and $I^n(r)$. We now decompose the cube $I^n(r)$ into R^n cubes of side length r/R. Suppose that we label these cubes which contain critical points of f as $D_1, ...D_N$. Let D_i contains a critical point a of f. Now if $y \in D$ then $|y - a| \leq \sqrt{n}r/R$ so using the Taylor's theorem remainder estimate above (4.1) with $y = a + h$ we see that $f(D_i)$ is contained in a cube $\widetilde{D}_i \subset \mathbb{R}^m$ of side

$$2c\left(\frac{\sqrt{n}r}{R}\right)^{k+1} = \frac{b}{R^{k+1}}$$

where the constant $b := (\sqrt{n}r)^{k+1}$ is independent of the particular cube D from the decomposition and depends only on f and $I^n(r)$. The sum of the volumes of all such cubes $\widetilde{D}_i$ is

$$S \leq R^n\left(\frac{b}{R^{k+1}}\right)^m$$

which, under the condition that $m(k+1) > n$, may be made arbitrarily small be choosing R large (refining the decomposition of $I^n(r)$). The result now follows. ■

Corollary 4.2 *If M and N are finite dimensional manifolds then the critical values of a smooth map $f : M \to N$ are dense in N.*

4.4 Immersions

Definition 4.3 *A map $f : M \to N$ is called an **immersion at** $p \in M$ iff $T_pf : T_pM \to T_{f(p)}N$ is a linear injection (see 26.43) at p. A map $f : M \to N$ is called an **immersion** if f is an immersion at every $p \in M$.*

Figure 4.3 shows a simple illustration of an immersion of $\mathbb{R}^2$ into $\mathbb{R}^3$. This example is also an injective immersion (as far as is shown) but an immersion can come back and cross itself. Being an immersion at p only requires that the restriction of the map to some small open neighborhood of p is injective. If an immersion is (globally) injective then we call it an immersed submanifold (see the definition 4.4 below).

Theorem 4.3 *Let $f : M^n \to N^d$ be a smooth function which is an immersion at p. Then $f : M^n \to N^d$ there exists charts $\mathtt{x} :: (M^n, p) \to (\mathbb{R}^n, 0)$ and $\mathtt{y} :: (N^d, f(p)) \to (\mathbb{R}^d, 0)$ such that*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^{d-n}$$

is given by $\mathsf{x} \mapsto (\mathsf{x}, 0)$ near $\mathsf{0}$. In other words, there is a open set $U \subset M$ such that $f(U)$ is a submanifold of N the expression for f is $(x^1, ..., x^n) \mapsto (x^1, ..., x^n, 0, ..., 0) \in \mathbb{R}^d$.

Proof. Follows easily from theorem 26.12. ■

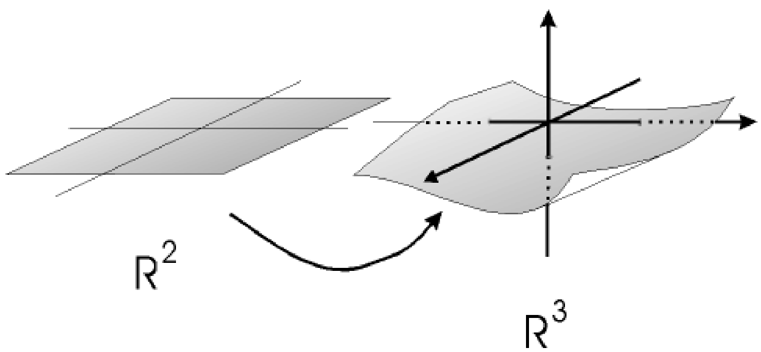


Figure 4.3: Embedding of the plane into 3d space.

Theorem 4.4 *If $f : M \to N$ is an immersion (so an immersion at every point) and if f is a homeomorphism onto its image $f(M)$ using the relative topology, then $f(M)$ is a regular submanifold of N. In this case we call $f : M \to N$ an* ***embedding****.*

Proof. Follows from the last theorem plus a little point set topology. ■

4.5 Immersed Submanifolds and Initial Submanifolds

Definition 4.4 *If $I : S \to M$ is an injective immersion then (S, I) is called an* ***immersed submanifold****.*

Exercise 4.2 *Show that every injective immersion of a compact manifold is an embedding.*

Theorem 4.5 *Suppose that M is an $n-$dimensional smooth manifold which has a finite atlas. Then there exist an injective immersion of M into $\mathbb{R}^{2n+1}$. Consequently, every* compact *$n-$dimensional smooth manifold can be embedded into $\mathbb{R}^{2n+1}$.*

Proof. Let M be a smooth manifold. Initially, we will settle for an immersion into $\mathbb{R}^D$ for some possibly very large dimension D. Let $\{O_i, \varphi_i\}_{i \in N}$ be an atlas with cardinality $N < \infty$. The cover $\{O_i\}$ cover may be refined to two other covers $\{U_i\}_{i \in N}$ and $\{V\}_{i \in N}$ such that $\overline{U_i} \subset V_i \subset \overline{V_i} \subset O_i$. Also, we may find smooth functions $f_i : M \to [0, 1]$ such that

$$\begin{aligned} f_i(x) &= 1 \text{ for all } x \in U_i \\ \operatorname{supp}(f_i) &\subset O_i. \end{aligned}$$

Next we write $\varphi_i = (x_i^1, ...x_i^n)$ so that $x_i^j : O_i \to \mathbb{R}$ is the $j-$th coordinate function of the i-th chart and then let

$$f_{ij} := f_i x_i^j \text{ (no sum)}$$

which is defined and smooth on all of M after extension by zero.

Now we put the functions f_i together with the functions f_{ij} to get a map $i : M \to \mathbb{R}^{n+Nn}$:

$$i = (f_1, ..., f_n, f_{11}, f_{12}, ..., f_{21}, ..., f_{nN}).$$

Now we show that i is injective. Suppose that $i(x) = i(y)$. Now $f_k(x)$ must be 1 for some k since $x \in U_k$ for some k. But then also $f_k(y) = 1$ also and this means that $y \in V_k$ (why?). Now then, since $f_k(x) = f_k(y) = 1$ it follows that $f_{kj}(x) = f_{kj}(y)$ for all j. Remembering how things were defined we see that x and y have the same image under $\varphi_k : O_k \to \mathbb{R}^n$ and thus $x = y$.

To show that $T_x i$ is injective for all $x \in M$ we fix an arbitrary such x and then $x \in U_k$ for some k. But then near this x the functions $f_{k1}, f_{k2}, ..., f_{kn,}$ are equal to $x_k^1, ...x_k^n$ and so the rank of i must be *at least* n and in fact equal to n since $\dim T_x M = n$.

So far we have an injective immersion of M into $\mathbb{R}^{n+Nn}$.

We show that there is a projection $\pi : \mathbb{R}^D \to L \subset \mathbb{R}^D$ where $L \cong \mathbb{R}^{2n+1}$ is a $2n+1$ dimensional subspace of $\mathbb{R}^D$, such that $\pi \circ f$ is an injective immersion. The proof of this will be inductive. So suppose that there is an injective immersion f of M into $\mathbb{R}^d$ for some d with $D \geq d > 2n+1$. We show that there is a projection $\pi_d : \mathbb{R}^d \to L^{d-1} \cong \mathbb{R}^{d-1}$ such that $\pi_d \circ f$ is still an injective immersion. To this end, define a map $h : M \times M \times \mathbb{R} \to \mathbb{R}^d$ by $h(x, y, t) := t(f(x) - f(y))$. Now since $d > 2n+1$, Sard's theorem implies that there is a vector $y \in \mathbb{R}^d$ which is neither in the image of the map h nor in the image of the map $df : TM \to \mathbb{R}^d$. This y cannot be 0 since 0 is certainly in the image of both of these maps. Now if $pr_{\perp y}$ is projection onto the orthogonal compliment of y then $pr_{\perp y} \circ f$ is injective; for if $pr_{\perp y} \circ f(x) = pr_{\perp y} \circ f(y)$ then $f(x) - f(y) = ay$ for some $a \in \mathbb{R}$. But suppose $x \neq y$. then since f is injective we must have $a \neq 0$. This state of affairs is impossible since it results in the equation $h(x, y, 1/a) = y$ which contradicts our choice of y. Thus $pr_{\perp y} \circ f$ is injective.

Next we examine $T_x(pr_{\perp y} \circ f)$ for an arbitrary $x \in M$. Suppose that $T_x(pr_{\perp y} \circ f)v = 0$. Then $d(pr_{\perp y} \circ f)|_x v = 0$ and since $pr_{\perp y}$ is linear this amounts to $pr_{\perp y} \circ df|_x v = 0$ which gives $df|_x v = ay$ for some number $a \in \mathbb{R}$ which cannot be 0 since f is assumed an immersion. But then $df|_x \frac{1}{a} v = y$ which also contradict our choice of y.

We conclude that $pr_{\perp y} \circ f$ is an injective immersion. Repeating this process inductively we finally get a composition of projections $pr : \mathbb{R}^D \to \mathbb{R}^{2n+1}$ such that $pr \circ f : M \to \mathbb{R}^{2n+1}$ is an injective immersion. ■

It might surprise the reader that an immersed submanifold *does not* necessarily have the following property:

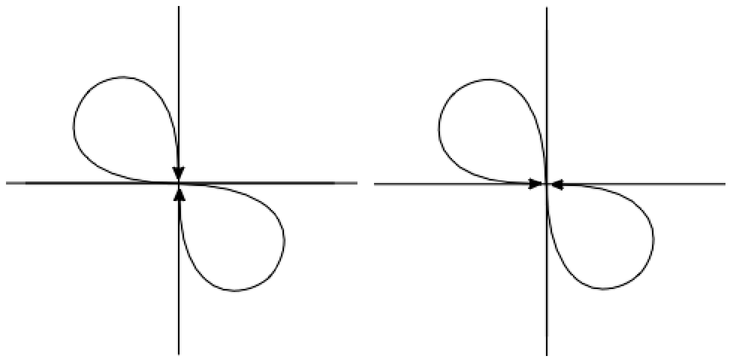


Figure 4.4: Counter example: consider the superposition.

Criterion 4.1 *Let S and M be a smooth manifolds. An injective immersion $I : S \to M$ is called **smoothly universal** if for any smooth manifold N, a mapping $f : N \to S$ is smooth if and only if $I \circ f$ is smooth.*

To see what goes wrong, imagine the map corresponding to superimposing one of the figure eights shown in figure 4.4 onto the other. If $I : S \to M$ is an embedding then it is also smoothly universal but this is too strong of a condition for our needs. For that we make the following definitions which will be especially handy when we study foliations.

Definition 4.5 *Let S be any subset of a smooth manifold M. For any $x \in S$ denote by $C_x(S)$ the set of all points of S that can be connected to x by a smooth curve with image entirely inside S.*

Definition 4.6 *A subset $S \subset M$ is called an **initial submanifold** if for each $s_0 \in S$ there exists a chart $U, \mathtt{x}$ centered at s_0 such that $\mathtt{x}(C_{s_0}(U \cap S)) = \mathtt{x}(U) \cap (\mathbb{R}^d \times \{0\})$ for some splitting $\mathbb{R}^n = \mathbb{R}^d \times \mathbb{R}^{n-d}$ (which is independent of s_0).*

The definition implies that if S is an initial submanifold of M then it has a unique smooth structure as a manifold modelled on $\mathbb{R}^d$ and it is also not hard to see that any initial submanifold S has the property that the inclusion map $S \hookrightarrow M$ is smoothly universal. Conversely, we have the following

Theorem 4.6 *If an injective immersion $I : S \to M$ is smoothly universal then the image $f(S)$ is an initial submanifold.*

Proof. Choose $s_0 \in S$. Since I is an immersion we may pick a coordinate chart $w : W \to \mathbb{R}^d$ centered at s_0 and a chart $v : V \to \mathbb{R}^n = \mathbb{R}^d \times \mathbb{R}^{n-d}$ centered at $I(s_0)$ such that we have

$$v \circ I \circ w^{-1}(\mathsf{y}) = (\mathsf{y}, 0).$$

Choose an $r > 0$ small enough that $B(0,r) \subset w(U)$ and $B(0,2r) \subset w(V)$. Let $U_0 = v^{-1}(B(0,r))$ and $W_0 = w^{-1}(V)$. We show that the coordinate chart $V_0, u := \varphi|_{V_0}$ satisfies the property of lemma 4.6.

$$\begin{aligned} u^{-1}(u(U_0) \cap (\mathbb{R}^d \times \{0\})) &= u^{-1}\{(\mathsf{y},0) : \|\mathsf{y}\| < r\} \\ &= I \circ w^{-1} \circ (u \circ I \circ w^{-1})^{-1}(\{(\mathsf{y},0) : \|\mathsf{y}\| < r\}) \\ &= I \circ w^{-1}(\{\mathsf{y} : \|\mathsf{y}\| < r\}) = I(W_0) \end{aligned}$$

Now $I(W_0) \subset U_0 \cap I(S)$ and since $I(W_0)$ is contractible we have $I(W_0) \subset C_{f(s_0)}(U_0 \cap I(S))$. Thus $u^{-1}(u(U_0) \cap (\ \mathbb{R}^d \times \{0\})) \subset C_{f(s_0)}(U_0 \cap I(S))$ or

$$u(U_0) \cap (\mathbb{R}^d \times \{0\}) \subset u(C_{I(s_0)}(U_0 \cap I(S))).$$

Conversely, let $z \in C_{I(s_0)}(U_0 \cap I(S))$. By definition there must be a smooth curve $c : [0,1] \to S$ starting at $I(s_0)$, ending at z and $c([0,1]) \subset U_0 \cap I(S)$. Since $I : S \to M$ is smoothly universal there is a unique smooth curve $c_1 : [0,1] \to S$ with $I \circ c_1 = c$.

Claim 4.1 $c_1([0,1]) \subset W_0$.

Assume not. Then there is some number $t \in [0,1]$ *with* $c_1(t) \in w^{-1}(\{r \leq \|\mathsf{y}\| < 2r\})$. *The*

$$\begin{aligned} (v \circ I)(c_1(t)) &\in (v \circ I \circ w^{-1})(\{r \leq \|\mathsf{y}\| < 2r\}) \\ &= \{(\mathsf{y},0) : r \leq \|\mathsf{y}\| < 2r\} \subset \{z \in \mathbb{R}^n : r \leq \|\mathsf{y}\| < 2r\}. \end{aligned}$$

Now this implies that $(v \circ I \circ c_1)(t) = (v \circ c)(t) \in \{z \in \mathbb{R}^n : r \leq \|\mathsf{y}\| < 2r\}$ *which in turn implies the contradiction* $c(t) \notin U_0$. *The claim is proven.*

Now the fact that $c_1([0,1]) \subset W_0$ *implies* $c_1(1) = I^{-1}(z) \in W_0$ *and so* $z \in I(W_0)$. *As a result we have* $C_{I(s_0)}(U_0 \cap I(S)) = I(W_0)$ *which together with the first half of the proof gives the result:*

$$\begin{aligned} I(W_0) = u^{-1}(u(U_0) \cap (\mathbb{R}^d \times \{0\})) &\subset C_{f(s_0)}(U_0 \cap I(S)) = I(W_0) \\ \implies \qquad & u^{-1}(u(U_0) \cap (\mathbb{R}^d \times \{0\})) = C_{f(s_0)}(U_0 \cap I(S)) \\ \implies \qquad & u(U_0) \cap (\mathbb{R}^d \times \{0\}) = u(C_{f(s_0)}(U_0 \cap I(S))). \end{aligned}$$

■

We say that two immersed submanifolds (S_1, I_1) and (S_2, I_2) are **equivalent** if there exist a diffeomorphism $\Phi : S_1 \to S_2$ such that $I_2 \circ \Phi = I_1$; i.e. so that the following diagram commutes

$$\begin{tikzcd} S_1 \arrow[rr, "\Phi"] \arrow[dr] & & S_2 \arrow[dl] \\ & M & \end{tikzcd}.$$

Now if $I : S \to M$ is smoothly universal so that $f(S)$ is an initial submanifold then it is not hard to see that (S, I) is equivalent to $(f(S), \iota)$ where ι is the inclusion map and we give $f(S)$ the unique smooth structure guaranteed by the fact that it is an initial submanifold. Thus we may as well be studying initial submanifolds rather than injective immersions that are smoothly universal. For this reason we will seldom have occasion to even use the terminology "smoothly universal" which the author now confesses to be a nonstandard terminology anyway.

4.6 Submersions

Definition 4.7 *A map $f : M \to N$ is called a **submersion at** $p \in M$ iff $T_p f : T_p M \to T_{f(p)} N$ is a (bounded) splitting surjection (see 26.43). $f : M \to N$ is called a **submersion** if f is a submersion at every $p \in M$.*

Example 4.1 *The map of the punctured space $\mathbb{R}^3 - \{0\}$ onto the sphere S^2 given by $x \mapsto |x|$ is a submersion. To see this use spherical coordinates and the map becomes $(\rho, \phi, \theta) \mapsto (\phi, \theta)$. Here we ended up with a projection onto a second factor $\mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ but this is clearly good enough.*

Theorem 4.7 *Let $f : M \to N$ be a smooth function which is an submersion at p. Then there exists charts $\mathtt{x} :: (M, p) \to (\mathbb{R}^{n-k} \times \mathbb{R}^k, 0) = (\mathbb{R}^n, 0)$ and $\mathtt{y} :: (N, f(p)) \to (\mathbb{R}^k, 0)$ such that*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: (\mathbb{R}^{n-k} \times \mathbb{R}^k, 0) \to (\mathbb{R}^k, 0)$$

is given by $(\mathsf{x}, \mathsf{y}) \mapsto \mathsf{x}$ near $0 = (0, 0)$.

Proof. Follows directly from theorem 26.14. ∎

Corollary 4.3 (Submanifold Theorem I) *Consider any smooth map $f : M \to N$ then if $q \in N$ is a regular value the inverse image set $f^{-1}(q)$ is a regular submanifold.*

Proof. If $q \in N$ is a regular value then f is a submersion at every $p \in f^{-1}(q)$. Thus for any $p \in f^{-1}(q)$ there exist charts $\psi :: (M, p) \to (\mathbb{R}^{n-k} \times \mathbb{R}^k, 0) = (\mathbb{R}^n, 0)$ and $\phi :: (N, f(p)) \to (\mathbb{R}^{n-k}, 0)$ such that

$$\phi \circ f \circ \psi^{-1} :: (\mathbb{R}^{n-k} \times \mathbb{R}^k, 0) \to (\mathbb{R}^{n-k}, 0)$$

is given by $(\mathsf{x}, \mathsf{y}) \mapsto \mathsf{x}$ near 0. We may assume that domains are of the nice form

$$U' \times V' \overset{\psi^{-1}}{\to} U \overset{f}{\to} V \overset{\phi}{\to} V'.$$

But $\phi \circ f \circ \psi^{-1}$ is just projection and since q corresponds to 0 under the diffeomorphism we see that $\psi(U \cap f^{-1}(q)) = (\phi \circ f \circ \psi^{-1})^{-1}(0) = pr_1^{-1}(0) = U' \times \{0\}$ so that $f^{-1}(q)$ has the submanifold property at p. Now $p \in f^{-1}(q)$ was arbitrary so we have a cover of $f^{-1}(q)$ by submanifold charts. ∎

Example 4.2 (The unit sphere) *The set* $S^{n-1} = \{x \in \mathbb{R}^n : x \cdot x = 1\}$ *is a codimension* 1 *submanifold of* $\mathbb{R}^n$ *since we can use the map* $(x, y) \mapsto x^2 + y^2$ *as our map and let* $q = 1$.

Given k functions $F^j(x, y)$ on $\mathbb{R}^n \times \mathbb{R}^k$ we define the locus

$$M := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^k : F^j(x, y) = c^j\}$$

where each c^j is a fixed number in the range of F^j. If the Jacobian determinant at $(x_0, y_0) \in M$;

$$\det \frac{\partial F}{\partial y}(x_0, y_0)$$

is not zero then near (x_0, y_0) then we can apply the theorem. We can see things more directly: Since the Jacobian determinant is nonzero, we can solve the equations $F^j(x, y) = c^j$ for $y^1, ..., y^n$ in terms of $x^1, ..., x^k$:

$$\begin{aligned} y^1 &= f^1(x^1, ..., x^k) \\ y^2 &= f^2(x^1, ..., x^k) \\ & \\ y^n &= f^n(x^1, ..., x^k) \end{aligned}$$

and this parameterizes M near (x_0, y_0) in such a nice way that the inverse is a chart for M. This latter statement is really the content of the inverse mapping theorem in this case. If the Jacobian determinant never vanishes on M then we have a cover by charts and M is a submanifold of $\mathbb{R}^n \times \mathbb{R}^k$.

It may help the understanding to recall that if $F^j(x, y)$ and (x_0, y_0) are as above then we can differentiate the expressions $F^j(x, y(x)) = c^j$ to get

$$\frac{\partial F^i}{\partial x^j} + \sum_s \frac{\partial F^i}{\partial y^s}\frac{\partial y^s}{\partial x^j} = 0$$

and then solve

$$\frac{\partial y^j}{\partial x^i} = -\sum [J^{-1}]^j_s \frac{\partial F^s}{\partial x^i}$$

where $[J^{-1}]^j_s$ is the matrix inverse of the Jacobian $\frac{\partial F}{\partial y}$ evaluated at points near (x_0, y_0).

Example 4.3 *The set of all square matrices* $M_{n\times n}$ *is a manifold by virtue of the obvious isomorphism* $M_{n\times n} \cong \mathbb{R}^{n^2}$. *The set* $\mathfrak{sym}(n, \mathbb{R})$ *of all symmetric matrices is an* $n(n+1)/2$*-dimensional manifold by virtue of the obvious 1-1 correspondence* $\mathfrak{sym}(n, \mathbb{R}) \cong \mathbb{R}^{n(n+1)/2}$ *given by using* $n(n+1)/2$ *independent entries in the upper triangle of the matrix as coordinates.*
Now the set $O(n, \mathbb{R})$ *of all* $n \times n$ *orthogonal matrices is a submanifold of* $M_{n\times n}$. *We can show this using Theorem 4.3 as follows. Consider the map* $f : M_{n\times n} \to \mathfrak{sym}(n, \mathbb{R})$ *given by* $A \mapsto A^t A$. *Notice that by definition of* $O(n, \mathbb{R})$ *we have* $f^{-1}(I) = O(n, \mathbb{R})$. *Let us compute the tangent map at any point* $Q \in f^{-1}(I) =$

$O(n,\mathbb{R})$. The tangent space of $\mathfrak{sym}(n,\mathbb{R})$ at I is $\mathfrak{sym}(n,\mathbb{R})$ itself since $\mathfrak{sym}(n,\mathbb{R})$ is a vector space. Similarly, $M_{n\times n}$ is its own tangent space. Under the identifications of section 3.6 we have

$$T_Q f \cdot v = \frac{d}{ds}(Q^t + sv^t)(AQ + sv) = v^t Q + Q^t v.$$

Now this map is clearly surjective onto $\mathfrak{sym}(n,\mathbb{R})$ when $Q = I$. On the other hand, for any $Q \in O(n,\mathbb{R})$ consider the map $L_{Q^{-1}} : M_{n\times n} \to M_{n\times n}$ given by $L_{Q^{-1}}(B) = Q^{-1}B$. The map $T_Q L_{Q^{-1}}$ is actually just $T_Q L_{Q^{-1}} \cdot v = Q^{-1}v$ which is a linear isomorphism since Q is a nonsingular. We have that $f \circ L_Q = f$ and so by the chain rule

$$\begin{aligned} T_Q f \cdot v &= T_I f \circ T_Q(L_{Q^{-1}}) \cdot v \\ &= T_I f \cdot Q^{-1} v \end{aligned}$$

which shows that $T_Q f$ is also surjective.

The following proposition shows an example of the simultaneous use of Sard's theorem and theorem4.3.

Proposition 4.2 *Let M be a connected submanifold of $\mathbb{R}^n$ and let S be a linear subspace of $\mathbb{R}^n$. Then there exist a vector $v \in \mathbb{R}^n$ such that $(v + S) \cap M$ is a submanifold of M.*

Proof. Start with a line l through the origin that is normal to S. Let $pr : \mathbb{R}^n \to S$ be orthogonal projection onto l . The restriction $\pi := pr|_M \to l$ is easily seen to be smooth. If $\pi(M)$ were just a single point x then $\pi^{-1}(x)$ would be all of M. Now $\pi(M)$ is connected and a connected subset of $l \cong \mathbb{R}$ must contain an interval which means that $\pi(M)$ has positive measure. Thus by Sard's theorem there must be a point $v \in l$ which is a regular value of π. But then 4.3 implies that $\pi^{-1}(v)$ is a submanifold of M. But this is the conclusion since $\pi^{-1}(v) = (v + S) \cap M$. ∎

4.7 Morse Functions

If we consider a smooth function $f : M \to \mathbb{R}$ and assume that M is a compact manifold (without boundary) then f must achieve both a maximum at one or more points of M and a minimum at one or more points of M. Let x_e be one of these points. The usual argument shows that $df|_e = 0$ (Recall that under the usual identification of $\mathbb{R}$ with any of its tangent spaces we have $df|_e = T_e f$). Now let x_0 be some point for which $df|_{x_0} = 0$. Does f achieve either a maximum or a minimum at x_0? How does the function behave in a neighborhood of x_0? As the reader may well be aware, these questions are easier to answer in case the second derivative of f at x_0 is nondegenerate. But what is the second derivative in this case? One could use a Riemannian metric and the corresponding Levi-Civita connection to be introduced later to given an invariant notion but for us it will suffice to work in a local coordinate system and restrict our attention to critical points. Under these conditions the following definition of nondegeneracy is well defined independent of the choice of coordinates:

Definition 4.8 *The **Hessian** matrix of f at one of its critical points x_0 and with respect to coordinates $\psi = (x^1, ..., x^n)$ is the matrix of second partials:*

$$H = \begin{bmatrix} \frac{\partial^2 f\circ\psi^{-1}}{\partial x^1 \partial x^1}(x_0) & \cdots & \frac{\partial^2 f\circ\psi^{-1}}{\partial x^1 \partial x^n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial^2 f\circ\psi^{-1}}{\partial x^n \partial x^1}(x_0) & \cdots & \frac{\partial^2 f\circ\psi^{-1}}{\partial x^n \partial x^n}(x_0) \end{bmatrix}$$

The critical point is called nondegenerate if H is nonsingular

Now any such matrix H is symmetric and by Sylvester's law of inertia this matrix is equivalent to a diagonal matrix whose diagonal entries are either 1 or -1. The number of -1 occurring is called the **index** of the critical point.

Exercise 4.3 *Show that the nondegeneracy is well defined.*

Exercise 4.4 *Show that nondegenerate critical points are isolated. Show by example that this need not be true for general critical points.*

The structure of a function near one of its nondegenerate critical points is given by the following famous theorem of M. Morse:

Theorem 4.8 (Morse Lemma) *If $f : M \to \mathbb{R}$ is a smooth function and x_0 is a nondegenerate critical point for f of index i. Then there is a local coordinate system $U, \mathtt{x}$ containing x_0 such that the local representative $f_U := f \circ \mathtt{x}^{-1}$ for has the form*

$$f_U(x^1, ..., x^n) = f(x_0) + \sum h_{ij} x^i x^j$$

and where it may be arranged that the matrix $h = (h_{ij})$ is a diagonal matrix of the form $diag(-1, ... - 1, 1, ..., 1)$ for some number (perhaps zero) of ones and minus ones. The number of minus ones is exactly the index i.

Proof. This is clearly a local problem and so it suffices to assume $f :: \mathbb{R}^n \to \mathbb{R}$ and also that $f(0) = 0$. Then our task is to show that there exists a diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$ such that $f \circ \phi(x) = x^t h x$ for a matrix of the form described. The first step is to observe that if $g : U \subset \mathbb{R}^n \to \mathbb{R}$ is any function defined on a convex open set U and $g(0) = 0$ then

$$\begin{aligned} g(u_1, ..., u_n) &= \int_0^1 \frac{d}{dt} g(tu_1, ..., tu_n) dt \\ &= \int_0^1 \sum_{i=1}^n u_i \partial_i g(tu_1, ..., tu_n) dt \end{aligned}$$

Thus g is of the form $g = \sum_{i=1}^{n} u_i g_i$ for certain smooth functions $g_i, 1 \leq i \leq n$ with the property that $\partial_i g(0) = g_i(0)$. Now we apply this procedure first to f to get $f = \sum_{i=1}^{n} u_i f_i$ where $\partial_i f(0) = f_i(0) = 0$ and then apply the procedure to each f_i and substitute back. The result is that

$$f(u_1, ..., u_n) = \sum_{i,j=1}^{n} u_i u_j h^{ij}(u_1, ..., u_n) \tag{4.2}$$

for some functions h^{ij} with the property that $h^{ij}()$ is nonsingular at and therefore near 0. Next we symmetrize (h^{ij}) by replacing it with $\frac{1}{2}(h^{ij} + h^{ji})$ if necessary. This leaves the expression 4.2 untouched. Now the index of the matrix $(h^{ij}(0))$ is i and this remains true in a neighborhood of 0. The trick is to find, for each x in the neighborhood a matrix $C(x)$ which effects the diagonalization guaranteed by Sylvester's theorem: $D = C(x)h(x)C(x)^{-1}$. The remaining details, including the fact that the matrix $C(x)$ may be chosen to depend smoothly on x, is left to the reader. ∎

We can generalize theorem 4.3 using the concept of transversality .

Definition 4.9 *Let $f : M \to N$ be a smooth map and $S \subset N$ a submanifold of N. We say that f is* ***transverse to*** *S if for every $p \in f^{-1}(S)$ the image of T_pM under the tangent map T_pf and $T_{f(p)}S$ together span all of $T_{f(p)}N$:*

$$T_{f(p)}N = T_{f(p)}S + T_pf(T_pM).$$

If f is ***transverse to*** *S we write $f \pitchfork S$.*

Theorem 4.9 *Let $f : M \to N$ be a smooth map and $S \subset N$ a submanifold of N and suppose that $f \pitchfork S$. Then $f^{-1}(S)$ is a submanifold of M. Furthermore we have $T_p(f^{-1}(S)) = T_{f(p)}f^{-1}(T_{f(p)}S)$ for all $p \in f^{-1}(S)$ and codim($f^{-1}(S)$)=codim(S).*

4.8 Problem set

1. Show that a submersion always maps open set to open set (it is an open mapping). Further show that if M is compact and N connected then a submersion $f : M \to N$ is a submersion must be surjective.

2. Show that the set of all symmetric matrices $Sym_{n\times n}(\mathbb{R})$ is a submanifold of $M_{n\times n}(\mathbb{R})$. Under the canonical identification of $T_S M_{n\times n}(\mathbb{R})$ with $M_{n\times n}(\mathbb{R})$ the tangent space of $Sym_{n\times n}(\mathbb{R})$ at the symmetric matrix S becomes what subspace of $M_{n\times n}(\mathbb{R})$? Hint: It's kind of a trick question.

3. Prove that in each case below the subset $f^{-1}(p_0)$ is a submanifold of M:

   a) $f : M = M_{n\times n}(\mathbb{R}) \to Sym_{n\times n}(\mathbb{R})$ and $f(Q) = Q^t Q$ with $p_0 = I$ the identity matrix.

   b) $f : M = \mathbb{R}^n \to \mathbb{R}^2$ and $f(x_1, ..., x_n) = ({x_1}^2 + \cdots + {x_n}^2, -{x_1}^2 + \cdots + {x_n}^2)$ with $p_0 = (0, 0)$.

4. Show that the set of all $m \times n$ matrices is a submanifold of $M_{m\times n}(\mathbb{R})$.

5. Show that if $p(x) = p(x_1, ..., x_n)$ is a homogeneous polynomial so that for some $m \in \mathbb{Z}_+$
$$p(tx_1, ..., tx_n) = t^m p(x_1, ..., x_n)$$
then as long as $c \neq 0$ the set $p^{-1}(c)$ is a $n-1$ dimensional submanifold of $\mathbb{R}^n$.

6. Suppose that $g : M \to N$ is transverse to a submanifold $W \subset N$. For another smooth map $f : Y \to M$ show that $f \pitchfork g^{-1}(N)$ if and only if $(g \circ f) \pitchfork W$.

7. Suppose that $c : [a, b] \to M$ is a smooth map. Show that given any compact subset $C \subset (a, b)$ and any $\epsilon > 0$ there is an immersion $\gamma : (a, b) \to M$ which agrees with c on the set C and such that
$$|\gamma(t) - c(t)| \leq \epsilon \text{ for all } t \in (a, b).$$

Chapter 5

Lie Groups I

5.1 Definitions and Examples

One approach to geometry is to view geometry as the study of invariance and symmetry. In our case we are interested in studying symmetries of differentiable manifolds, Riemannian manifolds, symplectic manifolds etc. Now the usual way to talk about symmetry in mathematics is by the use of the notion of a transformation group. The wonderful thing for us is that the groups that arise in the study of manifold symmetries are themselves differentiable (even analytic) manifolds. Perhaps the reader should keep in mind the prototypical example of the Euclidean motion group and the subgroup of rotations about a point in Euclidean space $\mathbb{R}^n$. The rotation group is usually represented as the group of orthogonal matrices $O(n)$ and then the action on $\mathbb{R}^n$ is given in the usual way as

$$x \mapsto Qx$$

for $Q \in O(n)$ and x a column vector from $\mathbb{R}^n$. It is pretty clear that this group must play a big role in physical theories. Of course, translation is important too and the combination of rotations and translations generate the group of Euclidean motions $E(n)$. We may represent a Euclidean motion as a matrix using the following trick: If we want to represent the transformation $x \to Qx+b$ where $x, b \in \mathbb{R}^n$ we can achieve this by letting column vectors

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \in \mathbb{R}^{n+1}$$

represent elements x. The we form

$$\begin{bmatrix} 1 & 0 \\ b & Q \end{bmatrix}$$

and notice that

$$\begin{bmatrix} 1 & 0 \\ b & Q \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ b+Qx \end{bmatrix}.$$

Thus we may identify $E(n)$ with the group of $(n+1) \times (n+1)$ matrices of the above form where $Q \in O(n)$. Because of the above observation we say that $E(n)$ is the semidirect product of $O(n)$ and $\mathbb{R}^n$ and write $O(n) \rightthreetimes \mathbb{R}^n$.

Exercise 5.1 *Make sense of the statement* $E(n)/O(n) \cong \mathbb{R}^n$.

Lie groups place a big role in classical mechanics (symmetry and conservation laws), fluid mechanic where the Lie group may be infinite dimensional and especially in particle physics (Gauge theory). In mathematics, Lie groups play a prominent role Harmonic analysis (generalized Fourier theory) and group representations as well as in many types of geometry including Riemannian geometry, algebraic geometry, Kähler geometry, symplectic geometry and also topology.

Definition 5.1 *A* C^∞ *differentiable manifold* G *is called a Lie group if it is a group (abstract group) such that the multiplication map* $\mu : G \times G \to G$ *and the inverse map* $\nu : G \to G$ *given by* $\mu(g,h) = gh$ *and* $\nu(g) = g^{-1}$ *are* C^∞ *functions.*

If G and H are Lie groups then so is the **product group** $G \times H$ where multiplication is $(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2)$. Also, if H is a subgroup of a Lie group G that is also a regular closed submanifold then H is a Lie group itself and we refer to H as a (regular) **Lie subgroup**. It is a nontrivial fact that an abstract subgroup of a Lie group that is also a closed subset is automatically a Lie subgroup (see 8.4).

Example 5.1 $\mathbb{R}$ *is a one-dimensional (abelian Lie group) were the group multiplication is addition. Similarly, any vector space, e.g.* $\mathbb{R}^n$, *is a Lie group and the vector addition.*

Example 5.2 *The circle* $S^1 = \{z \in \mathbb{C} : |z|^2 = 1\}$ *is a 1-dimensional (abelian) Lie group under complex multiplication. More generally, the torus groups are defined by* $T^n = \underbrace{S^1 \times \ldots \times S^1}_{n\text{-times}}$.

Now we have already introduced the idea of a discrete group action. We are also interested in more general groups action. Let us take this opportunity to give the relevant definitions.

Definition 5.2 *A left action of a Lie group* G *on a manifold* M *is a smooth map* $\lambda : G \times M \to M$ *such that* $\lambda(g_1, \lambda(g_2, m)) = \lambda(g_1 g_2, m))$ *for all* $g_1, g_2 \in G$. *Define the partial map* $\lambda_g : M \to M$ *by* $\lambda_g(m) = \lambda(g, m)$ *and then the requirement is that* $\widetilde{\lambda} : g \mapsto \lambda_g$ *is a group homomorphism* $G \to \mathrm{Diff}(M)$. *We often write* $\lambda(g, m)$ *as* $g \cdot m$.

Example 5.3 (The General Linear Group) *The set of all non-singular matrices* $\mathrm{GL}(n, \mathbb{R})$ *under multiplication.*

Example 5.4 (The Orthogonal Group) *The set of all orthogonal matrices* $O(n, \mathbb{R})$ *under multiplication. Recall that an orthogonal matrix* A *is one for* *which* $AA^t = I$*. Equivalently, an orthogonal matrix is one which preserves the standard inner product on* R^n*:*

$$Ax \cdot Ay = x \cdot y.$$

More generally, let $\mathsf{V}, b = \langle .,. \rangle$ *be any* n*- dimensional inner product space (*b *is just required to be nondegenerate) and let* k *be the maximum of the dimensions of subspaces on which* b *is positive definite. Then define*

$$O_{k,n-k}(\mathsf{V},b) = \{A \in \mathrm{GL}(\mathsf{V}) : \langle Av, Aw \rangle = \langle v, w \rangle \text{ for all } v, w \in \mathsf{V}\}.$$

$O_{k,n-k}(\mathsf{V},b)$ *turns out to be a Lie group. As an example, take inner product on* $\mathbb{R}^n$ *defined by*

$$\langle x, y \rangle := \sum_{i=1}^{k} x^i y^i - \sum_{i=1+1}^{n} x^i y^i$$

and denote the corresponding group by $O_{k,n-k}(\mathbb{R}^n)$*. We can identify this group with the group* $O(k, n-k)$ *of all matrices* A *such that* $A\Lambda A^t = \Lambda$ *where*

$$\Lambda = diag(\underset{k \ times}{1, ...,} \quad \underset{n-k \ times}{-1, ...}).$$

$O(1,3)$ *is called the Lorentz group. If we further require that the elements of our group have determinant equal to one we get the groups denoted* $SO(k, n-k)$*.*

Exercise 5.2 *Show that if* $0 < k < n$ *then* $SO(k, n-k)$ *has two connected components.*

Example 5.5 (The Special Orthogonal Group) *The set of all orthogonal matrices having determinant equal to one* $SO(n, \mathbb{R})$ *under multiplication. This is the* $k = n$ *case of the previous example. As a special case we have the rotation group* $SO(3, \mathbb{R}) = SO(3)$*.*

Example 5.6 (The Unitary Group) $U(2)$ *is the group of all* 2×2 *complex matrices* A *such that* $A\bar{A}^t = I$*.*

Example 5.7 (The Special Unitary Group) $SU(2)$ *is the group of all 2x2 complex matrices* A *such that* $A\bar{A}^t = I$ *and* $\det(A) = 1$*. There is a special relation between* $SO(3)$ *and* $SU(2)$ *that is important in quantum physics related to the notion of spin. We will explore this in detail later in the book.*

Notation 5.1 $A^* := \bar{A}^t$

Example 5.8 *The set of all* $A \in \mathrm{GL}(n, \mathbb{R})$ *with determinant equal to one is called the special linear group and is denoted* $SL(n, \mathbb{R})$*.*

Example 5.9 *Consider the matrix* $2n \times 2n$ *matrix*

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

J defines a skew-symmetric bilinear ω_0 form on $\mathbb{R}^{2n}$ by $\omega_0(v,w) = v^t Jw$. The set of all matrices in $\mathbb{M}_{2n\times 2n}$ that preserve this bilinear form is called the ***symplectic group*** *$Sp(n,\mathbb{R})$ and can be defined also as*

$$Sp(n,\mathbb{R}) = \{S \in \mathbb{M}_{2n\times 2n} : S^t JS = J.$$

This group is important in Hamiltonian mechanics. If $e_1, ..., e_n, ..., e_{2n}$ is a the standard basis of $\mathbb{R}^{2n}$ then $\omega_0 = \sum_{i=1}^{n} e_i \wedge e_{i+n}$. We will see that in some basis every nondegenerate 2-form on $\mathbb{R}^{2n}$ takes this form.

Exercise 5.3 *Show that $SU(2)$ is simply connected.*

Exercise 5.4 *Let $g \in G$ (a Lie group). Show that each of the following maps $G \to G$ is a diffeomorphism:*
1) $L_g : x \mapsto gx$ (left translation)
2) $R_g : x \mapsto xg$ (right translation)
3) $C_g : x \mapsto gxg^{-1}$ (conjugation).
4) $inv : x \mapsto x^{-1}$ (inversion)

5.2 Lie Group Homomorphisms

The following definition is manifestly natural:

Definition 5.3 *A smooth map $f : G \to H$ is called a* ***Lie group homomorphism*** *if*

$$f(g_1 g_2) = f(g_1)f(g_2) \text{ for all } g_1, g_2 \in G \text{ and}$$
$$f(g^{-1}) = f(g)^{-1} \text{ for all } g \in G.$$

and an isomorphism in case it has an inverse which is also a Lie group homomorphism. A Lie group isomorphism $G \to G$ is called a Lie group ***automorphism.***

Example 5.10 *The inclusion $SO(n,\mathbb{R}) \hookrightarrow Gl(n,\mathbb{R})$ is a Lie group homomorphism.*

Example 5.11 *The circle $S^1 \subset \mathbb{C}$ is a Lie group under complex multiplication and the map*

$$z = e^{i\theta} \to \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is a Lie group homomorphism into $SO(n)$.

Example 5.12 *The conjugation map $C_g : G \to G$ is a Lie group automorphism.*

Exercise 5.5 (*) *Show that the multiplication map $\mu : G\times G \to G$ has tangent map at the identity $(e,e) \in G \times G$ given as $T_{(e,e)}\mu(v,w) = v + w$. Recall that we identify $T_{(e,e)}(G \times G)$ with $T_eG \times T_eG$.*

Exercise 5.6 *Show that $Gl(n,\mathbb{R})$ is an open subset of the vector space of all $n \times n$ matrices and that it is a Lie group. Using the natural identification of $T_eGl(n,\mathbb{R})$ with $M_{n\times n}(\mathbb{R})$ show that as a map $M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ we have*

$$T_eC_g : x \mapsto gxg^{-1}$$

where $g \in Gl(n,\mathbb{R})$ and $x \in M_{n\times n}(\mathbb{R})$.

Example 5.13 *The map $t \mapsto e^{it}$ is a Lie group homomorphism from $\mathbb{R}$ to $S^1 \subset \mathbb{C}$.*

Definition 5.4 *A **closed (Lie) subgroup** H of a Lie group G is a closed regular submanifold that is also a Lie group with respect to the submanifold differentiable structure.*

Example 5.14 *S^1 embedded as $S^1 \times \{1\}$ in the torus $S^1 \times S^1$ is a subgroup.*

Example 5.15 *$SO(n)$ is a subgroup of both $O(n)$ and $\mathrm{GL}(n)$.*

Definition 5.5 *An **immersed subgroup** of a Lie group G is a smooth injective immersion (or it's image) which is also a Lie group homomorphism.*

Remark 5.1 *It is an unfortunate fact that in this setting a map is referred to as a "submanifold". This is a long standing tradition but we will avoid this terminology as much as possible preferring to call maps like those in the last definition simply **injective homomorphisms**.*

Definition 5.6 *A homomorphism from the additive group $\mathbb{R}$ into a Lie group is called a **one-parameter subgroup**.*

Example 5.16 *The torus $S^1 \times S^1$ is a Lie group under multiplication given by $(e^{i\tau_1}, e^{i\theta_1})(e^{i\tau_2}, e^{i\theta_2}) = (e^{i(\tau_1+\tau_2)}, e^{i(\theta_1+\theta_2)})$. Every homomorphism of $\mathbb{R}$ into $S^1 \times S^1$, that is, ever one parameter subgroup of $S^1 \times S^1$ is of the form $t \mapsto (e^{tai}, e^{tbi})$ for some pair of real numbers $a, b \in \mathbb{R}$.*

Example 5.17 *The map $R : \mathbb{R} \to SO(3)$ given by*

$$\theta \mapsto \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is a one parameter subgroup. (an hence a homomorphism).

Example 5.18 *Given an element g of the group $SU(2)$ we define the map $Ad_g : \mathfrak{su}(2) \to \mathfrak{su}(2)$ by $Ad_g : x \mapsto gxg^{-1}$. Now the skew-Hermitian matrices of zero trace can be identified with $\mathbb{R}^3$ by using the following matrices as a basis:*

$$\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

These are just $-i$ times the ***Pauli matrices*** *$\sigma_1, \sigma_2, \sigma_3$ [1] and so the correspondence $\mathfrak{su}(2) \to \mathbb{R}^3$ is $-xi\sigma_1 - yi\sigma_2 - iz\sigma_3 \mapsto (x, y, z)$. Under this correspondence the inner product on $\mathbb{R}^3$ becomes the inner product $(A, B) = trace(AB)$. But then*

$$\begin{aligned}(\mathrm{Ad}_g A, \mathrm{Ad}_g B) &= \mathrm{trace}(gAgg^{-1}Bg^{-1}) \\ &= \mathrm{trace}(AB) = (A, B)\end{aligned}$$

so actually Ad_g can be thought of as an element of $O(3)$. More is true; Ad_g acts as an element of $SO(3)$ and the map $g \mapsto Ad_g$ is then a homomorphism from $SU(2)$ to $SO(\mathfrak{su}(2)) \cong SO(3)$.

The set of all Lie groups together with Lie group homomorphisms forms an important category which has much of the same structure as the category of all groups. The notions of kernel, image and quotient show up as expected.

Theorem 5.1 *Let $h : G \to H$ be a Lie group homomorphism. The subgroups $Ker(h) \subset G$, $Img(h) \subset H$ are Lie subgroups.*

[1] check this

Chapter 6

Fiber Bundles and Vector Bundles I

First of all lets get the basic idea down. Think about this: A function defined on the circle with range in the interval $(0, 1)$ can be thought of in terms of its graph. The latter is a subset, a cross section, of the product $S^1 \times (0, 1)$. Now, what would a similar cross section of the Mobius band signify? This can't be the same thing as before since a continuous cross section of the Mobius band would have to cross the center and this need not be so for $S^1 \times (0, 1)$. Such a cross section would have to be a sort of twisted function. The Mobius band (with projection onto its center line) provides us with our first nontrivial example of a fiber bundle. The cylinder $S^1 \times (0, 1)$ with projection onto one the factors is a trivial example. Projection onto S^1 gives a topological line bundle while projection onto the interval is a circle bundle. Often what we call the Mobius band will be the slightly different object which is a twisted version of $S^1 \times \mathbb{R}^1$. Namely, the space obtained by identifying one edge of $[0, 1] \times \mathbb{R}^1$ with the other but with a twist.

A fiber bundle is to be though of as a bundle of -or parameterized family of- spaces $E_x = \pi^{-1}(x) \subset E$ called the fibers. Nice bundles have further properties that we shall usually assume without explicit mention. The first one is simply that the spaces are Hausdorff and paracompact. The second one is called local triviality. In order to describe this we need a notion of a bundle map and the ensuing notion of equivalence of fiber bundles.

Most of what we do here will work either for the general topological category or for the smooth category so we once again employ the conventions of 2.6.1.

Definition 6.1 *A general C^r- bundle is a triple $\xi = (E, \pi, X)$ where $\pi : E \to M$ is a surjective C^r-map of C^r-spaces (called the bundle projection). For each $p \in X$ the subspace $E_p := \pi^{-1}(p)$is called the fiber over p. The space E is called the* ***total space*** *and X is the* ***base space****. If $S \subset X$ is a subspace we can always form the restricted bundle (E_S, π_S, S) where $E_S = \pi^{-1}(S)$ and $\pi_S = \pi|_S$ is the restriction.*

Definition 6.2 *A (C^r-) **section** of a general bundle $\pi_E : E \to M$ is a (C^r-) map $s : M \to E$ such that $\pi_E \circ s = id_M$. In other words, the following diagram must commute:*

$$\begin{array}{ccccc} s & & E & & \\ & \nearrow & \downarrow & \pi_E & \\ M & \rightarrow & M & & . \\ & id & & & \end{array}$$

The set of all C^r-sections of a general bundle $\pi_E : E \to M$ is denoted by $\Gamma^k(M, E)$. We also define the notion of a section over an open set U in M is the obvious way and these are denoted by $\Gamma^k(U, E)$.

Notation 6.1 *We shall often abbreviate to just $\Gamma(U, E)$ or even $\Gamma(E)$ whenever confusion is unlikely. This is especially true in case $k = \infty$ (smooth case) or $k = 0$ (continuous case).*

Now there are two different ways to treat bundles as a category:

The Category *Bun*.

Actually, we should define the Categories Bun_k ;$k = 0, 1, ..., \infty$ and then abbreviate to just "Bun" in cases where a context has been establish and confusion is unlikely. The objects of Bun_k are C^r-fiber bundles.

Definition 6.3 *A morphism from $\mathrm{Hom}_{Bun_k}(\xi_1, \xi_2)$, also called a bundle map from a C^r-fiber bundle $\xi_1 := (E_1, \pi_1, X_1)$ to another fiber bundle $\xi_2 := (E_2, \pi_2, X_2)$ is a pair of C^r-maps $(\bar{f}, f)$ such that the following diagram commutes:*

$$\begin{array}{ccc} E_1 & \xrightarrow{\bar{f}} & E_2 \\ \downarrow & & \downarrow \\ X_1 & \xrightarrow{f} & X_2 \end{array}$$

If both maps are C^r-isomorphisms we call the map a (C^r-) bundle isomorphism.

Definition 6.4 *Two fiber bundles $\xi_1 := (E_1, \pi_1, X_1)$ and $\xi_2 := (E_2, \pi_2, X_2)$ are **equivalent** in Bun_k or **isomorphic** if there exists a bundle isomorphism from ξ_1 to ξ_2.*

Definition 6.5 *Two fiber bundles $\xi_1 := (E_1, \pi_1, X_1)$ and $\xi_2 := (E_2, \pi_2, X_2)$ are said to be **locally equivalent** if for any $y \in E_1$ there is an open set U containing p and a bundle equivalence $(f, \bar{f})$ of the restricted bundles:*

$$\begin{array}{ccc} E_1|\,U & \xrightarrow{\bar{f}} & E_2|\,f(U) \\ \downarrow & & \downarrow \\ U & \xrightarrow{f} & f(U) \end{array}$$

The Category $Bun_k(X)$

Definition 6.6 *A morphism from* $\mathrm{Hom}_{Bun_k(X)}(\xi_1, \xi_2)$*, also called a bundle map over* X *from a* C^r-*fiber bundle* $\xi_1 := (E_1, \pi_1, X)$ *to another fiber bundle* $\xi_2 := (E_2, \pi_2, X)$ *is a* C^r-*map* $\bar{f}$ *such that the following diagram commutes:*

$$\begin{array}{ccc} E_1 & \xrightarrow{\bar{f}} & E_2 \\ & \searrow \quad \swarrow & \\ & X & \end{array}$$

If both maps are C^r*-isomorphisms we call the map a (*C^r*-)* ***bundle isomorphism over*** X *(also called a* ***bundle equivalence****).*

Definition 6.7 *Two fiber bundles* $\xi_1 := (E_1, \pi_1, X)$ *and* $\xi_2 := (E_2, \pi_2, X)$ *are* ***equivalent*** *in* $Bun_k(X)$ *or* ***isomorphic*** *if there exists a (*C^r*-) bundle isomorphism over* X *from* ξ_1 *to* ξ_2*.*

By now the reader is no doubt tired of the repetitive use of the index C^r so from now on we will simple refer to space (or manifolds) and maps where the appropriate smoothness C^r will not be explicitly stated unless something only works for a specific value of r.

Definition 6.8 *Two fiber bundles* $\xi_1 := (E_1, \pi_1, X)$ *and* $\xi_2 := (E_2, \pi_2, X)$ *are said to be* ***locally equivalent*** *(over* X*) if for any* $y \in E_1$ *there is an open set* U *containing p and a bundle equivalence* $(\bar{f}, f)$ *of the restricted bundles:*

$$\begin{array}{ccc} E_1|U & \xrightarrow{\bar{f}} & E_2|U \\ \downarrow & & \downarrow \\ U & \xrightarrow{id} & U \end{array}$$

Now for any space X **the trivial bundle** with fiber F is the triple $(X \times F, pr_1, X)$ where pr_1 always denoted the projection onto the first factor. Any bundle over X that is bundle equivalent to $X \times F$ is referred to as a trivial bundle. We will now add in an extra condition that we will usually need:

Definition 6.9 *A* $(-)$ *fiber bundle* $\xi := (E, \pi, X)$ *is said to be* ***locally trivial*** *(with fiber* F*) if every for every* $x \in X$ *has an open neighborhood* U *such that* $\xi_U := (E_U, \pi_U, U)$ *is isomorphic to the trivial bundle* $(U \times F, pr_1, U)$*. Such a fiber bundle is called a* ***locally trivial fiber bundle****.*

We immediately make the following convention: **All fiber bundles in the book will be assumed to be locally trivial unless otherwise stated.** Once we have the local triviality it follows that each fiber $E_p = \pi^{-1}(p)$ is homeomorphic (in fact, $-$diffeomorphic) to F.

Notation 6.2 *We shall take the liberty of using a variety of notions when talking about bundles most of which are quite common and so the reader may as well get used to them. One, perhaps less common, notation which is very suggestive* *is writing $\xi := (F \hookrightarrow E \xrightarrow{\pi} X)$ to refer to a fiber bundle with typical fiber F. The notation suggests that F may be embedded into E as one of the fibers. This embedding is not canonical in general.*

It follows from the definition that there is a cover of E by **bundle charts** or (local) **trivializing maps** $f_\alpha : E_{U_\alpha} \to U_\alpha \times F$ such that

$$\begin{array}{ccc} E_{U_\alpha} & \xrightarrow{f_\alpha} & U_\alpha \times F \\ \pi \searrow & & \swarrow \\ & U_\alpha & \end{array}.$$

which in turn means that the so called overlap maps $f_\alpha \circ f_\beta^{-1} : U_\alpha \cap \grave{U} \times F \to U_\alpha \cap \grave{U} \times F$ must have the form $f_\alpha \circ f_\beta^{-1}(x,u) = (x, f_{\beta\alpha,x}(u))$ for maps $x \to f_{\beta\alpha,x} \in Diff^r(F)$ defined on each nonempty intersection $U_\alpha \cap \grave{U}$. These are called **transition functions** and we will assume them to be $-$maps whenever there is enough structure around to make sense of the notion. Such a cover by bundle charts is called a **bundle atlas**.

Definition 6.10 *It may be that there exists a C^r-group G (a Lie group in the smooth case) and a representation ρ of G in $Diff^r(F)$ such that for each nonempty $U_\alpha \cap U_\beta$ we have $f_{\beta\alpha,x} = \rho(g_{\alpha\beta}(x))$ for some C^r-map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$. In this case we say that G serves as a **structure group** for the bundle via the representation ρ. In case the representation is a faithful one then we may take G to be a subgroup of $Diff^r(F)$ and then we simply have $f_{\beta\alpha,x} = g_{\alpha\beta}(x)$. Alternatively, we may speak in terms of group actions so that G acts on F by $-$diffeomorphisms.*

The maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ must satisfy certain consistency relations:

$$\begin{aligned} g_{\alpha\alpha}(x) &= id \text{ for } x \in U_\alpha \\ g_{\alpha\beta}(x)g_{\beta\alpha}(x) &= id \text{ for } x \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(x)g_{\beta\gamma}(x)g_{\gamma\alpha}(x) &= id \text{ for } x \in U_\alpha \cap U_\beta \cap U_\gamma. \end{aligned} \tag{6.1}$$

A system of maps $g_{\alpha\beta}$ satisfying these relations is called a **cocycle** for the cover $\{U_\alpha\}$.

Definition 6.11 *A fiber bundle $\xi := (F \hookrightarrow E \xrightarrow{\pi} X)$ together with a G action on F is called a G-bundle if there exists a bundle atlas $\{(U_\alpha, f_\alpha)\}$ for ξ such that the overlap maps have the form $f_\alpha \circ f_\beta^{-1}(x,y) = (x, g_{\alpha\beta}(x) \cdot y))$ cocycle $\{g_{\alpha\beta}\}$ for the cover $\{U_\alpha\}$.*

Theorem 6.1 *Let G have C^r-action on F and suppose we are given cover of a C^r-space M and cocycle $\{g_{\alpha\beta}\}$ for a some cover $\{U_\alpha\}_{\alpha\in A}$. Then there exists a G-bundle with an atlas $\{(U_\alpha, f_\alpha)\}$ satisfying $f_\alpha \circ f_\beta^{-1}(x,y) = (x, g_{\alpha\beta}(x) \cdot y))$ on the overlaps.*

Proof. On the union $\Sigma := \bigcup_\alpha \{\alpha\} \times U_\alpha \times F$ define an equivalence relation such that

$$(\alpha, u, v) \in \{\alpha\} \times U_\alpha \times F$$

is equivalent to $(\beta, x, y) \in \{\beta\} \times U_\beta \times F$ iff $u = x$ and $v = g_{\alpha\beta}(x) \cdot y$.

The total space of our bundle is then $E := \Sigma/\sim$. The set Σ is essentially the disjoint union of the product spaces $U_\alpha \times F$ and so has an obvious topology. We then give $E := \Sigma/\sim$ the quotient topology. The bundle projection π_E is induced by $(\alpha, u, v) \mapsto u$. Notice that $\pi_E^{-1}(U_\alpha)$ To get our trivializations we define

$$f_\alpha(e) := (u, v) \text{ for } e \in \pi_E^{-1}(U_\alpha)$$

where (u, v) is the unique member of $U_\alpha \times F$ such that $(\alpha, u, v) \in e$. The point here is that $(\alpha, u_1, v_1) \sim (\alpha, u_2, v_2)$ only if $(u_1, v_1) = (u_2, v_2)$. Now suppose $U_\alpha \cap U_\beta \neq \emptyset$. Then for $x \in U_\alpha \cap U_\beta$ the element $f_\beta^{-1}(x, y)$ is in $\pi_E^{-1}(U_\alpha \cap U_\beta) = \pi_E^{-1}(U_\alpha) \cap \pi_E^{-1}(U_\beta)$ and so $f_\beta^{-1}(x, y) = [(\beta, x, y)] = [(\alpha, u, v)]$.

Which means that $x = u$ and $v = g_{\alpha\beta}(x) \cdot y$. From this it is not hard to see that

$$f_\alpha \circ f_\beta^{-1}(x, y) = (x, g_{\alpha\beta}(x) \cdot y)).$$

We leave the question of the regularity of the maps and the C^r structure to the reader. ∎

An important tool in the study of fiber bundles is the notion of a pullback bundle. We shall see that construction time and time again. Let $\xi = (F \hookrightarrow E \xrightarrow{\pi} M)$ be a -fiber bundle and suppose we have a -map $f : X \to M$. We want to define a fiber bundle $f^*\xi = (F \hookrightarrow f^*E \to X)$. As a set we have

$$f^*E = \{(x, e) \in X \times E : f(x) = \pi(e)\}.$$

The projection $f^*E \to X$ is the obvious one:$(x, e) \mapsto x \in N$.

Exercise 6.1 *Exhibit fiber bundle charts for f^*E.*

If a cocycle $\{g_{\alpha\beta}\}$ for a some cover $\{U_\alpha\}_{\alpha \in A}$ determine a bundle $\xi = (F \hookrightarrow E \xrightarrow{\pi} M)$ and $f : X \to M$ as above, then $\{g_{\alpha\beta} \circ f\} = \{f^* g_{\alpha\beta}\}$ is a cocycle for the same cover and the bundle determined by this cocycle is (up to isomorphism) none other than the pullback bundle $f^*\xi$.

$$\{g_{\alpha\beta}\} \rightsquigarrow \xi$$
$$\{f^* g_{\alpha\beta}\} \rightsquigarrow f^*\xi$$

The verification of this is an exercise that is easy but constitutes important experience so the reader should not skip the next exercise:

Exercise 6.2 *Verify that above claim.*

Exercise 6.3 *Show that if $A \subset M$ is a subspace of the base space of a bundle $\xi = (F \hookrightarrow E \xrightarrow{\pi} M)$ and $\iota : A \hookrightarrow M$ then $\iota^{-1}(\xi)$ is naturally isomorphic to the restricted bundle $\xi_A = (F \hookrightarrow E_A \to A)$.*

An important class of fiber bundles often studied on their own is the vector bundles. Roughly, a vector bundle is a fiber bundle with fibers being vector spaces. More precisely, we make the following definition:

Definition 6.12 *A real (or complex)* $(-)$ ***vector bundle*** *is a* $(-)$ *fiber bundle* (E, π, X) *such that*

(i) Each fiber $E_x := \pi^{-1}(x)$ *has the structure of a real (resp. complex) vector space.*

(ii) There exists a cover by bundle charts (U_α, f_α) *such that each restriction* $f_\alpha|_{E_x}$ *is a real (resp. complex) vector space isomorphism. We call these* ***vector bundle charts*** *or VB-charts. .*

The set of all vector bundles is a category **Vect**. Once again we need to specify the appropriate morphisms in this category and the correct choice should be obvious. A **vector bundle morphism** (or vector bundle map) between $\xi_1 := (E_1, \pi_1, X_1)$ and $\xi_2 := (E_2, \pi_2, X_2)$ is a bundle map $(\bar{f}, f)$:

$$\begin{array}{ccc} E_1 & \stackrel{\bar{f}}{\to} & E_2 \\ \downarrow & & \downarrow \\ X_1 & \stackrel{f}{\to} & X_2 \end{array}$$

which is linear on fibers. That is $\bar{f}\big|_{\pi_1^{-1}(x)}$ is a linear map from $\pi_1^{-1}(x)$ into the fiber $\pi_2^{-1}(x)$. We also have the category $Vect(X)$ consisting of all vector bundles over the fixed space X. Here the morphisms are bundle maps of the form (F, id_X). Two vector bundles $\xi_1 := (E_1, \pi_1, X)$ and $\xi_2 := (E_2, \pi_2, X)$ over the same space X are **isomorphic** (over X) if there is a bundle isomorphism over X from ξ_1 to ξ_2 which is a linear isomorphisms when restricted to each fiber. Such a map is called a vector bundle isomorphism.

In the case of vector bundles the transition maps are given by a representation of a Lie group G as a subgroup of $Gl(n, \mathbb{F})$. More precisely, if $\xi = (\mathbb{F}^k \hookrightarrow E \stackrel{\pi}{\to} M)$ there is a Lie group homomorphism $\rho : G \to Gl(k, \mathbb{F})$ such that for some VB-atlas $\{U_\alpha, f_\alpha\}$ we have the overlap maps

$$f_\alpha \circ f_\beta^{-1} : U_\alpha \cap U_\beta \times \mathbb{F}^k \to U_\alpha \cap U_\beta \times \mathbb{F}^k$$

are given by $f_\alpha \circ f_\beta^{-1}(x, v) = (x, \rho(g_{\alpha\beta}(x))v)$ for a cocycle $\{g_{\alpha\beta}\}$. In a great many cases, the representation is faithful and we may as well assume that $G \subset Gl(k, \mathbb{F})$ and that the representation is the standard one given by matrix multiplication $v \mapsto gv$. On the other hand we cannot restrict ourselves to this case because of our interest in the phenomenon of spin. A simple observation that gives a hint of what we are talking about is that if $G \subset G(n, \mathbb{R})$ acts on $\mathbb{R}^k$ by matrix multiplication and $h : \widetilde{G} \to G$ is a covering group homomorphism (or any Lie group homomorphism) then $v \mapsto g \cdot v := h(g)v$ is also action. Put another way, if we define $\rho_h : \widetilde{G} \to G(n, \mathbb{R})$ by $\rho_h(g) = h(g)v$ then ρ_h is representation of $\widetilde{G}$. The reason we care about this seemingly trivial fact only becomes apparent when we try to globalize this type of lifting as well will see when we study spin structures later on. To summarize this point we may say that whenever we have a VB-atlas $\{U_\alpha, f_\alpha\}$ we have the transition functions $\{f_{\alpha\beta}\} = \{x \mapsto f_{\beta\alpha,x}\}$ which are given straight from the overlap maps $f_\alpha \circ f_\beta^{-1}(x,u)$ by $f_\alpha \circ f_\beta^{-1}(x,u) = (x, f_{\beta\alpha,x}(u))$. Of course, the transition functions $\{f_{\alpha\beta}\}$ certainly form a cocycle for the cover $\{U_\alpha\}$ but there may be cases when we want a (not necessarily faithful) representation $\rho : G \to Gl(k, \mathbb{F}^n)$ of some group G not necessarily a subgroup of $Gl(k, \mathbb{F}^n)$ together with some $G-$valued cocycle $\{g_{\alpha\beta}\}$ such that $f_{\beta\alpha,x}(u) = \rho(g_{\alpha\beta}(x))$. Actually, we may first have to replace $\{U_\alpha, f_\alpha\}$ by a "refinement"; a notion we now define:

Definition 6.13 *A **refinement** of a VB-atlas $\{U_\alpha, f_\alpha\}$ is a VB-atlas $\{V_{\alpha'}, h_{\alpha'}\}_{\alpha' \in A'}$ such that for every $\alpha' \in A'$ there is a $\{U_\alpha, f_\alpha\}$ such that $U_\alpha \subset V_{\alpha'}$ and $h_{\alpha'}|_{U_\alpha} = f_\alpha$.*

We consider the refined atlas to be equivalent to the original. In fact, there is more to the notion of equivalence as the next result shows. For this theorem we comment that any two C^r-compatible atlases on a vector bundle have a common refinement (C^r-compatible) which the first two).

Theorem 6.2 *Let $\xi_1 = (\mathbb{F}^k \hookrightarrow E_1 \overset{\pi_1}{\to} M)$ and $\xi_2 = (\mathbb{F}^k \hookrightarrow E_2 \overset{\pi_2}{\to} M)$ be two vector bundles over M. Let $\{f_{\alpha\beta}\}$ and $\{h_{\alpha\beta}\}$ be the respective transition maps (which we assume to be over the same cover). Then ξ_2 is vector bundle isomorphic to ξ_1 if and only if there exist functions $\Delta_\alpha : U_\alpha \to Gl(k, \mathbb{F}^n)$ such that*

$$f_{\alpha\beta} = \Delta_\alpha h_{\alpha\beta} \Delta_\beta$$

whenever $U_\alpha \cap U_\beta$.

Example 6.1 (Canonical line bundle) *Recall that $P^n(\mathbb{R})$ is the set of all lines through the origin in $\mathbb{R}^{n+1}$. Define the subset $\mathbb{L}(P^n(\mathbb{R}))$ of $P^n(\mathbb{R}) \times \mathbb{R}^{n+1}$ consisting of all pairs (l, v) such that $v \in l$ (think about this). This set together with the map $\pi_{P^n(\mathbb{R})} : (l, v) \mapsto l$ is a rank one vector bundle. The bundle charts are of the form $\pi_{P^n(\mathbb{R})}^{-1}(U_i), \widetilde{\psi}_i$ where $\widetilde{\psi}_i : (l, v) \mapsto (\psi_i(l), pr_i(v)) \in \mathbb{R}^n \times \mathbb{R}$.*

Example 6.2 (Tautological Bundle) *Let $G(n, k)$ denote the Grassmannian manifold of k-planes in $\mathbb{R}^n$. Let $\gamma_{n,k}$ be the subset of $G(n, k) \times \mathbb{R}^n$ consisting of pairs (P, v) where P is a k-plane (k-dimensional subspace) and v is a vector in the plane P. The projection of the bundle is simply $(P, v) \mapsto P$. We leave it to the reader to discover an appropriate VB-atlas.*

Note well that these vector bundles are not just trivial bundles and in fact their topology (for large n) is of the up most importance for the topology of other vector bundles. One may take the inclusions $...R^n \subset R^{n+1} \subset ... \subset R^\infty$ to construct inclusions $...G(n, k) \subset G(n+1, k)$...and $...\gamma_{n,k} \subset \gamma_{n+1,k}$..from which a "universal bundle" $\gamma_n \to G(n)$ is constructed with the property that every rank k vector bundle E over X is the pull back by some map $f : X \to G(n)$:

$$\begin{array}{ccccc} E \cong & f^*\gamma_n & \to & \gamma_n \\ & \downarrow & & \downarrow \\ & X & \xrightarrow{f} & G(n) \end{array}$$

6.1 Transitions Maps and Structure

Now we come to an important point. Suppose we have a (locally trivial) fiber bundle $\xi := (E, \pi, X)$.We have seen that we can find a covering of the base space X by open sets U_α with corresponding trivializing maps $f_\alpha : E_{U_\alpha} \to U_\alpha \times F$. We have also noted that whenever $U_\alpha \cap U_\beta \neq \emptyset$ we have, for every point $x \in U_\alpha \cap U_\beta$, transition maps $f_{\alpha,\beta,x} : F \to F$. These may be may be seen to arise through the following composition:

$$y \mapsto (x, y) \mapsto f_\alpha \circ f_\beta|_{E_y}^{-1}(x, y) = (x, f_{\alpha\beta,x}(y)) \mapsto f_{\alpha\beta,x}(y)$$

Thus $f_{\alpha,\beta,x}$ is essentially the restriction of $f_\alpha \circ f_\beta^{-1}$ to $\{x\} \times F$. If the maps $f_{\alpha,\beta} : U_\alpha \cap U_\beta \to Diff^r(F)$ given by $x \mapsto f_{\alpha,\beta,x}$ often take values in some subgroup $G \subset Diff^r(F)$ which exists because the fiber F has some structure. For example it might be a finite dimensional $\mathbb{F}$-vector space (say $\mathbb{F}^k$) and then we hope to find a cover by trivializations (f_α, U_α) resulting in transition maps taking values in $Gl(k, \mathbb{F})$.

Exercise 6.4 *Show that if we have this situation;* $f_{\alpha,\beta} : U_\alpha \cap U_\beta \to Gl(k, \mathbb{F})$ *whenever* $U_\alpha \cap U_\beta \neq \emptyset$ *then there is defined a natural vector space structure on each fiber* E_x *and that we have a vector bundle as defined above in 6.12.*

As we indicated above, for a smooth (or C^r , $r > 0$) vector bundle require that all the maps are smooth (or C^r , $r > 0$) and in particular we require that $f_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(k, \mathbb{F})$ are all smooth.

Exercise 6.5 *To each point on a sphere attach the space of all vectors normal to the sphere at that point. Show that this* normal bundle *is in fact a (smooth) vector bundle. Also, in anticipation of the next section, do the same for the union of tangent planes to the sphere.*

Exercise 6.6 *Let* $Y = \mathbb{R} \times \mathbb{R}$ *and define let* $(x_1, y_1) \sim (x_2, y_2)$ *if and only if* $x_1 = x_2 + jk$ *and* $y_1 = (-1)^{jk} y_2$ *for some integer* k. *Show that* $E := Y/\sim$ *is a vector bundle of rank 1 which is trivial if and only if* j *is even. Convince yourself that this is the Mobius band when* j *is odd.*

6.2 Useful ways to think about vector bundles

It is often useful to think about vector bundles as a family of vector spaces parameterized by some topological space. So we just have a continuously varying

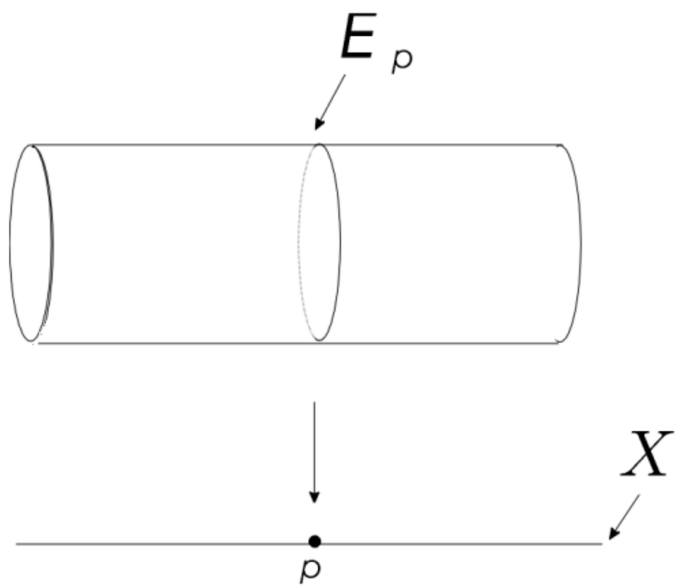


Figure 6.1: Circle bundle. Schematic for fiber bundle.

family $V_x : x \in X$. The fun comes when we explore the consequences of how these V_x fit together in a topological sense. We will study the idea of a tangent bundle below but for the case of a surface such as a sphere we can define the tangent bundle rather directly. For example, we may let $TS^2 = \{(x, v) : x \in S^2$ and $(x, v) = 0\}$. Thus the tangent vectors are pairs (x, v) where the x just tells us which point on the sphere we are at. The topology we put on TS^2 is the one we get by thinking of TS^2 as a subset of all pair, namely as a subset of $\mathbb{R}^3 \times \mathbb{R}^3$. The projection map is $\pi : (x, v) \mapsto x$.

Now there is another way we can associate a two dimensional vector space to each point of S^2. For this just associate a fixed copy of $\mathbb{R}^2$ to each $x \in S^2$ by taking the product space $S^2 \times \mathbb{R}^2$. This too is just a set of pairs but the topology is the product topology. Could it be that $S^2 \times \mathbb{R}^2 \to S^2$ and TS^2 are equivalent as vector bundles? The answer is no. The reason is that $S^2 \times \mathbb{R}^2 \to S^2$ is the trivial bundle and so have a nowhere zero section:

$$x \mapsto (x, v)$$

where v is any nonzero vector in $\mathbb{R}^2$. On the other hand it can be shown using the techniques of algebraic topology that TS^2 does not support a nowhere zero section. It follows that there can be no bundle isomorphism between the tangent bundle of the sphere $\pi : TS^2 \to S^2$ and the trivial bundle $pr_1 : S^2 \times \mathbb{R}^2 \to S^2$.

Recall that our original definition of a vector bundle (of rank k) did not include explicit reference to transition functions but they exist nonetheless.

Accordingly, another way to picture a vector bundle is to think about it as a bunch of trivial bundles $U_\alpha \times \mathbb{R}^k$ (or $U_\alpha \times \mathbb{C}^k$ for a complex bundle)) together with "change of coordinate" maps: $U_\alpha \times \mathbb{R}^k \to U_\beta \times \mathbb{R}^k$ given by $(x, v) \mapsto (x, g_{\alpha\beta}(x)v)$. So a point (vector) in a vector bundle from $\alpha' s$ view is a pair (x, v) while if $U_\alpha \cap U_\beta \neq \emptyset$ then from $\beta' s$ viewpoint it is the pair (x, w) where $w = g_{\alpha\beta}(x)v$. Now if this really does describe a vector bundle then the maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to Gl(k, \mathbb{R})$ (or $Gl(k, \mathbb{C})$) must satisfy certain consistency relations:

$$\begin{aligned} g_{\alpha\alpha}(x) &= id \text{ for } x \in U_\alpha \\ g_{\alpha\beta}(x)g_{\beta\alpha}(x) &= id \text{ for } x \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(x)g_{\beta\gamma}(x)g_{\gamma\alpha}(x) &= id \text{ for } x \in U_\alpha \cap U_\beta \cap U_\gamma. \end{aligned} \tag{6.2}$$

A system of maps $g_{\alpha\beta}$ satisfying these relations is called a cocycle for the bundle atlas.

Before After

Example 6.3 *Here is how the tangent bundle of the sphere looks in this picture:*

Let $U_{xy,+} = \{(x, y, z) \in S^2 : z > 0\}$

$U_{xy,+} = \{(x,y,z) \in S^2 : z > 0\}$

$$\begin{aligned}
U_{z+} &= \{(x,y,z) \in S^2 : z > 0\} \\
U_{z-} &= \{(x,y,z) \in S^2 : z < 0\} \\
U_{y+} &= \{(x,y,z) \in S^2 : y > 0\} \\
U_{y+} &= \{(x,y,z) \in S^2 : y < 0\} \\
U_{x+} &= \{(x,y,z) \in S^2 : x > 0\} \\
U_{x-} &= \{(x,y,z) \in S^2 : x < 0\}
\end{aligned}$$

Then for example $g_{y+,z+}(x,y,z)$ is the matrix

$$\begin{pmatrix} \frac{\partial}{\partial x} x & \frac{\partial}{\partial y} x \\ \frac{\partial}{\partial x}\sqrt{1-x^2-y^2} & \frac{\partial}{\partial y}\sqrt{1-x^2-y^2} \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 \\ -x/z & -y/z \end{pmatrix}$$

the rest are easily calculated using the Jacobian of the change of coordinate maps.

Or using just two open sets coming from stereographic projection: let U_1 be the open subset $\{(x,y,z) \in S^2 : z \neq -1\}$ and let U_2 be the open subset $\{(x,y,z) \in S^2 : z \neq 1\}$. Now we take the two trivial bundles $U_1 \times \mathbb{R}^2$ and $U_2 \times \mathbb{R}^2$ and then describe $g_{12}(p)$ for $p = (x,y,z) \in U_1 \cap U_2$.

$$g_{12}(p).(v_1,v_2)^t = (w_1,w_2)^t$$

where $h_{12}(p)$ is the matrix

$$\begin{pmatrix} -(x^2-y^2) & -2xy \\ -2xy & x^2-y^2 \end{pmatrix}.$$

This last matrix is not so easy to recognize. Can you see where I got it?

Theorem 6.3 *Let M be a C^r-space and suppose there is given an open cover $\{U_\alpha\}_{\alpha\in A}$ of M together with C^r-maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathbb{F}^k)$ satisfying the cocycle relations. Then there is a vector bundle $\pi_E : E \to M$ that has these as transition maps.*

6.3 Sections of a Vector Bundle

It is a general fact that if $\{V_x\}_{x\in X}$ is any family of $\mathbb{F}-$vector spaces and $\mathcal{F}(X,\mathbb{F})$ is the vector space of all $\mathbb{F}$-valued functions on the index set X then the set of maps (sections) of the form $\sigma : x \mapsto \sigma(x) \in V_x$ is a module over $\mathcal{F}(X,\mathbb{F})$ the operation being the obvious one. For us, the index set is some C^r-space and the family has the structure of a vector bundle. Let us restrict attention to C^∞ bundles. The space of sections $\Gamma(\xi)$ of a bundle $\xi = (\mathbb{F}^k \hookrightarrow E \longrightarrow M)$ is a vector space over $\mathbb{F}$ but also a over module over the ring of smooth functions

$C^\infty(M)$. The scalar multiplication is the obvious one: $(f\sigma)(x) := f(x)\sigma(x)$. Of course we also have local sections $\sigma : U \to E$ defined only on some open set. Piecing together local sections into global sections is whole topic in itself and leads to several topological constructions and in particular sheaf theory and sheaf cohomology. The first thing to observe about global sections of a vector bundle is that we always have plenty of them. But as we have seen, it is a different matter entirely if we are asking for global sections that never vanish; that is, sections σ for which $\sigma(x)$ is never the zero element of the fiber E_x.

6.4 Sheaves,Germs and Jets

In this section we introduce some formalism that will not only provide some convenient language but also provides conceptual tools. The objects we introduce here form an whole area of mathematics and may be put to uses far more sophisticated than we do here. Some professional mathematicians may find it vaguely offensive that we introduce these concepts without an absolute necessity for some specific and central problems. However, the conceptual usefulness of the ideas go a long way toward helping us organize our thoughts on several issues. So we persist.

There is a interplay in geometry and topology between local and global data. To see what the various meanings of the word local might be let consider a map $\digamma : C^\infty(M) \to C^\infty(M)$ which is not necessarily linear. For any $f \in C^\infty(M)$ we have a function $\digamma(f)$ and its value $(\digamma(f))(p) \in \mathbb{R}$ at some point $p \in M$. Let us consider in turn the following situations:

1. It just might be the case that whenever f and g agree on some neighborhood of p then $(\digamma(f))(p) = (\digamma(g))(p)$. So all that matters for determining $(\digamma(f))(p)$ is the behavior of f in any arbitrarily small open set containing p. To describe this we say that $(\digamma(f))(p)$ only depends on the "germ" of f at p.

2. Certainly if f and g agree on some neighborhood of p then they have the same Taylor expansion at p. The reverse is not true however. Suppose that whenever two functions f and g have Taylor series which agree up to and including terms of order $|x|^k$ then $(\digamma(f))(p) = (\digamma(g))(p)$. Then we say that $(\digamma(f))(p)$ depends only on the $k-$jet of f at p.

3. If $(\digamma(f))(p) = (\digamma(g))(p)$ whenever $df|_p = dg|_p$ we have a special case of the previous situation since $df|_p = dg|_p$ exactly when f and g have Taylor series which agree up to and including terms of order 1. So we are talking about the "$1-$jet".

4. Finally, it might be the case that $(\digamma(f))(p) = (\digamma(g))(p)$ exactly when p.

Of course it is also possible that none of the above hold at any point. Notice as we go down the list we are saying that the information needed to determine $(\digamma(f))(p)$ is becoming more and more local; even to the point of being infinitesimal.

Here is a simple example. Let $F : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be defined by $F(f) := f'$. Then it is pretty clear that $F(f)(t)$ depends only on the 1-jet at t (for any t).

Exercise 6.7 *Let $(a^{ij})_{1\leq i,j\leq n}$ be a matrix of smooth functions on $\mathbb{R}^n$ and also let $b^1, ...b^n, a$ be smooth functions on $\mathbb{R}^n$. Observe that the map $L : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ by*

$$L(f) = \sum_{i,j} a^{ij} \frac{\partial^2 f}{\partial x^i \partial x^j} + \sum_i b^i \frac{\partial f}{\partial x^i} + a$$

is such that for any $x \in \mathbb{R}^n$ the value $L(f)(x)$ depends only on the $2-$jet of f at x. Show that any linear *map $L : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ which has this property must have the above form.*

Now let us consider a section $\sigma : M \to E$ of a vector bundle E. Given any open set $U \subset M$ we may always produce the restricted section $\sigma|_U : U \to E$. This gives us a family of sections; on for each open set U. To reverse the situation, suppose that we have a family of sections $\sigma_U : U \to E$ where U varies over the open sets (or just a cover of M). When is it the case that such a family is just the family of restrictions of some (global) section $\sigma : M \to E$? This is one of the basic questions of sheaf theory. .

Definition 6.14 *A* ***presheaf of abelian groups (resp. rings etc.)*** *on a manifold (or more generally a topological space M is an assignment $\mathcal{M}(U)$ to each open set $U \subset M$ together with a family of abelian group (resp. ring etc.) homomorphisms $r^U_V : \mathcal{M}(U) \to \mathcal{M}(V)$ for each nested pair $V \subset U$ of open sets and such that*

Presheaf 1 $r^V_W \circ r^U_V = r^U_W$ whenever $W \subset V \subset U$.

Presheaf 2 $r^V_V = \mathrm{id}_V$ for all open $V \subset M$.

Definition 6.15 *Let $\mathcal{M}$ be a presheaf and $\mathcal{R}$ a presheaf of rings. If for each open $U \subset M$ we have that $\mathcal{M}(U)$ is a module over the ring $\mathcal{R}(U)$ and if the multiplication map $\mathcal{R}(U)\times\mathcal{M}(U) \to \mathcal{M}(U)$ commutes with the restriction maps r^U_W then we say that $\mathcal{M}$ is a presheaf of modules over $\mathcal{R}$.*

The best and most important example of a presheaf is the assignment $U \mapsto \Gamma(E, U)$ for some vector bundle $E \to M$ and where by definition $r^U_V(s) = s|_V$ for $s \in \Gamma(E, U)$. In other words r^U_V is just the restriction map. Let us denote this presheaf by $\mathcal{S}_E$ so that $\mathcal{S}_E(U) = \Gamma(E, U)$.

Definition 6.16 *A presheaf homomorphism $h : \mathcal{M}_1 \to \mathcal{M}_2$ is an is an assignment to each open set $U \subset M$ an abelian group (resp. ring, module, etc.) morphism $h_U : \mathcal{M}_1(U) \to \mathcal{M}_2(U)$ such that whenever $V \subset U$ then the following diagram commutes:*

$$\begin{array}{ccc} \mathcal{M}_1(U) & \overset{h_U}{\to} & \mathcal{M}_2(U) \\ r^U_V \downarrow & & r^U_V \downarrow \\ \mathcal{M}_1(V) & \overset{h_V}{\to} & \mathcal{M}_2(V) \end{array} .$$

Note we have used the same notation for the restriction maps of both presheaves.

Definition 6.17 *We will call a presheaf a* ***sheaf*** *if the following properties hold whenever $U = \bigcup_{U_\alpha \in \mathcal{U}} U_\alpha$ for some collection of open sets $\mathcal{U}$.*

Sheaf 1 *If $s_1, s_2 \in \mathcal{M}(U)$ and $r^U_{U_\alpha} s_1 = r^U_{U_\alpha} s_2$ for all $U_\alpha \in \mathcal{U}$ then $s_1 = s_2$.*

Sheaf 2 *If $s_\alpha \in \mathcal{M}(U_\alpha)$ and whenever $U_\alpha \cap U_\beta \neq \emptyset$ we have*

$$r^U_{U_\alpha \cap U_\beta} s_\alpha = r^U_{U_\alpha \cap U_\beta} s_\beta$$

then there exists a $s \in \mathcal{M}(U)$ such that $r^U_{U_\alpha} s = s_\alpha$.

If we need to indicate the space M involved we will write $\mathcal{M}_M$ instead of $\mathcal{M}$.

It is easy to see that all of the following examples are sheaves. In each case the maps $r^U_{U_\alpha}$ are just the restriction maps.

Example 6.4 (Sheaf of smooth functions) $\mathcal{C}^\infty_M(U) = C^\infty(U)$. *Notice that $\mathcal{C}^\infty_M$ is a sheaf of modules over itself. Also $\mathcal{C}^\infty_M(U)$ is sometimes denoted by $\mathcal{F}_M(U)$.*

Example 6.5 (Sheaf of holomorphic functions) *Sheaf theory really shows its strength in complex analysis. This example is one of the most studied. However, we have not yet defined the notion of a complex manifold and so this example is for those readers with some exposure to complex manifolds. Let M be a complex manifold and let $\mathcal{O}_M(U)$ be the algebra of holomorphic functions defined on U. Here too, $\mathcal{O}_M$ is a sheaf of modules over itself. Where as the sheaf $\mathcal{C}^\infty_M$ always has global sections, the same is not true for $\mathcal{O}_M$. The sheaf theoretic approach to the study of obstructions to the existence of global sections has been very successful.*

Example 6.6 (Sheaf of continuous functions) $\mathcal{C}(U) = C(U)$

Example 6.7 (Sheaf of smooth sections) $\mathcal{S}^E_M(U) = \Gamma(U, E)$ *for a vector bundle $E \to M$. Here $\mathcal{S}^E_M$ is a sheaf of modules over $\mathcal{C}^\infty_M$.*

Remark 6.1 (Notational convention !) *Even though sheaf theory is a deep subject, one of our main purposes for introducing it is for notational (and conceptual) convenience. Each of the above presheaves is a presheaf of sections. As we proceed into the theory of differentiable manifolds we will meet many instances where theorems stated for sections are for both globally defined sections and for sections over opens sets and many times the results will be natural in one way or another with respect to restrictions. For this reason we might simply write $\mathcal{S}^\infty_E, \mathcal{C}^\infty$ etc. whenever we are proving something that would work globally or locally in a natural way. In other words, when the reader sees $\mathcal{S}^E_M$ or $\mathcal{C}^\infty_M$ he or she should think about both $\mathcal{S}^E(U)$ and $\mathcal{S}^E(M)$ or both $\mathcal{C}^\infty(U)$ and $\mathcal{C}^\infty(M)$ and so on.*

We say that $s_1 \in \mathcal{S}^E(U)$ and $s_2 \in \mathcal{S}^E(V)$ determine the same germ of sections at p if there is an open set $W \subset U \cap V$ such that $r^U_W s_1 = r^V_W s_2$. Now on the union

$$\bigcup_{p \in U} \mathcal{S}^E(U)$$

we impose the equivalence relation $s_1 \sim s_2$ iff s_1 and s_2 determine the same germ of sections at p. The set of equivalence classes (called germs of section at p) a ring and is denoted $\mathcal{S}^E_p$. The set $\mathcal{S}^E((U)) = \bigcup_{p \in U} \mathcal{S}^E_p$ is called the sheaf of germs and can be given a topology so that the projection $pr_{\mathcal{S}} : \mathcal{S}^E((U)) \to M$ defined by the requirement that $pr_{\mathcal{S}}([s]) = p$ iff $[s] \in \mathcal{S}^E_p$ is a local homeomorphism.

More generally, let $\mathcal{M}$ be a presheaf of abelian groups on M. For each $p \in M$ we define the direct limit group

$$\mathcal{M}_p = \varinjlim_{p \in U} \mathcal{M}(U)$$

with respect to the restriction maps r^U_V.

Definition 6.18 *$\mathcal{M}_p$ is a set of equivalence classes called germs at p. Here $s_1 \in \mathcal{M}(U)$ and $s_2 \in \mathcal{M}(V)$ determine the same germ of sections at p if there is an open set $W \subset U \cap V$ containing p such that $r^U_W s_1 = r^V_W s_2$. The germ of $s \in \mathcal{M}(U)$ at p is denoted s_p.*

Now we take to union $\widetilde{\mathcal{M}} = \bigcup_{p \in M} \mathcal{M}_p$ and define a surjection $\pi : \widetilde{\mathcal{M}} \to M$ by the requirement that $\pi(s_p) = p$ for $s_p \in \mathcal{M}_p$. The space $\widetilde{\mathcal{M}}$ is called the **sheaf of germs** generated by $\mathcal{M}$. We want to topologize $\widetilde{\mathcal{M}}$ so that π is continuous and a local homeomorphism but first a definition.

Definition 6.19 (étalé space) *A topological space Y together with a continuous surjection $\pi : Y \to M$ which is a local homeomorphism is called an* ***étalé space****. A local section of an étalé space over an open subset $U \subset M$ is a map $s_U : U \to Y$ such that $\pi \circ s_U = \mathrm{id}_U$. The set of all such sections over U is denoted $\Gamma(U, Y)$.*

Definition 6.20 *For each $s \in \mathcal{M}(U)$ we can define a map (of sets) $\widetilde{s} : U \to \widetilde{\mathcal{M}}$ by*

$$\widetilde{s}(x) = s_x$$

and we give $\widetilde{\mathcal{M}}$ the smallest topology such that the images $\widetilde{s}(U)$ for all possible U and s are open subsets of $\widetilde{\mathcal{M}}$.

With the above topology $\widetilde{\mathcal{M}}$ becomes an étalé space and all the sections $\widetilde{s}$ are continuous open maps. Now if we let $\widetilde{\mathcal{M}}(U)$ denote the sections over U for this étalé space, then the assignment $U \to \widetilde{\mathcal{M}}(U)$ is a presheaf which is always a sheaf.

Proposition 6.1 *If $\mathcal{M}$ was a sheaf then $\widetilde{\mathcal{M}}$ is isomorphic as a sheaf to $\mathcal{M}$.*

The notion of a germ not only makes sense of sections of a bundle but we also have the following definition:

Definition 6.21 *Let $F_p(M, N)$ be the set of all smooth maps $f :: M, p \to N$ which are locally define near p. We will say that two such functions f_1 and f_2 have the same **germ** at p if they are equal on some open subset of the intersection of their domains which also contains the point p. In other words they agree in some neighborhood of p. This defines an equivalence relation on $F_p(M, N)$ and the equivalence classes are called germs of maps. The set of all such germs of maps from M to N is denoted by $\mathcal{F}_p(M, N)$.*

It is easy to see that for $[f_1], [f_2] \in \mathcal{F}_p(M, \mathbb{F})$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, the product $[f_1][f_2] = [f_1 f_2]$ is well defined as are the linear operations $a[f_1] + b[f_2] = [af_1 + bf_2]$. One can easily see that $\mathcal{F}_p(M, \mathbb{F})$ is a commutative ring. In fact, it is an algebra over $\mathbb{F}$. Notice also that in any case we have a well defined evaluation $[f](p) = f(p)$.

Exercise 6.8 *Reformulate Theorem 26.12 in terms of germs of maps.*

6.5 Jets and Jet bundles

A map $f :: \mathsf{E}, \mathsf{x} \to \mathsf{F}, \mathsf{y}$ is said to have k-th order contact at x with a map $g :: \mathsf{E}, \mathsf{x} \to \mathsf{F}, \mathsf{y}$ if f and g have the same Taylor polynomial of order k at x. This notion defines an equivalence class on the set of (germs) of functions $\mathsf{E}, \mathsf{x} \to \mathsf{F}, \mathsf{y}$ and the equivalence classes are called k-jets. The equivalence class of a function f is denoted by $j_x^k f$ and the space of k-jets of maps $\mathsf{E}, \mathsf{x} \to \mathsf{F}, \mathsf{y}$ is denoted $J_x^k(\mathsf{E}, \mathsf{F})_y$. We have the following disjoint unions:

$$J_x^k(\mathsf{E}, \mathsf{F}) = \bigcup_{y \in \mathsf{F}} J_x^k(\mathsf{E}, \mathsf{F})_y$$
$$J^k(\mathsf{E}, \mathsf{F})_y = \bigcup_{x \in \mathsf{E}} J_x^k(\mathsf{E}, \mathsf{F})_y$$
$$J^k(\mathsf{E}, \mathsf{F}) = \bigcup_{x \in \mathsf{E}, y \in \mathsf{F}} J_x^k(\mathsf{E}, \mathsf{F})_y$$

Now for a (germ of a) smooth map $f :: M, x \to N, y$ we say that f has k-th order contact at x with a another map $g :: M, x \to N, y$ if $\phi \circ f \circ \psi$ has k-th order contact at $\psi(x)$ with $\phi \circ\, g \circ \psi$ for some (and hence all) charts ϕ and ψ defined in a neighborhood of $x \in M$ and $y \in N$ respectively. We can then define sets $J_x^k(M, N)$, $J^k(M, N)_y$, and $J^k(M, N)$. The space $J^k(M, N)$ is called the space of jets of maps from M to N and turns out to be a smooth vector bundle. In fact, a pair of charts as above determines a chart for $J^k(M, N)$ defined by

$$J^k(\psi, \phi) : j_x^k f \mapsto (\phi(x), \psi(f(x)), D(\phi \circ f \circ \psi), ..., D^k(\phi \circ f \circ \psi))$$

where $D^j(\phi \circ f \circ \psi) \in L^j_{sym}(\mathbb{R}^n, \mathsf{N})$. In finite dimensions, the chart looks like

$$J^k(\psi, \phi) : j^k_x f \mapsto (x^1, ..., x^n, y^1, ..., y^n, \frac{\partial f^i}{\partial x^j}(x), ..., \frac{\partial^\alpha f^i}{\partial x^\alpha}(x))$$

where $|\alpha| = k$. Notice that the chart has domain $U_\psi \times U_\phi$.

Exercise 6.9 *What vector space is naturally the typical fiber of this vector bundle?*

Definition 6.22 *The* k***-jet extension*** *of a map* $f : M \to N$ *is defined by* $j^k f : x \mapsto j^k_x f \in J^k(M, N)$. *There is a strong transversality theorem that uses the concept of a jet space:*

Theorem 6.4 (Strong Transversality) *Let* S *be a closed submanifold of* $J^k(M, N)$. *Then the set of maps such that the* k*-jet extensions* $j^k f$ *are transverse to* S *is an open everywhere dense subset in* $C^\infty(M, N)$.

Chapter 7

Vector Fields and 1-Forms

7.1 Definition of vector fields and 1-forms

Definition 7.1 *A smooth* **vector field** *is a smooth map $X : M \to TM$ such that $X(p) \in T_pM$ for all $p \in M$. We often write $X(p) = X_p$. In other words, a vector field on M is a smooth section of the tangent bundle $\tau_M : TM \to M$.*

The map X being smooth is equivalent to the requirement that $Xf : M \to \mathbb{R}$ given by $p \mapsto X_pf$ is smooth whenever $f : M \to \mathbb{R}$ is smooth.

If $(\mathtt{x}, U)$ is a chart and X a vector field defined on U then the local representation of X is $\mathsf{x} \mapsto (\mathsf{x}, \mathsf{X}(\mathsf{x}))$ where the **principal representative** (or **principal part**) X is given by projecting $T\mathtt{x} \circ X \circ \mathtt{x}^{-1}$ onto the second factor in $T\mathsf{E} = \mathsf{E} \times \mathsf{E}$:

$$\begin{aligned} \mathsf{x} \mapsto \mathtt{x}^{-1}(\mathsf{x}) = p \mapsto X(p) \mapsto T\mathtt{x} \cdot X(p) \\ = (\mathtt{x}(p), \mathsf{X}(\mathtt{x}(p))) = (\mathsf{x}, \mathsf{X}(\mathsf{x})) \mapsto \mathsf{X}(\mathsf{x}) \end{aligned}$$

In finite dimensions one can write $\mathsf{X}(\mathsf{x}) = (v_1(\mathsf{x}), ..., v_n(\mathsf{x}))$.

Notation 7.1 *The set of all smooth vector fields on M is denoted by $\Gamma(M, TM)$ or by the common notation $\mathfrak{X}(M)$. Smooth vector fields may at times be defined only on some open set so we also have the notation $\mathfrak{X}(U) = \mathfrak{X}_M(U)$ for these fields. The map $U \mapsto \mathfrak{X}_M(U)$ is a presheaf (in fact a sheaf).*

A (smooth) section of the cotangent bundle is called a **covector field** or also a smooth **1-form** . The set of all C^r 1-forms is denoted by $\mathfrak{X}^{r*}(M)$ with the smooth 1-forms denoted by $\mathfrak{X}^*(M)$.

$\mathfrak{X}^*(M)$ is a module over the ring of functions $C^\infty(M)$ with a similar statement for the C^r case.

Definition 7.2 *Let $f : M \to \mathbb{R}$ be a smooth function with $r \geq 1$. The map $df : M \to T^*M$ defined by $p \mapsto df(p)$ where $df(p)$ is the differential at p as defined in 3.3. is a 1-form called the differential of f.*

7.2 Pull back and push forward of functions and 1-forms

If $\phi : N \to M$ is a C^r map with $r \geq 1$ and $f : M \to \mathbb{R}$ a C^r function we define the **pullback** of f by ϕ as

$$\phi^* f = f \circ \phi$$

and the **pullback** of a 1-form $\alpha \in \mathfrak{X}^*(M)$ by $\phi^*\alpha = \alpha \circ T\phi$. To get a clear picture of what is going on we could view things at a point and we have $\phi^*\alpha|_p \cdot v = \alpha|_{\phi(p)} \cdot (T_p\phi \cdot v)$.

The pull-back of a function or 1-form is defined whether $\phi : N \to M$ happens to be a diffeomorphism or not. On the other hand, when we define the pull-back of a vector field in a later section we will only be able to do this if the map that we are using is a diffeomorphism. Push-forward is another matter.

Definition 7.3 *If $\phi : N \to M$ is a C^r diffeomorphism with $r \geq 1$. The **push-forward** of a function $\phi_* f$ by $\phi_* f(p) := f(\phi^{-1}(p))$. We can also define the push-forward of a 1-form as $\phi_*\alpha = \alpha \circ T\phi^{-1}$.*

It should be clear that the pull-back is the more natural of the two when it comes to forms and functions but in the case of vector fields this is not true.

Lemma 7.1 *The differential is natural with respect to pullback. In other words, if $\phi : N \to M$ is a C^r map with $r \geq 1$ and $f : M \to \mathbb{R}$ a C^r function with $r \geq 1$ then $d(\phi^* f) = \phi^* df$. Consequently, differential is also natural with respect to restrictions*

Proof. Let v be a curve such that $\dot{c}(0) = v$. Then

$$\begin{aligned} d(\phi^* f)(v) &= \left.\frac{d}{dt}\right|_0 \phi^* f(c(t)) = \left.\frac{d}{dt}\right|_0 f(\phi(c(t))) \\ &= df \left.\frac{d}{dt}\right|_0 \phi(c(t)) = df(T\phi \cdot v) \end{aligned}$$

As for the second statement (besides being obvious from local coordinate expressions) notice that if U is open in M and $\iota : U \hookrightarrow M$ is the inclusion map (identity map id_M) restricted to U) then $f|_U = \iota^* f$ and $df|_U = \iota^* df$ this part follows from the first part. ■

We also have the following familiar looking formula in the finite dimensional case

$$df = \sum \frac{\partial f}{\partial x^i} dx^i$$

which means that at each $p \in U_\alpha$

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

In general, if we have a chart $U, \mathtt{x}$ then we may write

$$df = \frac{\partial f}{\partial \mathtt{x}} d\mathtt{x}$$

We have seen this before. All that has happened is that p is allowed to vary so we have a field.

For any open set $U \subset M$, the set of smooth functions defined $C^\infty(U)$ on U is an algebra under the obvious linear structure $(af + bg)(p) := af(p) + bg(p)$ and obvious multiplication; $(fg)(p) := f(p)g(p)$. When we think of $C^\infty(U)$ in this way we sometimes denote it by $\mathcal{C}^\infty(U)$.The assignment $U \mapsto \mathfrak{X}_M(U)$ is a presheaf of modules over $\mathcal{C}^\infty$.

7.3 Frame Fields

If $U, \mathtt{x}$ is a chart on a smooth n-manifold then writing $\mathtt{x} = (x^1, ..., x^n)$ we have vector fields defined on U by

$$\frac{\partial}{\partial x^i} : p \mapsto \left.\frac{\partial}{\partial x^i}\right|_p$$

such that the together the $\frac{\partial}{\partial x^i}$ form a basis at each tangent space at point in U. We call the set of fields $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$ a holonomic frame field over U. If X is a vector field defined on some set including this local chart domain U then for some smooth functions X^i defined on U we have

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

or in other words

$$X|_U = \sum X^i \frac{\partial}{\partial x^i}.$$

Notice also that $dx^i : p \mapsto \left.dx^i\right|_p$ defines a field of co-vectors such that $\left.dx^1\right|_p, ..., \left.dx^n\right|_p$ forms a basis of T_p^*M for each $p \in U$. The fields form what is called a **holonomic**[1] **co-frame** over U. In fact, the functions X^i are given by $dx^i(X) : p \mapsto \left.dx^i\right|_p(X_p)$.

Notation 7.2 *We will not usually bother to distinguish X from its restrictions and so we just write $X = \sum X^i \frac{\partial}{\partial x^i}$ or using the Einstein summation convention $X = X^i \frac{\partial}{\partial x^i}$.*

It is important to realize that it is possible to have family of fields that are linearly independent at each point in their mutual domain and yet are not necessarily of the form $\frac{\partial}{\partial x^i}$ for any coordinate chart. This leads to the following

[1] The word holonomic comes from mechanics and just means that the frame field derives from a chart. A related fact is that $[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = 0$.

Definition 7.4 *Let $F_1, F_2, ..., F_n$ be smooth vector fields defined on some open subset U of a smooth n-manifold M. If $F_1(p), F_2(p), ..., F_n(p)$ from a basis for T_pM for each $p \in U$ then we say that $F_1, F_2, ..., F_n$ is a (non-holonomic)* ***frame field*** *over U.*

If $F_1, F_2, ..., F_n$ is frame field over $U \subset M$ and X is a vector field defined on U then we may write

$$X = \sum X^i F_i \text{ on } U$$

for some functions X^i defined on U. Taking the dual basis in T_p^*M for each $p \in U$ we get a (non-holonomic) co-frame field $F^1, ..., F^n$ and then $X^i = F^i(X)$.

Definition 7.5 *A derivation on $\mathcal{C}^\infty(U)$ is a linear map $\mathcal{D} : \mathcal{C}^\infty(U) \to \mathcal{C}^\infty(U)$ such that*

$$\mathcal{D}(fg) = \mathcal{D}(f)g + f\mathcal{D}(g).$$

A C^∞ vector field on U may be considered as a derivation on $\mathfrak{X}(U)$ where we view $\mathfrak{X}(U)$ as a module over the ring of smooth functions $\mathcal{C}^\infty(U)$.

Definition 7.6 *To a vector field X on U we associate the map $\mathcal{L}_X : \mathfrak{X}_M(U) \to \mathfrak{X}_M(U)$ defined by*

$$(\mathcal{L}_X\ f)(p) := X_p \cdot f$$

and called the ***Lie derivative*** *on functions.*

It is easy to see, based on the Leibnitz rule established for vectors X_p in individual tangent spaces, that $\mathcal{L}_X$ is a derivation on $\mathcal{C}^\infty(U)$. We also define the symbolism "Xf", where $X \in \mathfrak{X}(U)$, to be an abbreviation for the function $\mathcal{L}_X\ f$. We often leave out parentheses and just write $Xf(p)$ instead of the more careful $(Xf)(p)$ and so, for example the derivation law (Leibnitz rule) reads

$$X(fg) = fXg + gXf.$$

7.4 Lie Bracket

Lemma 7.2 *Let $U \subset M$ be an open set. If $\mathcal{L}_X\ f(p) = 0$ for all $f \in C^\infty(U)$ and all $p \in U$ then $X|_U = 0$.*

Proof. Working locally in a chart $(\mathtt{x}, U)$, let X be the principal representative of X (defined in section 7.1). Suppose that $\ell : \mathsf{E} \to \mathbb{R}$ is a continuous linear map such that $\ell(\mathsf{X}(p)) \neq 0$ (Hahn-Banach). Then $D_p\ell = \ell$ and $d(\mathtt{x}_\alpha^*\ell_p)(X(p)) = D_p\ell \cdot \mathsf{X} \neq 0$ ■

Theorem 7.1 *For every pair of vector fields $X, Y \in \mathfrak{X}(M)$ there is a unique vector field $[X, Y]$ which for any open set $U \subset M$ and $f \in \mathcal{C}^\infty(U)$ we have $[X, Y]_{(p)}f = X_p(Yf) - Y_p(Xf)$ and such that in a local chart $U, \mathtt{x}$ the vector field $[X, Y]$ has principal part given by*

$$D\mathsf{Y} \cdot \mathsf{X} - D\mathsf{X} \cdot \mathsf{Y}$$

or more fully if $f \in \mathcal{C}^{\infty}(U)$ then letting $\mathsf{f} := f \circ \mathtt{x}^{-1}$ be the local representative we have

$$([\mathsf{X}, \mathsf{Y}]\mathsf{f})(\mathsf{x}) = D\mathsf{f} \cdot (D\mathsf{Y}(\mathsf{x}) \cdot \mathsf{X}(\mathsf{x}) - D\mathsf{X}(\mathsf{x}) \cdot \mathsf{Y}(\mathsf{x}))$$

where $\mathsf{x} \sim p$ and $[\mathsf{X}, \mathsf{Y}]\mathsf{f}$ means the local representative of $[X, Y]f$.

Proof. We sketch the main points and let the reader fill in the details. One can prove that the formula $f \mapsto Df(D\mathsf{Y} \cdot \mathsf{X} - D\mathsf{X} \cdot \mathsf{Y})$ defines a derivation locally. Also, $D\mathsf{Y} \cdot \mathsf{X} - D\mathsf{X} \cdot \mathsf{Y}$ transforms correctly and so gives a vector at each point. Thus by pulling back via charts we get a vector field on each chart domain. But these agree on overlaps because they are all coming from the local representations of a single derivation $f \mapsto X_p(Yf) - Y_p(Xf)$. In the finite dimensional case we can see the result from another point of view. Namely, every derivation of $C^{\infty}(M)$ is $\mathcal{L}_X$ for some X so we can define $[X, Y]$ as the unique vector field such that $\mathcal{L}_{[X,Y]} = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$. ∎

We ought to see what the local formula for the Lie derivative looks like in the finite dimensional case where we may employ classical notation. Suppose we have $X = \sum_{i=1}^{m} X^i \frac{\partial}{\partial x^i}$ and $Y = \sum_{i=1}^{m} Y^i \frac{\partial}{\partial x^i}$. Then $[X, Y] = \sum_i \left(\sum_j \frac{\partial Y^i}{\partial x^j} X^j - \frac{\partial X^i}{\partial x^j} Y^j \right) \frac{\partial}{\partial x^i}$

Exercise 7.1 *Check this.*

Definition 7.7 *The vector field $[X, Y]$ from the previous theorem is called the **Lie bracket** of X and Y.*

The following properties of the Lie Bracket are checked by direct calculation. For any $X, Y, Z \in \mathfrak{X}(M)$,

1. $[X, Y] = -[Y, X]$

2. $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$

Definition 7.8 (Lie Algebra) *A vector space $\mathfrak{a}$ is called a **Lie algebra** if it is equipped with a bilinear map $\mathfrak{a} \times \mathfrak{a} \to \mathfrak{a}$ (a multiplication) denoted $v, w \mapsto [v, w]$ such that*

$$[v, w] = -[w, v]$$

and such that we have a Jacobi identity

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$$

for all $x, y, z \in \mathfrak{a}$.

We have seen that $\mathfrak{X}(M)$ (or $\mathfrak{X}(U)$) is a Lie algebra with the Bracket defined in definition 7.7.

7.5 Localization

It is a fact that if M is finite dimensional then every derivation of $\mathcal{C}^\infty(M)$ is given by some vector field and in fact the association $X \mapsto \mathcal{L}_X$ is a bijection from $\mathfrak{X}(M)$ onto the set of all derivations of the algebra of $C^\infty = \mathfrak{F}(M)$ functions $\mathcal{C}^\infty(M)$. Actually, this is true not just for finite dimensional manifolds but also for manifolds modelled on a special class of Banach spaces which include Hilbert spaces. In order to get an isomorphism between the set of all derivations of the algebra $\mathcal{C}^\infty(M) = \mathfrak{F}(M)$ we need to be able to construct appropriate bump functions or more specifically cut-off functions.

Definition 7.9 *Let K be a closed subset of M contained in an open subset $U \subset M$. A* ***cut-off function*** *for the nested pair $K \subset U$ is a C^∞ function $\beta : M \to \mathbb{R}$ such that $\beta|_K \equiv 1$ and $\beta|_{M\backslash U} \equiv 0$.*

Definition 7.10 *A manifold M is said to* ***admit cut-off functions*** *if given any point $p \in M$ and any open neighborhood U of p, there is another neighborhood V of p such that $\overline{V} \subset U$ and a cut-off function $\beta_{\overline{V},U}$ for the nested pair $\overline{V} \subset U$.*

Lemma 7.3 *All finite dimensional smooth manifolds admit cut-off functions.*

Proof. Exercise ■

Definition 7.11 *Let E be a Banach space and suppose that the norm on E is smooth (resp.C^r). The we say that E is a* ***smooth (resp. C^r) Banach space****.*

Lemma 7.4 *If E is a smooth (resp. C^r) Banach space and $B_r \subset B_R$ nested open balls then there is a smooth (resp. C^r) function β defined on all of E which is identically equal to 1 on the closure $\overline{B_r}$ and zero outside of B_R.*

Proof. We assume with out loss of generality that the balls are centered at the origin $0 \in \mathsf{E}$. Let

$$\phi_1(s) = \frac{\int_{-\infty}^{s} g(t)dt}{\int_{-\infty}^{\infty} g(t)dt}$$

where

$$g(t) = \begin{cases} \exp(-1/(1-|t|^2) & \text{if} \quad |t| < 1 \\ \\ 0 & \text{otherwise} \end{cases}.$$

This is a smooth function and is zero if $s < -1$ and 1 if $s > 1$ (verify). Now let $\beta(\mathsf{x}) = g(2 - |\mathsf{x}|)$. Check that this does the job using the fact that $\mathsf{x} \mapsto |\mathsf{x}|$ is assumed to be smooth (resp. C^r). ■

Corollary 7.1 *If a manifold M is modelled on a smooth (resp. C^r) Banach space E, (in particular, if M is an n-manifold), then for every $\alpha_p \in T^*M$, there is a (global) smooth (resp. C^r) function f such that $Df|_p = \alpha_p$.*

Proof. Let $\mathsf{x}_0 = \psi(p) \in \mathbb{R}^n$ for some chart ψ, U. Then the local representative $\bar{\alpha}_{x_0} = (\psi^{-1})^* \alpha_p$ can be considered a linear function on $\mathbb{R}^n$ since we have the canonical identification $\mathbb{R}^n \cong \{\mathsf{x}_0\} \times \mathbb{R}^n = \mathsf{T}_{\mathsf{x}_0} \mathbb{R}^n$. Thus we can define

$$\varphi(\mathsf{x}) = \begin{cases} \beta(\mathsf{x})\bar{\alpha}_{\mathsf{x}_0}(\mathsf{x}) & \text{for} \quad \mathsf{x} \in B_R(\mathsf{x}_0) \\ 0 & \text{otherwise} \end{cases}$$

and now making sure that R is small enough that $B_R(\mathsf{x}_0) \subset \psi(U)$ we can transfer this function back to M via ψ^{-1} and extend to zero outside of U get f. Now the differential of φ at x_0 is $\bar{\alpha}_{x_0}$ and so we have for $v \in T_pM$

$$\begin{aligned} df(p) \cdot v &= d(\psi^* \varphi)(p) \cdot v \\ &= (\psi^* d\varphi)(p) v \\ & d\varphi(T_p \psi \cdot v) \\ &= \bar{\alpha}_{x_0}(T_p \psi \cdot v) = (\psi^{-1})^* \alpha_p (T_p \psi \cdot v) \\ &= \alpha_p (T \psi^{-1} T_p \psi \cdot v) = \alpha_p(v) \end{aligned}$$

so $df(p) = \alpha_p$ ■

Lemma 7.5 *The map from $\mathfrak{X}(M)$ to the vector space of derivations $\mathsf{Der}(M)$ given by $X \mapsto \mathcal{L}_X$ is a module monomorphism if M is modelled on a C^∞ Banach space.*

Proof. The fact that the map is a module map is straightforward. We just need to get the injectivity. For that, suppose $\mathcal{L}_X f = 0$ for all $f \in \mathcal{C}^\infty(M)$. Then $Df|_p X_p = 0$ for all $p \in M$. Thus by corollary 7.1 $\alpha_p(X_p) = 0$ for all $\alpha_p \in T_p^* M$. By the Hahn Banach theorem this means that $X_p = 0$. Since p was arbitrary we concluded that $X = 0$. ■

Theorem 7.2 *Let $L : \mathfrak{X}(M) \to \mathcal{C}^\infty(M)$ be a $\mathcal{C}^\infty(M)-$linear function on vector fields. If M admits cut off functions then $L(X)(p)$ depends only on the germ of X at p.*

If M is finite dimensional then $L(X)(p)$ depends only on the value of X at p.

Proof. Suppose $X = 0$ in a neighborhood U and let $p \in U$ be arbitrary. Let O be a smaller open set containing with closure inside U. Then letting β be a function that is 1 on a neighborhood of p contained in O and identically zero outside of O then $(1 - \beta)X = X$. Thus we have

$$\begin{aligned} L(X)(p) &= L((1 - \beta)X)(p) \\ &= (1 - \beta(p))L(X)(p) = 0 \times L(X)(p) \\ &= 0. \end{aligned}$$

Applying this to $X - Y$ we see that if two fields X and Y agree in an open set then $L(X) = L(Y)$ on the same open set. The result follows from this.

Now suppose that M is finite dimensional and suppose that $X(p) = 0$. Write $X = X^i \frac{\partial}{\partial x^i}$ in some chart domain containing p with smooth function X^i satisfying $X^i(p) = 0$. Letting β be as above we have

$$\beta^2 L(X) = \beta X^i L(\beta \frac{\partial}{\partial x^i})$$

which evaluated at p gives

$$L(X)(p) = 0$$

since $\beta(p) = 1$. Applying this to $X - Y$ we see that if two fields X and Y agree at p then $L(X)(p) = L(Y)(p)$. ■

Corollary 7.2 *If M is finite dimensional and $L : \mathfrak{X}(M) \to \mathcal{C}^\infty(M)$ is a $\mathcal{C}^\infty(M)-$linear function on vector fields then there exists an element $\alpha \in \mathfrak{X}^*(M)$ such that $\alpha(X) = L(X)$ for all $X \in \mathfrak{X}(M)$.*

Remark 7.1 (Convention) *Even though a great deal of what we do does not depend on the existence of cut off functions there are several places where we would like to be able to localize.*

7.6 Action by pullback and push-forward

Given a diffeomorphism $\phi : M \to N$ we define the **pull back** $\phi^* Y \in \mathfrak{X}(M)$ for $Y \in \mathfrak{X}(N)$ and the **push-forward** $\phi_* X \in \mathfrak{X}(N)$ of $X \in \mathfrak{X}(M)$ via ϕ by defining

$$\phi^* Y = T\phi^{-1} \circ Y \circ \phi \text{ and}$$
$$\phi_* X = T\phi \circ X \circ \phi^{-1}.$$

In other words, $(\phi^* Y)(p) = T\phi^{-1} \cdot Y_{\phi p}$ and $(\phi_* X)(p) = T\phi \cdot X_{\phi^{-1}(p)}$. Notice that $\phi^* Y$ and $\phi_* X$ are both smooth vector fields. Let $\phi : M \to N$ be a smooth map of manifolds. The following commutative diagrams summarize some of the concepts:

$$\begin{array}{ccc} T_pM & \stackrel{T_p\phi}{\to} & T_{\phi p}N \\ \downarrow & & \downarrow \\ M, p & \stackrel{\phi}{\to} & N, \phi(p) \end{array}$$

$$\begin{array}{ccc} TM & \stackrel{T\phi}{\to} & TN \\ \downarrow & & \downarrow \\ M & \stackrel{\phi}{\to} & N \end{array}$$

and if ϕ is a diffeomorphism then

$$\begin{array}{ccc} \mathfrak{X}(M) & \stackrel{\phi_*}{\to} & \mathfrak{X}(N) \\ \downarrow & & \downarrow \\ M & \stackrel{\phi}{\to} & N \end{array}$$

and also

$$\begin{array}{ccc} \mathfrak{X}(M) & \overset{\phi^*}{\leftarrow} & \mathfrak{X}(N) \\ \downarrow & & \downarrow \\ M & \overset{\phi}{\rightarrow} & N \end{array}.$$

Notice the arrow reversal. The association $\phi \mapsto \phi^*$ is said to be contravariant as opposed to $\phi \mapsto \phi_*$ which is covariant. In fact we have the following facts concerning a composition of smooth diffeomorphisms $M \overset{\phi}{\rightarrow} N \overset{f}{\rightarrow} P$:

$$\begin{array}{ll} (\phi \circ f)_* = \phi_* \circ f_* & \text{covariant} \\ (\phi \circ f)^* = f^* \circ \phi^* & \text{contravariant} \end{array}$$

If $M = N$, this gives a right and left pair of actions[2] of the diffeomorphism group $\mathrm{Diff}(M)$ on the space of vector fields: $\mathfrak{X}(M) = \Gamma(M, TM)$.

$$\begin{aligned} \mathrm{Diff}(M) \times \mathfrak{X}(M) &\rightarrow \mathfrak{X}(M) \\ (\phi_*, X) &\mapsto \phi_* X \end{aligned}$$

and

$$\begin{aligned} \mathfrak{X}(M) \times \mathrm{Diff}(M) &\rightarrow \mathfrak{X}(M) \\ (X, \phi) &\mapsto \phi^* X \end{aligned}$$

Now even if $\phi : M \rightarrow N$ is not a diffeomorphism it still may be that there is a vector field $Y \in \mathfrak{X}(N)$ such that

$$T\phi \circ X = Y \circ \phi.$$

Or on other words, $T\phi \cdot X_p = Y_{\phi(p)}$ for all p in M. In this case we say that Y is **ϕ-related** to X and write $X \sim_\phi Y$.

Theorem 7.3 *The Lie derivative on functions is natural with respect to pull-back and push-forward by diffeomorphisms. In other words, if $\phi : M \rightarrow N$ is a diffeomorphism and $f \in C^\infty(M)$, $g \in C^\infty(N)$, $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$ then*

$$\mathcal{L}_{\phi^* Y} \phi^* g = \phi^* \mathcal{L}_Y g$$

and

$$\mathcal{L}_{\phi_* X} \phi_* f = \phi_* \mathcal{L}_X f$$

Proof.

$$\begin{aligned} (\mathcal{L}_{\phi^* Y} \phi^* g)(p) &= d(\phi^* g)(\phi^* Y)(p) \\ = (\phi^* dg)(T\phi^{-1} Y(\phi p)) &= dg(T\phi T\phi^{-1} Y(\phi p)) \\ &= dg(Y(\phi p)) = (\phi^* \mathcal{L}_Y g)(p) \end{aligned}$$

The second statement follows from the first since $(\phi^{-1})^* = \phi_*$. ■

In case the map $\phi : M \rightarrow N$ is not a diffeomorphism we still have a result when two vector fields are **ϕ-related**.

[2]For the definition of group action see the sections on Lie groups below.

Theorem 7.4 *Let $\phi : M \to N$ be a smooth map and suppose that $X \sim_\phi Y$. Then we have for any $g \in C^\infty(N)$ $\mathcal{L}_X \phi^* g = \phi^* \mathcal{L}_Y g$.*

The proof is similar to the previous theorem and is left to the reader.

7.7 Flows and Vector Fields

All flows of vector fields near points where the field doesn't vanish look the same.

A family of diffeomorphisms $\Phi_t : M \to M$ is called a **(global) flow** if $t \mapsto \Phi_t$ is a group homomorphism from the additive group $\mathbb{R}$ to the diffeomorphism group of M and such that $\Phi_t(x) = \Phi(t,x)$ gives a smooth map $\mathbb{R}\times M \to M$. A **local flow** is defined similarly except that $\Phi(t,x)$ may not be defined on all of $\mathbb{R}\times M$ but rather on some open set in $\mathbb{R}\times M$ and so we explicitly require that

1. $\Phi_t \circ \Phi_s = \Phi_{t+s}$ and
2. $\Phi_t^{-1} = \Phi_{-t}$

for all t and s such that both sides of these equations are defined.

Using a smooth local flow we can define a vector field X^Φ by

$$X^\Phi(p) = \left.\frac{d}{dt}\right|_0 \Phi(t,p) \in T_pM.$$

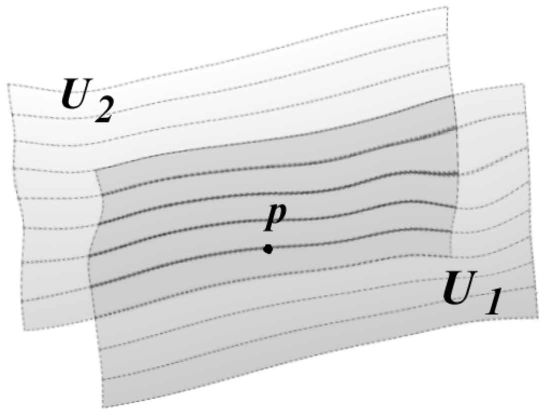
If one computes the velocity vector $\dot{c}(0)$ of the curve $c : t \mapsto \Phi(t,x)$ one gets $X^\Phi(x)$. On the other hand, if we are given a smooth vector field X in open set $U \subset M$ then we say that $c : (a,b) \to M$ is an **integral curve** for X if $\dot{c}(t) = X(c(t))$ for $t \in (a,b)$.

Our study begins with a quick recounting of a basic existence and uniqueness theorem for differential equations stated here in the setting of Banach spaces. The proof may be found in Appendix B.

Theorem 7.5 *Let E be a Banach space and let $X : U \subset E \to E$ be a smooth map. Given any $x_0 \in U$ there is a smooth curve $c : (-\epsilon,\epsilon) \to U$ with $c(0) = x_0$ such that $c'(t) = X(c(t))$ for all $t \in (-\epsilon,\epsilon)$. If $c_1 : (-\epsilon_1,\epsilon_1) \to U$ is another such curve with $c_1(0) = x_0$ and $c_1'(t) = X(c(t))$ for all $t \in (-\epsilon_1,\epsilon_1)$ then $c = c_1$ on the intersection $(-\epsilon_1,\epsilon_1) \cap (-\epsilon,\epsilon)$. Furthermore, there is an open set V with $x_0 \in V \subset U$ and a smooth map $\Phi : V \times (-a,a) \to U$ such that $t \mapsto c_x(t) := \Phi(x,t)$ is a curve satisfying $c'(t) = X(c(t))$ for all $t \in (-a,a)$.*

We will now use this theorem to obtain similar but more global results on smooth manifolds. First of all we can get a more global version of uniqueness:

Lemma 7.6 *If $c_1 : (-\epsilon_1,\epsilon_1) \to M$ and $c_2 : (-\epsilon_2,\epsilon_2) \to M$ are integral curves of a vector field X with $c_1(0) = c_2(0)$ then $c_1 = c_2$ on the intersection of their domains.*



Proof. Let $K = \{t \in (-\epsilon_1, \epsilon_1) \cap (-\epsilon_2, \epsilon_2) : c_1(t) = c_2(t)\}$. The set K is closed since M is Hausdorff. If follows from the local theorem 7.5 that K contains a (small) open interval $(-\epsilon, \epsilon)$. Now let t_0 be any point in K and consider the translated curves $c_1^{t_0}(t) = c_1(t_0 + t)$ and $c_2^{t_0}(t) = c_2(t_0 + t)$. These are also integral curves of X and agree at $t = 0$ and by 7.5 again we see that $c_1^{t_0} = c_2^{t_0}$ on some open neighborhood of 0. But this means that c_1 and c_2 agree in this neighborhood so in fact this neighborhood is contained in K implying K is also open since t_0 was an arbitrary point in K. Thus, since $I = (-\epsilon_1, \epsilon_1) \cap (-\epsilon_2, \epsilon_2)$ is connected, it must be that $I = K$ and so c_1 and c_2 agree on $I = (-\epsilon_1, \epsilon_1) \cap (-\epsilon_2, \epsilon_2)$. ∎

Flow box and Straightening

Let X be a C^r vector field on M with $r \geq 1$. A **flow box** for X at a point $p_0 \in M$ is a triple (U, a, Fl^X) where

1. U is an open set in M containing p.
2. $\mathrm{Fl}^X : U \times (-a, a) \to M$ is a C^r map and $0 < a \leq \infty$.
3. For each $p \in M$ the curve $t \mapsto c_p(t) = \mathrm{Fl}^X(p, t)$ is an integral curve of X with $c_p(0) = p$.
4. The map $\mathrm{Fl}_t^X : U \to M$ given by $\mathrm{Fl}_t^X(p) = \mathrm{Fl}^X(p, t)$ is a diffeomorphism onto its image for all $t \in (-a, a)$.

Now before we prove that flow boxes actually exist, we make the following observation: If we have a triple that satisfies 1-3 above then both $c_1 : t \mapsto \mathrm{Fl}_{t+s}^X(p)$ and $c_2 : t \mapsto \mathrm{Fl}_t^X(\mathrm{Fl}_s^X(p))$ are integral curves of X with $c_1(0) = c_2(0) = \mathrm{Fl}_s^X(p)$ so by uniqueness (Lemma 7.6) we conclude that $\mathrm{Fl}_t^X(\mathrm{Fl}_s^X(p)) = \mathrm{Fl}_{t+s}^X(p)$ as long as they are defined. This also shows that

$$\mathrm{Fl}_s^X \circ \mathrm{Fl}_t^X = \mathrm{Fl}_{t+s}^X = \mathrm{Fl}_t^X \circ \mathrm{Fl}_s^X$$

whenever defined. This is the local group property, so called because if Fl_t^X were defined for all $t \in \mathbb{R}$ (and X a global vector field) then $t \mapsto \mathrm{Fl}_t^X$ would be a group homomorphism from $\mathbb{R}$ into $\mathrm{Diff}(M)$. Whenever this happens we say that X is a **complete vector field**. The group property also implies that $\mathrm{Fl}_t^X \circ \mathrm{Fl}_{-t}^X = \mathrm{id}$ and so Fl_t^X must at least be a locally defined diffeomorphism with inverse Fl_{-t}^X.

Theorem 7.6 (Flow Box) *Let X be a C^r vector field on M with $r \geq 1$. Then for every point $p_0 \in M$ there exists a flow box for X at p_0. If $(U_1, a_1, \mathrm{Fl}_1^X)$ and $(U_2, a_2, \mathrm{Fl}_2^X)$ are two flow boxes for X at p_0,then $\mathrm{Fl}_1^X = \mathrm{Fl}_2^X$ on $(-a_1, a_1) \cap (-a_2, a_2) \times U_1 \cap U_2$.*

Proof. First of all notice that the U in the triple (U, a, Fl^X) does not have to be contained in a chart or even homeomorphic to an open set in the model space. However, to prove that there are flow boxes at any point we can work in the domain of a chart U_α, ψ_α and so we might as well assume that the vector field is defined on an open set in the model space as in 7.5. Of course, we may have to choose a to be smaller so that the flow stays within the range of the chart map ψ_α. Now a vector field in this setting can be taken to be a map $U \to E$ so the theorem 7.5 provides us with the flow box data (V, a, Φ) where we have taken $a > 0$ small enough that $V_t = \Phi(t, V) \subset U_\alpha$ for all $t \in (-a, a)$. Now the flow box is transferred back to the manifold via ψ_α

$$U = \psi_\alpha^{-1}(V)$$
$$\mathrm{Fl}^X(t, p) = \Phi(t, \psi_\alpha(p)).$$

Now if we have two such flow boxes $(U_1, a_1, \mathrm{Fl}_1^X)$ and $(U_2, a_2, \mathrm{Fl}_2^X)$ then by lemma 7.6 we have for any $x \in U_1 \cap U_2$ we must have $\mathrm{Fl}_1^X(t, x) = \mathrm{Fl}_2^X(t, x)$ for all $t \in (-a_1, a_1) \cap (-a_2, a_2)$.

Finally, since both $\mathrm{Fl}_t^X = \mathrm{Fl}^X(t, .)$ and $\mathrm{Fl}_{-t}^X = \mathrm{Fl}^X(-t, .)$ are both smooth and inverse of each other we see that Fl_t^X is a diffeomorphism onto its image $U_t = \psi_\alpha^{-1}(V_t)$. ■

Definition 7.12 *Let X be a C^r vector field on M with $r \geq 1$. For any given $p \in M$ let $(T_{p,X}^-, T_{p,X}^+) \subset \mathbb{R}$ be the largest interval such that there is an integral curve $c : (T_{p,X}^-, T_{p,X}^+) \to M$ of X with $c(0) = p$. The **maximal flow** Fl^X is defined on the open set (called the maximal flow domain)*

$$\mathcal{FD}_X = \bigcup (T_{p,X}^-, T_{p,X}^+) \times \{p\}.$$

Remark 7.2 *Up until now we have used the notation Fl^X ambiguously to refer to any (local or global) flow of X and now we have used the same notation for the* unique *maximal flow defined on $\mathcal{FD}_X$. We could have introduced notation such as $\mathrm{Fl}_{\max}^X$ but prefer not to clutter up the notation to that extent unless necessary. We hope that the reader will be able to tell from context what we are referring to when we write Fl^X.*

Exercise 7.2 *Show that what we have proved so far implies that the maximal interval $(T^-_{p,X}, T^+_{p,X})$ exists for all $p \in M$ and prove that $\mathcal{FD}_X$ is an open subset of $\mathbb{R} \times M$.*

Definition 7.13 *We say that X is a* ***complete vector field*** *iff $\mathcal{FD}_X = \mathbb{R} \times M$.*

Notice that we always have $\mathrm{Fl}^X_s \circ \mathrm{Fl}^X_t = \mathrm{Fl}^X_{t+s} = \mathrm{Fl}^X_t \circ \mathrm{Fl}^X_s$ whenever s, t are such that all these maps are defined but if X is a complete vector field then this equation is true for all s,$t \in \mathbb{R}$.

Definition 7.14 *The support of a vector field X is the closure of the set $\{p : X(p) \neq 0\}$ and is denoted supp(X).*

Lemma 7.7 *Every vector field that has compact support is a complete vector field. In particular if M is compact then every vector field is complete.*

Proof. Let c^X_p be the maximal integral curve through p and $(T^-_{p,X}, T^+_{p,X})$ its domain. It is clear that for any $t \in (T^-_{p,X}, T^+_{p,X})$ the image point $c^X_p(t)$ must always lie in the support of X. But we show that if $T^+_{p,X} < \infty$ then given any compact set $K \subset M$, for example the support of X, there is an $\epsilon > 0$ such that for all $t \in (T^+_{p,X} - \epsilon, T^+_{p,X})$ the image $c^X_p(t)$ is outside K. If not then we may take a sequence t_i converging to $T^+_{p,X}$ such that $c^X_p(t_i) \in K$. But then going to a subsequence if necessary we have $x_i := c^X_p(t_i) \to x \in K$. Now there must be a flow box (U, a, x) so for large enough k, we have that t_k is within a of $T^+_{p,X}$ and $x_i = c^X_p(t_i)$ is inside U. We then a guaranteed to have an integral curve $c^X_{x_i}(t)$ of X that continues beyond $T^+_{p,X}$ and thus can be used to extend c^X_p a contradiction of the maximality of $T^+_{p,X}$. Hence we must have $T^+_{p,X} = \infty$. A similar argument give the result that $T^-_{p,X} = -\infty$. ∎

Exercise 7.3 *Let $a > 0$ be any positive real number. Show that if for a given vector field X the flow Fl^X is defined on $(-a, a) \times M$ then in fact the (maximal) flow is defined on $\mathbb{R} \times M$ and so X is a complete vector field.*

7.8 Lie Derivative

Let X be a vector field on M and let $\mathrm{Fl}^X(p, t) = \mathrm{Fl}^X_p(t) = \mathrm{Fl}^X_t(p)$ be the flow so that

$$\left.\frac{d}{dt}\right|_0 \mathrm{Fl}^X_p(t) = T_0\mathrm{Fl}^X_p \left.\frac{\partial}{\partial t}\right|_0 = X_p$$

Recall our definition of the Lie derivative of a function (7.6). The following is an alternative definition.

Definition 7.15 *For a smooth function $f : M \to \mathbb{R}$ and a smooth vector field $X \in \mathfrak{X}(M)$ define the Lie derivative $\mathcal{L}_X$ of f with respect to X by*

$$\begin{aligned}\mathcal{L}_X f(p) &= \left.\frac{d}{dt}\right|_0 f \circ \mathrm{Fl}^X(p, t) \\ &= X_p f\end{aligned}$$

Exercise 7.4 *Show that this definition is compatible with definition 7.6.*

Discussion: Notice that if X is a complete vector field then Fl_t^X is a diffeomorphism $M \to M$ and we may define $(\mathrm{Fl}_t^{X*}Y)(p) = (T_p\mathrm{Fl}_t^X)^{-1}Y(\mathrm{Fl}_t(p))$ or

$$\mathrm{Fl}_t^{X*}Y = (T\mathrm{Fl}_t^X)^{-1} \circ Y \circ \mathrm{Fl}_t. \tag{7.1}$$

On the other hand, if X is not complete then we **cannot** say that Fl_t^X is a diffeomorphism of M since for any specific t there might be points for which Fl_t^X is not even defined! To see this one just needs to realize that the notation implies that we have a map $t \mapsto \mathrm{Fl}_t^X \in \mathrm{Diff}(M)$. But what is the domain? Suppose that ϵ is in the domain. Then it follows that for all $0 \leq t \leq \epsilon$ the map Fl_t^X is defined on all of M and $\mathrm{Fl}_t^X(p)$ exists for $0 \leq t \leq \epsilon$ independent of p. But now a standard argument shows that $t \mapsto \mathrm{Fl}_t^X(p)$ is defined for all t which means that X is a complete vector field. If X is not complete we really have no business writing down the above equation without some qualification. Despite this it has become common to write this expression anyway especially when we are taking a derivative with respect to t. Whether or not this is just a mistake or liberal use of notation is not clear. Here is what we can say. Given any relatively compact open set $U \subset M$ the map Fl_t^X will be defined at least for all $t \in (-\varepsilon, \varepsilon)$ for some ε depending only on X and the choice of U. Because of this, the expression $\mathrm{Fl}_t^{X*}Y = (T_p\mathrm{Fl}_t^X)^{-1} \circ Y \circ \mathrm{Fl}_t$ is a well defined map on U for all $t \in (-\varepsilon, \varepsilon)$. Now if our manifold has a cover by relatively compact open sets $M = \bigcup U_i$ then we can make sense of $\mathrm{Fl}_t^{X*}Y$ on as large of a relatively compact set we like as long as t is small enough. Furthermore, if $\mathrm{Fl}_t^{X*}Y\big|_{U_i}$ and $\mathrm{Fl}_t^{X*}Y\big|_{U_j}$ are both defined for the same t then they both restrict to $\mathrm{Fl}_t^{X*}Y\big|_{U_i \cap U_j}$. So $\mathrm{Fl}_t^{X*}Y$ makes sense point by point for small enough t. It is just that "small enough" may never be uniformly small enough over M so literally speaking Fl_t^X is just not a map on M since t has to be *some* number and for a vector field that is not complete, no t will be small enough (see the exercise 7.3 above). At any rate $t \mapsto (\mathrm{Fl}_t^{X*}Y)(p)$ has a well defined germ at $t = 0$. With this in mind the following definition makes sense even for vector fields that are not complete as long as we take a loose interpretation of $\mathrm{Fl}_t^{X*}Y$.

Definition 7.16 *Let X and Y be smooth vector fields on M. Then the **Lie derivative** L_XY defined by $L_XY := \frac{d}{dt}\big|_0 \mathrm{Fl}_t^{\mathbf{X}*}Y$.*

The following slightly different but equivalent definition is perfectly precise in any case.

Definition 7.17 (version2) *Let X and Y be smooth vector fields on M or some open subset of M. Then the **Lie derivative** L_XY defined by*

$$L_XY(p) = \lim_{t \to 0} \frac{T\mathrm{Fl}_{-t}^X \cdot Y_{\mathrm{Fl}_t^X(p)} - Y}{t} \tag{7.2}$$

for any p in the domain of X.

Figure 7.1: Lie derivative defined by flow.

In the sequel we will sometimes need to keep in mind our comments above in order to make sense of certain expressions.

If $X, Y \in \mathfrak{X}(M)$ where M is modelled on a C^∞ Banach space then we have defined $[X, Y]$ via the correspondence between vector fields and derivations on $C^\infty(M) = \mathfrak{F}(M)$. Even in the general case $[X, Y]$ makes sense as a derivation and if we already know that there is a vector field that acts as a derivation in the same way as $[X, Y]$ then it is unique by 7.5.

Theorem 7.7 *If $X, Y \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$ then $(L_X Y)f = [X, Y]f$. That is $L_X Y$ and $[X, Y]$ are equal as derivations. Thus $L_X Y = [X, Y]$ as vector fields. In other words, $L_X Y$ is the unique vector field such that*

$$(L_X Y)(p) = X_p(Yf) - Y_p(Xf)$$

Proof. We shall show that both sides of the equation act in the same way as derivations. Consider the map $\alpha : I \times I \to \mathbb{R}$ given by $\alpha(s, t) = Y(\mathrm{Fl}^X(p, s))(f \circ \mathrm{Fl}^{\mathbf{X}}_t)$. Notice that $\alpha(s, 0) = Y(\mathrm{Fl}^X(p, s))(f)$ and $\alpha(0, t) = Y(p)(f \circ \mathrm{Fl}^X_t)$ so that we have

$$\frac{\partial}{\partial s}\alpha(0,0) = \left.\frac{\partial}{\partial s}\right|_0 Y_{\mathrm{Fl}^{\mathbf{X}}(p,s)} f = \left.\frac{\partial}{\partial s}\right|_0 Yf \circ Fl^{\mathbf{X}}(p, s) = X_p Y f$$

and similarly

$$\frac{\partial}{\partial t}\alpha(0,0) = Y_p X f.$$

Subtracting we get $[X, Y](p)$. On the other hand we also have that

$$\frac{\partial}{\partial s}\alpha(0,0) - \frac{\partial}{\partial t}\alpha(0,0) = \left.\frac{d}{dr}\right|_0 \alpha(r, -r) = (L_X Y)_p$$

so we are done. To prove the last statement we just use lemma 7.5. ■

Theorem 7.8 *Let $X, Y \in \mathfrak{X}(M)$*

$$\frac{d}{dt}\mathrm{Fl}^{\mathbf{X}*}_{\mathbf{t}} Y = \mathrm{Fl}^{X*}_t(L_X Y)$$

Proof.

$$\begin{aligned}
\left.\frac{d}{dt}\right|_t \mathrm{Fl}^{X*}_t Y &= \left.\frac{d}{ds}\right|_0 \mathrm{Fl}^{X*}_{t+s} Y \\
&= \left.\frac{d}{ds}\right|_0 \mathrm{Fl}^{X*}_t(\mathrm{Fl}^{X*}_s Y) \\
&= \mathrm{Fl}^{X*}_t \left.\frac{d}{ds}\right|_0 (\mathrm{Fl}^{X*}_s Y) \\
&= \mathrm{Fl}^{X*}_t L_X Y
\end{aligned}$$

Now we can see that the infinitesimal version of the action

$$\begin{aligned}\Gamma(M,TM)\times \mathrm{Diff}(M) &\to \Gamma(M,TM)\\ (X,\Phi) &\mapsto \Phi^* X\end{aligned}$$

is just the Lie derivative. As for the left action of $\mathrm{Diff}(M)$ on $\mathfrak{X}(M)=\Gamma(M,TM)$ we have for $X,Y\in\mathfrak{X}(M)$

$$\begin{aligned}\left.\frac{d}{dt}\right|_0 (\mathrm{Fl}^X_{t*}Y)(p) &= \left.\frac{d}{dt}\right|_0 T\mathrm{Fl}^X_t(Y(\mathrm{Fl}_t^{-1}(p)))\\ &= -\left.\frac{d}{dt}\right|_0 (T\mathrm{Fl}^X_{-t})^{-1}Y(\mathrm{Fl}_{-t}(p))\\ &= -\left(L_X Y\right) = -[X,Y]\end{aligned}$$

It is easy to see that the Lie derivative is linear in both variables and over the reals. ■

Proposition 7.1 *Let $X\in\mathfrak{X}(M)$ and $Y\in\mathfrak{X}(N)$ be ϕ-related vector fields for a smooth map $\phi: M\to N$. Then*

$$\phi\circ\mathrm{Fl}^X_t = \mathrm{Fl}^Y_t\circ\phi$$

whenever both sides are defined. Suppose that $\phi: M\to M$ is a diffeomorphism and $X\in\mathfrak{X}(M)$. Then the flow of $\phi_* X=(\phi^{-1})^*X$ is $\phi\circ\mathrm{Fl}^X_t\circ\phi^{-1}$ and the flow of ϕ^*X is $\phi^{-1}\circ\mathrm{Fl}^X_t\circ\phi$.

Proof. Differentiating we have $\frac{d}{dt}(\phi\circ\mathrm{Fl}^X_t)=T\phi\circ\frac{d}{dt}\mathrm{Fl}^X_t=T\phi\circ X\circ\mathrm{Fl}^X_t=Y\circ\phi\circ\mathrm{Fl}^X_t$. But $\phi\circ\mathrm{Fl}^X_0(x)=\phi(x)$ and so $t\mapsto\phi\circ\mathrm{Fl}^X_t(x)$ is an integral curve of Y starting at $\phi(x)$. By uniqueness we have $\phi\circ\mathrm{Fl}^X_t(x)=\mathrm{Fl}^Y_t(\phi(x))$. ■

Lemma 7.8 *Suppose that $X\in\mathfrak{X}(M)$ and $\widetilde{X}\in\mathfrak{X}(N)$ and that $\phi: M\to N$ is a smooth map. Then $X\sim_\phi\widetilde{X}$ iff*

$$\widetilde{X}f\circ\phi = X(f\circ\phi)$$

for all $f\in C^\infty(U)$ and all open sets $U\subset N$.

Proof. We have $(\widetilde{X}f\circ\phi)(p)=df_{\phi(p)}\widetilde{X}(\phi(p))$. Using the chain rule we have

$$\begin{aligned}X(f\circ\phi)\,(p) &= d(f\circ\phi)(p)X_p\\ &= df_{\phi(p)}T_p\phi\cdot X_p\end{aligned}$$

and so if $X\sim_\phi\widetilde{X}$ then $T\phi\circ X=\widetilde{X}\circ\phi$ and so we get $\widetilde{X}f\circ\phi=X(f\circ\phi)$. On the other hand, if $\widetilde{X}f\circ\phi=X(f\circ\phi)$ for all $f\in C^\infty(U)$ and all open sets $U\subset N$ then we can pick U to be a chart domain and f the pull back of a linear functional on the model space $\mathbb{R}^n$. So assuming that we are actually on an open set $U \subset \mathbb{R}^n$ and $f = \alpha$ is any functional from $(\mathbb{R}^n)^*$ we would have $\widetilde{X}\alpha \circ \phi(p) = X(\alpha \circ \phi)$ or $d\alpha(\widetilde{X}_{\phi p}) = d(\alpha \circ \phi)X_p$ or again

$$\begin{aligned} d\alpha_{\phi p}(\widetilde{X}_{\phi p}) &= d\alpha_{\phi p}(T\phi \cdot X_p) \\ \alpha_{\phi p}(\widetilde{X}_{\phi p}) &= \alpha_{\phi p}(T\phi \cdot X_p) \end{aligned}$$

so by the Hahn-Banach theorem $\widetilde{X}_{\phi p} = T\phi \cdot X_p$. Thus since p was arbitrary $X \sim_\phi \widetilde{X}$. ■

Theorem 7.9 *Let $\phi : M \to N$ be a smooth map, $X, Y \in \mathfrak{X}(M)$, $\widetilde{X}, \widetilde{Y} \in \mathfrak{X}(N)$ and suppose that $X \sim_\phi \widetilde{X}$ and $Y \sim_\phi \widetilde{Y}$. Then*

$$[X, Y] \sim_\phi [\widetilde{X}, \widetilde{Y}]$$

In particular, if ϕ is a diffeomorphism then $[\phi_* X, \phi_* Y] = \phi_*[X, Y]$.

Proof. By lemma 7.8 we just need to show that for any open set $U \subset N$ and any $f \in C^\infty(U)$ we have $([\widetilde{X}, \widetilde{Y}]f) \circ \phi = [X, Y](f \circ \phi)$. We calculate using 7.8:

$$\begin{aligned} ([\widetilde{X}, \widetilde{Y}]f) \circ \phi &= \widetilde{X}(\widetilde{Y}f) \circ \phi - \widetilde{Y}(\widetilde{X}f) \circ \phi \\ &= X((\widetilde{Y}f) \circ \phi) - Y((\widetilde{X}f) \circ \phi) \\ = X(Y(f \circ \phi)) - Y(X(f \circ \phi)) &= [X, Y] \circ \phi. \end{aligned}$$

■

Theorem 7.10 *For $X, Y \in \mathfrak{X}(M)$ each of the following are equivalent:*

1. $\mathcal{L}_X Y = 0$
2. $(\mathrm{Fl}_t^X)^* Y = Y$
3. *The flows of X and Y commute:*

$$\mathrm{Fl}_t^X \circ \mathrm{Fl}_s^Y = \mathrm{Fl}_s^Y \circ \mathrm{Fl}_t^X \text{ whenever defined.}$$

Proof. The equivalence of 1 and 2 is follows easily from the proceeding results and is left to the reader. The equivalence of 2 and 3 can be seen by noticing that $\mathrm{Fl}_t^X \circ \mathrm{Fl}_s^Y = \mathrm{Fl}_s^Y \circ \mathrm{Fl}_t^X$ is true and defined exactly when $\mathrm{Fl}_s^Y = \mathrm{Fl}_t^X \circ \mathrm{Fl}_s^Y \circ \mathrm{Fl}_t^X$ which happens exactly when

$$\mathrm{Fl}_s^Y = \mathrm{Fl}_s^{(\mathrm{Fl}_t^X)^* Y}$$

and in turn exactly when $Y = (\mathrm{Fl}_t^X)^* Y$. ■

Proposition 7.2 $[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X$.

This is a straightforward calculation.

The Lie derivative and the Lie bracket are essentially the same object and are defined in for local sections $X \in \mathfrak{X}_M(U)$ as well as global sections. As is so often the case for operators in differential geometry, the Lie derivative is natural with respect to restriction so we have the commutative diagram

$$\begin{array}{ccc} \mathfrak{X}_M(U) & \stackrel{L_{X_U}}{\to} & \mathfrak{X}_M(U) \\ r_V^U \downarrow & & \downarrow r_V^U \\ \mathfrak{X}_M(V) & \stackrel{L_{X_V}}{\to} & \mathfrak{X}_M(V) \end{array}$$

where $X_U = X|_U$ denotes the restriction of $X \in \mathfrak{X}_M$ to the open set U and r_V^U is the map that restricts from U to $V \subset U$.

7.9 Time Dependent Fields

Consider a small charged particle pushed along by the field set up by a large stationary charged body. The particle will follow integral curves of the field. What if while the particle is being pushed along, the larger charged body responsible for the field is put into motion? The particle must still have a definite trajectory but now the field it *time dependent*. To see what difference this makes, consider a time dependent vector field $X(t, .)$ on a manifold M. For each $x \in M$ let $\phi_t(x)$ be the point at which a particle that was at x at time 0,ends up after time t. Will it be true that $\phi_s \circ \phi_t(x) = \phi_{s+t}(x)$? The answer is the in general this equality does not hold. The flow of a time dependent vector field is not a 1-parameter group. On the other hand, if we define $\Phi_{s,t}(x)$ to be the location of the particle which was at x at time s at the later time t then we expect

$$\Phi_{s,r} \circ \Phi_{r,t} = \Phi_{s,t}$$

which is called the Chapman-Kolmogorov law. If in a special case $\Phi_{r,t}$ depends only on $s - t$ then setting $\phi_t := \Phi_{0,t}$ we recover a flow corresponding to a time-independent vector field. The formal definitions are as follows:

Definition 7.18 *A C^r **time dependent vector field** on M is a C^r map $X : (a, b) \times M \to TM$ such that for each fixed $t \in (a, b) \subset \mathbb{R}$ the map $X_t : M \to TM$ given by $X_t(x) := X(t, x)$ is a C^r vector field.*

Definition 7.19 *Let X be a time dependent vector field. A curve $c : (a, b) \to M$ is called an **integral curve** of X iff*

$$\dot{c}(t) = X(t, c(t)) \text{ for all } t \in (a, b).$$

*The **evolution operator** $\Phi^X_{t,s}$ for X is defined by the requirement that*

$$\frac{d}{dt}\Phi^X_{t,s}(x) = X(t, \Phi^X_{t,s}(x)) \text{ and } \Phi^X_{s,s}(x) = x.$$

In other words, $t \mapsto \Phi^X_{t,s}(x)$ is the integral curve which goes through x at time s.

We have chosen to use the term "evolution operator" as opposed to "flow" in order to emphasize that the local group property does not hold in general. Instead we have the following

Theorem 7.11 *Let X be a time dependent vector field. Suppose that $X_t \in \mathfrak{X}^r(M)$ for each t and that $X : (a,b) \times M \to TM$ is continuous. Then $\Phi^X_{t,s}$ is C^r and we have $\Phi_{s,a} \circ \Phi_{a,t} = \Phi_{s,t}$ whenever defined.*

Theorem 7.12 *Let X and Y be smooth time dependent vector fields and let $f : \mathbb{R}\times M \to \mathbb{R}$ be smooth. We have the following formulas:*

$$\frac{d}{dt}(\Phi^X_{t,s})^* f_t = (\Phi^X_{t,s})^* (\frac{\partial f}{\partial t} + X_t f_t)$$

where $f_t() := f(t,.)$, and

$$\frac{d}{dt}(\Phi^X_{t,s})^* Y_t = (\Phi^X_{t,s})^* (\frac{\partial Y}{\partial t} + [X_t, Y_t]).$$

Chapter 8

Lie Groups II

Definition 8.1 *For a Lie group G and a fixed $g \in G$, the maps L_g :$G \to G$ and $R_g : G \to G$ are defined by*

$$L_g x = gx \text{ for } x \in G$$
$$R_g x = xg \text{ for } x \in G$$

*and are called **left translation** and **right translation** respectively.*

Definition 8.2 *A vector field $X \in \mathfrak{X}(G)$ is called **left invariant** iff $(L_g)_* X = X$ for all $g \in G$. A vector field $X \in \mathfrak{X}(G)$ is called **right invariant** iff $(R_g)_* X = X$ for all $g \in G$. The set of left invariant (resp. right invariant) vectors Fields is denoted $\mathfrak{L}(G)$ or $\mathfrak{X}^L(G)$ (resp. $\mathfrak{R}(G)$ or $\mathfrak{X}^R(G)$).*

Note that $X \in \mathfrak{X}(G)$ is left invariant iff the following diagram commutes

$$\begin{array}{ccc} TG & \stackrel{TL_g}{\to} & TG \\ \downarrow & & \downarrow \\ G & \stackrel{L_g}{\to} & G \end{array}$$

for every $g = g \in G$. There is a similar diagram for right invariance.

Lemma 8.1 *$\mathfrak{X}^L(G)$ is closed under the Lie Bracket operation.*

Proof. Suppose that $X, Y \in \mathfrak{X}^L(G)$. Then by 7.9 we have

$$\begin{aligned}(L_g)_*[X,Y] &= \mathcal{L}_X Y = \mathcal{L}_{L_{g*}X} L_{g*} Y \\ &= [L_{g*}X, L_{g*}Y] = [X,Y]\end{aligned}$$

∎

Corollary 8.1 *$\mathfrak{X}^L(G)$ is an n-dimensional Lie algebra under the Bracket of vector fields (see definition **??**).*

Lemma 8.2 *Given a vector $v \in T_eG$ we can define a smooth left (resp. right) invariant vector field L^v (resp. R^v) such that $L^v(e) = v$ (resp. $R^v(e) = v$) by the simple prescription*

$$L^v(g) = TL_g \cdot v$$
$$R^v(g) = (\text{resp. } TR_g \cdot v)$$

and furthermore the map $v \mapsto L^v$ is a linear isomorphism from T_eG onto $\mathfrak{X}^L(G)$.

Proof. The proof that this gives the invariant vector fields as prescribed is easy and left as an exercise. ■

Definition 8.3 *To every $v \in T_eG$ we associate a left (resp. right) invariant vector field via the map*

$$L : v \mapsto L^v \text{ (resp. } R : v \mapsto R^v\text{)}.$$

Now we can transfer the Lie algebra structure to T_eG by defining a bracket operation on T_eG by using the bracket of the corresponding left invariant vector fields.

Definition 8.4 *For a Lie group G,define the bracket of any two elements $v, w \in T_eG$ by*

$$[v, w] := [L^v, L^w](e).$$

The Lie algebras thus defined are isomorphic by construction both of them are often referred to as the Lie algebra of the Lie group G and denoted by $\mathfrak{Lie}(G)$ or $\mathfrak{g}$. Of course we are implying that $\mathfrak{Lie}(H)$ is denoted $\mathfrak{h}$ and $\mathfrak{Lie}(K)$ by $\mathfrak{k}$ etc. Our specific convention will be that $\mathfrak{g} = \mathfrak{Lie}(G) := T_eG$ with the bracket defined above and then occasionally identify this with the left invariant fields $\mathfrak{X}^L(G)$ with the vector field Lie bracket defined in definition **??**.

Let us compute the form of the Lie bracket for the Lie algebra of the matrix general linear group. First of all this Lie algebra is $T_I\mathrm{GL}(n)$ which is canonically isomorphic to the vector space of all matrices $\mathbb{M}_{n\times n}$ so we set

$$\mathfrak{gl}(n) = \mathbb{M}_{n\times n}.$$

Now a global coordinate system for $\mathrm{GL}(n)$ is given by the maps

$$x^k_l : (a^i_j) \mapsto a^k_l.$$

Thus any vector fields $X, Y \in \mathfrak{X}(\mathrm{GL}(n))$ can be written

$$X = f^i_j \frac{\partial}{\partial x^i_j}$$
$$Y = g^i_j \frac{\partial}{\partial x^i_j}$$

and hence for some functions f^i_j and g^i_j. Under the canonical isomorphism we have two matrix valued functions $F = (f^i_j)$ and $G = (g^i_j)$. The bracket is then

$$[X, Y] = (f^i_j \frac{\partial g^k_l}{\partial x^i_j} - f^i_j \frac{\partial g^k_l}{\partial x^i_j}) \frac{\partial}{\partial x^k_l}$$

Now the left invariant vector fields are given by $g \mapsto gA$ for constant matrix A. In coordinates, left invariant vector fields corresponding to $A, B \in \mathfrak{gl}(n)$ would be

$$X = x^i_k a^k_j \frac{\partial}{\partial x^i_j} \text{ and}$$
$$Y = x^i_k b^k_j \frac{\partial}{\partial x^i_j}$$

and then the bracket

$$[X, Y] = \left(a^r_s \frac{\partial(x^i_k b^k_j)}{\partial x^r_s} - b^r_s \frac{\partial(x^i_k a^k_j)}{\partial x^r_s} \right) \frac{\partial}{\partial x^i_j}$$
$$= (a^i_k b^k_j - b^i_k a^k_j) \frac{\partial}{\partial x^i_j}$$

which corresponds to the matrix $AB - BA$. So we have

Proposition 8.1 *Under the identification of* $T_I \mathrm{GL}(n)$ *with* $\mathbb{M}_{n\times n}$ *the bracket is the commutator bracket*

$$[A, B] = AB - BA$$

Similarly, under the identification of $T_{\mathrm{id}} Gl(\mathrm{V})$ *with* $\mathrm{End}(\mathrm{V})$ *the bracket is*

$$[A, B] = A \circ B - B \circ A.$$

Now we can use the maps $TL_{g^{-1}}$ and $TR_{g^{-1}}$ to identify vectors in $T_g G$ with unique vectors in $T_e G = \mathfrak{g}$: Define the maps $\omega_G : TG \to \mathfrak{g}$ (resp. $\omega^{right}_G : TG \to \mathfrak{g}$) by

$$\omega_G(X_g) = TL_{g^{-1}} \cdot X_g$$
$$(\text{resp. } \omega^{right}_G(X_g) = TR_{g^{-1}} \cdot X_g)$$

ω_G is a $\mathfrak{g}$ valued 1-form called the (left-) Maurer Cartan form. We will call ω^{right}_G the right-Maurer Cartan form but we will not be using it to the extent of ω_G. Now we can define maps $\mathrm{triv}_L : TG \to G \times \mathfrak{g}$ and $\mathrm{triv}_R : TG \to G \times \mathfrak{g}$ by

$$\mathrm{triv}_L(v_g) = (g, \omega_G(v_g))$$
$$\mathrm{triv}_R(v_g) = (g, \omega^{right}_G(v_g))$$

for $v_g \in T_g G$. These maps are both vector bundle isomorphisms. Thus we have the following:

Proposition 8.2 *The tangent bundle of a Lie group is trivial: $TG \cong G \times \mathfrak{g}$.*

Proof. Follows from the above discussion. ∎

Lemma 8.3 (Left-right lemma) *For any $v \in \mathfrak{g}$ the map $g \mapsto \operatorname{triv}_L^{-1}(g, v)$ is a left invariant vector field on G while $g \mapsto \operatorname{triv}_R^{-1}(g, v)$ is right invariant. Also, $\operatorname{triv}_R \circ \operatorname{triv}_L^{-1}(g, v) = (g, \operatorname{Ad}_g(v))$.*

Proof. The invariance is easy to check and is left as an exercise. Now the second statement is also easy:

$$\begin{aligned}
&\operatorname{triv}_R \circ \operatorname{triv}_L^{-1}(v) \\
&= (g, TR_{g^{-1}}TL_g v) = (g, T(R_{g^{-1}}L_g) \cdot v) \\
&= (g, \operatorname{Ad}_g(v)).
\end{aligned}$$

∎

Example 8.1 *Consider the group $SU(2)$ defined above. Suppose that $g(t)$ is a curve in $SU(2)$ with $g(0) = e = I$. Then $g(t)\overline{g(t)}^t = I$ and so*

$$\begin{aligned}
0 &= \left.\frac{d}{dt}\right|_0 g(t)\overline{g(t)}^t \\
&= \overline{g'(0)}^t + g'(0)
\end{aligned}$$

so $g'(0)$ is skew-Hermitian: $\overline{g'(0)}^t = -g'(0)$. Thus every element of $T_eSU(2) = \mathfrak{su}(2)$ is a skew-Hermitian matrix. But we also have the restriction that $\det(g(t)) = 1$ and this means that $g'(0)$ also has to have trace zero. In fact, if A is any skew-Hermitian matrix with trace zero then $g(t) = e^{tA}$ is a curve with $g'(0) = A$ so the Lie algebra $\mathfrak{su}(2)$ is the set vector space of skew-Hermitian matrices with trace zero. We can also think of the Lie algebra of $SU(2)$ as Hermitian matrices of trace zero since $A \longleftrightarrow iA$ is an isomorphism between the skew-Hermitian matrices and the Hermitian matrices.

Proposition 8.3 *Given a Lie group homomorphism $h : G_1 \to G_2$ we have that the map $T_eh : \mathfrak{g}_1 \to \mathfrak{g}_2$ is a Lie algebra homomorphism called the **Lie differential** which is often denoted in this context by $dh : \mathfrak{g}_1 \to \mathfrak{g}_2$ or by $Lh : \mathfrak{g}_1 \to \mathfrak{g}_2$.*

Proof. For $v \in \mathfrak{g}_1$ and $x \in G$ we have

$$\begin{aligned}
T_xh \cdot L^v(x) &= T_xh \cdot (T_eL_x \cdot v) \\
&= T_e(h \circ L_x) \cdot v \\
&= T_e(L_{h(x)} \circ h) \cdot v \\
&= T_e(L_{h(x)}) \circ T_eh \cdot v \\
&= L^{dh(v)}(h(x))
\end{aligned}$$

so $L^v \backsim_h L^{dh(v)}$. Thus by 7.9 we have for any $v, w \in \mathfrak{g}_1$ that $L^{[v,w]} \backsim_h [L^{dh(v)}, L^{dh(w)}]$ or in other words, $[L^{dh(v)}, L^{dh(w)}] \circ h = Th \circ L^{[v,w]}$ which at e gives

$$[dh(v), dh(w)] = [v, w].$$

■

Theorem 8.1 *Let invariant vector fields are complete. The integral curves through the identify element are the one-parameter subgroups.*

Proof. Let X be a left invariant vector field and $c : (a, b) \to G$ be the integral curve of X with $c(0) = X(p)$. Let $a < t_1 < t_2 < b$ and choose an element $g \in G$ such that $gc(t_1) = c(t_2)$. Let $\Delta t = t_2 - t_1$ and define $\bar{c} : (a + \Delta t, b + \Delta t) \to G$ by $\bar{c}(t) = gc(t - \Delta t)$. Then we have

$$\begin{aligned} \bar{c}'(t) = TL_g \cdot c'(t - \Delta t) &= TL_g \cdot X(c(t - \Delta t)) \\ &= X(gc(t - \Delta t)) = X(\bar{c}(t)) \end{aligned}$$

and so $\bar{c}$ is also an integral curve of X. Now on the intersection $(a + \Delta t, b)$ of the their domains, c and $\bar{c}$ are equal since they are both integral curve of the same field and since $\bar{c}(t_2) = gc(t_1) = c(t_2)$. Thus we can concatenate the curves to get a new integral curve defined on the larger domain $(a, b + \Delta t)$. Since this extension can be done again for a fixed Δt we see that c can be extended to (a, ∞). A similar argument gives that we can extend in the negative direction to get the needed extension of c to $(-\infty, \infty)$.

Next assume that c is the integral curve with $c(0) = e$. The proof that $c(s+t) = c(s)c(t)$ proceeds by considering $\gamma(t) = c(s)^{-1}c(s+t)$. Then $\gamma(0) = e$ and also

$$\begin{aligned} \gamma'(t) &= TL_{c(s)^{-1}} \cdot c'(s + t) = TL_{c(s)^{-1}} \cdot X(c(s + t)) \\ &= X(c(s)^{-1}c(s + t)) = X(\gamma(t)). \end{aligned}$$

By the uniqueness of integral curves we must have $c(s)^{-1}c(s + t) = c(t)$ which implies the result. Conversely, if $c : \mathbb{R} \to G$ is a one parameter subgroup the let $X_e = \dot{c}(0)$ then there is a left invariant vector field X such that $X(e) = X_e$. We must show that the integral curve through e of the field X is exactly c. But for this we only need that $\dot{c}(t) = X(c(t))$ for all t. Now $c(t + s) = c(s)$ or $c(t + s) = L_{c(t)}c(s)$. Thus

$$\dot{c}(t) = \left.\frac{d}{ds}\right|_0 c(t + s) = (T_{c(t)}L).\dot{c}(0) = X(c(t))$$

and we are done. ■

Lemma 8.4 *Let $v \in \mathfrak{g} = T_eG$ and the corresponding left invariant field L^v. Then with $Fl^v := Fl^{L^v}$ we have that*

$$Fl^v(st) = Fl^{sv}(t) \tag{8.1}$$

A similar statement holds with R^v replacing L^v.

Proof. Let $u = st$. We have that $\frac{d}{dt}\big|_{t=0} Fl^v(st) = \frac{du}{dt}\frac{d}{du}\big|_{t=0} Fl^v(u)\frac{du}{dt} = sv$ and so by uniqueness $Fl^v(st) = Fl^{sv}(t)$. ■

Definition 8.5 *For any $v \in \mathfrak{g} = T_eG$ we have the corresponding left invariant field L^v which has an integral curve through e which we denote by $\exp(tv)$. Thus the map $t \to \exp(tv)$ is a Lie group homomorphism from $\mathbb{R}$ into G which is a one-parameter subgroup. The map $v \mapsto \exp(1v) = \exp^G(v)$ is referred to as the* ***exponential map*** $\exp^G : \mathfrak{g} \to G$.

Lemma 8.5 $\exp^G : \mathfrak{g} \to G$ *is smooth.*

Proof. Consider the map $\mathbb{R} \times G \times \mathfrak{g} \to G \times \mathfrak{g}$ given by $(t, g, v) \mapsto (g \cdot \exp^G(tv), v)$. This map is easily seen to be the flow on $G \times \mathfrak{g}$ of the vector field $\widetilde{X} : (g, v) \mapsto (L^v(g), 0)$ and so is smooth. Now the restriction of this smooth flow to the submanifold $\{1\} \times \{e\} \times \mathfrak{g}$ is $(1, e, v) \mapsto (\exp^G(v), v)$ is also smooth which clearly implies that $\exp^G(v)$ is smooth also. ■

Theorem 8.2 *The tangent map of the exponential map $\exp^G : \mathfrak{g} \to G$ is the identity at the origin $0 \in T_eG = \mathfrak{g}$ and $\exp$ is a diffeomorphism of some neighborhood of the origin onto its image in G.*

Proof. By lemma 8.5 we know that $\exp^G : \mathfrak{g} \to G$ is a smooth map. Also, $\frac{d}{dt}\big|_0 \exp^G(tv) = v$ which means the tangent map is $v \mapsto v$. If the reader thinks through the definitions carefully, he or she will discover that we have here used the natural identification of $\mathfrak{g}$ with $T_0\mathfrak{g}$. ■

Remark 8.1 *The "one-parameter subgroup" $\exp^G(tv)$ corresponding to a vector $v \in \mathfrak{g}$ is actually a homomorphism rather than a subgroup but the terminology is conventional.*

Proposition 8.4 *For a (Lie group) homomorphism $h : G_1 \to G_2$ the following diagram commutes:*

$$\begin{array}{ccc} \mathfrak{g}_1 & \overset{dh}{\to} & \mathfrak{g}_2 \\ \exp^{G_1}\downarrow & & \exp^{G_2}\downarrow \\ G_1 & \overset{h}{\to} & G_2 \end{array}$$

Proof. The curve $t \mapsto h(\exp^{G_1}(tv))$ is clearly a one parameter subgroup. Also,

$$\frac{d}{dt}\bigg|_0 h(\exp^{G_1}(tv)) = dh(v)$$

so by uniqueness of integral curves $h(\exp^{G_1}(tv)) = \exp^{G_2}(tdh(v))$. ■

Remark 8.2 *We will sometimes not index the maps and shall just write* $\exp$ *for any Lie group.*

The reader may wonder what happened to right invariant vector fields and how do they relate to one parameter subgroups. The following theorem give various relationships.

Theorem 8.3 *For a smooth curve $c : \mathbb{R} \to G$ with $c(0) = e$ and $\dot{c}(0) = v$, the following are all equivalent:*

1. $c(t) = Fl_t^{L^v}(e)$
2. $c(t) = Fl_t^{R^v}(e)$
3. *c is a one parameter subgroup.*
4. $Fl_t^{L^v} = R_{c(t)}$
5. $Fl_t^{R^v} = L_{c(t)}$

Proof. By definition $Fl_t^{L^v}(e) = \exp(tv)$. We have already shown that 1 implies 3. The proof of 2 implies 3 would be analogous. We have also already shown that 3 implies 1.

Also, 4 implies 1 since then $Fl_t^{L^v}(e) = R_{c(t)}(e) = c(t)$. Now assuming 1 we have

$$\begin{aligned} c(t) &= Fl_t^{L^v}(e) \\ \left.\frac{d}{dt}\right|_0 c(t) &= L^v(e) \\ \left.\frac{d}{dt}\right|_0 gc(t) &= \left.\frac{d}{dt}\right|_0 L_g(c(t)) \\ &= TL_g v = L^v(g) \text{ for any } g \\ \left.\frac{d}{dt}\right|_0 R_{c(t)} g &= L^v(g) \text{ for any } g \\ R_{c(t)} &= Fl_t^{L^v} \end{aligned}$$

The rest is left to the reader. ∎

The tangent space at the identity of a Lie group G is a vector space and is hence a manifold. Thus exp is a smooth map between manifolds. As is usual we identify the tangent space $T_v(T_e\dot{G})$ at some $v \in T_eG$ with T_eG itself. The we have the following

Lemma 8.6 $T_e \exp = \mathrm{id} : T_eG \to T_eG$

Proof. $T_e \exp \cdot v = \left.\frac{d}{dt}\right|_0 \exp(tv) = v.$ ∎

The Lie algebra of a Lie group and the group itself are closely related in many ways. One observation is the following:

Proposition 8.5 *If G is a connected Lie group then for any open neighborhood $V \subset \mathfrak{g}$ of 0 the group generated by $\exp(V)$ is all of G.*

sketch of proof. Since $T_e \exp = \mathrm{id}$ we have that exp is an open map near 0. The group generated by $\exp(V)$ is a subgroup containing an open neighborhood of e. The complement is also open. ∎

Now we prove a remarkable theorem which shows how an algebraic assumption can have implications in the differentiable category. First we need some notation.

Notation 8.1 *If S is any subset of a Lie group G then we define*

$$S^{-1} = \{s^{-1} : s \in S\}$$

and for any $x \in G$ we define

$$xS = \{xs : s \in S\}.$$

Theorem 8.4 *An abstract subgroup H of a Lie group G is a (regular) submanifold iff H is a closed set in G. If follows that H is a (regular) Lie subgroup of G.*

Proof. First suppose that H is a (regular) submanifold. Then H is locally closed. That is, every point $x \in H$ has an open neighborhood U such that $U \cap H$ is a relatively closed set in H. Let U be such a neighborhood of the identity element e. We seek to show that H is closed in G. Let $y \in \overline{H}$ and $x \in yU^{-1} \cap H$. Thus $x \in H$ and $y \in xU$. Now this means that $y \in \overline{H} \cap xU$, and thence $x^{-1}y \in \overline{H} \cap U = H \cap U$. So $y \in H$ and we have shown that H is closed.

Now conversely, let us suppose that H is a closed abstract subgroup of G. Since we can always use the diffeomorphism to translate any point to the identity it suffices to find a neighborhood U of e such that $U \cap H$ is a submanifold. The strategy is to find out what $\mathrm{Lie}(H) = \mathfrak{h}$ is likely to be and then exponentiate a neighborhood of $e \in \mathfrak{h}$.

First we will need to have an inner product on T_eG so choose any such. Then norms of vectors in T_eG makes sense. Choose a small neighborhood $\widetilde{U}$ of $0 \in T_eG = \mathfrak{g}$ on which exp is a diffeomorphism say $\exp : \widetilde{U} \to U$ with inverse denoted by $\log_U$. Define the set $\widetilde{H}$ in $\widetilde{U}$ by $\widetilde{H} = \log_U(H \cap U)$.

Claim 8.1 *If h_n is a sequence in $\widetilde{H}$ converging to zero and such that $u_n = h_n / |h_n|$ converges to $v \in \mathfrak{g}$ then $\exp(tv) \in H$ for all $t \in \mathbb{R}$.*

Proof of claim: Note that $th_n / |h_n| \to tv$ while $|h_n|$ converges to zero. But since $|h_n| \to 0$ we must be able to find a sequence $k(n) \in \mathbb{Z}$ such that $k(n)|h_n| \to t$. From this we have $\exp(k(n)h_n) = \exp(k(n)|h_n| \frac{h_n}{|h_n|}) \to \exp(tv)$. But by the properties of exp proved previously we have $\exp(k(n)h_n) = (\exp(h_n))^{k(n)}$. But $\exp(h_n) \in H \cap U \subset H$ and so $(\exp(h_n))^{k(n)} \in H$. But since H is closed we have $\exp(tv) = \lim_{n\to\infty}(\exp(h_n))^{k(n)} \in H$.

Claim 8.2 *The set W of all tv where v can be obtained as a limit $h_n / |h_n| \to v$ with $h_n \in \widetilde{H}$ is a vector space.*

Proof of claim: It is enough to show that if $h_n/|h_n| \to v$ and $h'_n/|h'_n| \to w$ with $h'_n, h_n \in \widetilde{H}$ then there is a sequence of elements h''_n from $\widetilde{H}$ with

$$h''_n/|h''_n| \to \frac{v+w}{|v+w|}.$$

This will follow from the observation that

$$h(t) = \log_U(\exp(tv)\exp(tw))$$

is in $\widetilde{H}$ and by exercise 5.5 we have that

$$\lim_{t \downarrow 0} h(t)/t = v + w$$

and so

$$\frac{h(t)/t}{|h(t)/t|} \to \frac{v+w}{|v+w|}.$$

The proof of the next claim will finish the proof of the theorem.

Claim: *Let W be the set from the last claim. Then $\exp(W)$ contains an open neighborhood of e in H. Let $W^\perp$ be the orthogonal compliment of W with respect to the inner product chosen above. Then we have $T_eG = W^\perp \oplus W$. It is not difficult to show that the map $\Sigma : W \oplus W^\perp \to G$ defined by*

$$v + w \mapsto \exp(v)\exp(w)$$

is a diffeomorphism in a neighborhood of the origin in T_eG. Now suppose that $\exp(W)$ does not contain an open neighborhood of e in H. Then we can choose a sequence $(v_n, w_n) \in W \oplus W^\perp$ with $(v_n, w_n) \to 0$ and $\exp(v_n)\exp(w_n) \in H$ and yet $w_n \neq 0$. The space $W^\perp$ and the unit sphere in $W^\perp$ is compact so after passing to a subsequence we may assume that $w_n/|w_n| \to w \in W^\perp$ and of course $|w| = 1$. Since $\exp(v_n) \in H$ and H is at least an algebraic subgroup, $\exp(v_n)\exp(w_n) \in H$, it must be that $\exp(w_n) \in H$ also. But then by the definition of W we have that $w \in W$ which contradicts the fact that $|w| = 1$ and $w \in W^\perp$.

■

8.1 Spinors and rotation

The matrix Lie group $SO(3)$ is the group of orientation preserving rotations of $\mathbb{R}^3$ acting by matrix multiplication on column vectors. The group $SU(2)$ is the group of complex 2×2 unitary matrices of determinant 1. We shall now expose an interesting relation between these groups. First recall the Pauli matrices:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The real vector space spanned by $\sigma_1, \sigma_2, \sigma_3$ is isomorphic to $\mathbb{R}^3$ and is the space of traceless Hermitian matrices. Let us temporarily denote the latter by $\widehat{\mathbb{R}^3}$. Thus we have a linear isomorphism $\mathbb{R}^3 \to \widehat{\mathbb{R}^3}$ given by $(x^1, x^2, x^3) \mapsto x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3$ which abbreviate to $\vec{x} \mapsto \widehat{x}$. Now it is easy to check that $\det(\widehat{x})$ is just $-|\vec{x}|^2$. In fact, we may introduce an inner product on $\widehat{\mathbb{R}^3}$ by the formula $\langle \widehat{x}, \widehat{y} \rangle := -\frac{1}{2} tr(\widehat{x}\widehat{y})$ and then we have that $\vec{x} \mapsto \widehat{x}$ is an isometry. Next we notice that $SU(2)$ acts on $\widehat{\mathbb{R}^3}$ by $(g, \widehat{x}) \mapsto g\widehat{x}g^{-1} = g\widehat{x}g^*$ thus giving a representation ρ of $SU(2)$ in $\widehat{\mathbb{R}^3}$. It is easy to see that $\langle \rho(g)\widehat{x}, \rho(g)\widehat{y} \rangle = \langle \widehat{x}, \widehat{y} \rangle$ and so under the identification $\mathbb{R}^3 \leftrightarrow \widehat{\mathbb{R}^3}$ we see that $SU(2)$ act on $\mathbb{R}^3$ as an element of $O(3)$.

Exercise 8.1 *Show that in fact, the map $SU(2) \to O(3)$ is actually a group map onto $SO(3)$ with kernel $\{\pm I\} \cong \mathbb{Z}_2$.*

Exercise 8.2 *Show that the algebra generated by the matrices $\sigma_0, i\sigma_1, i\sigma_2, i\sigma_3$ is isomorphic to the quaternion algebra and that the set of matrices $i\sigma_1, i\sigma_2, i\sigma_3$ span a real vector space which is equal as a set to the traceless skew Hermitian matrices $\mathfrak{su}(2)$.*

Let $I = i\sigma_1$, $J = i\sigma_2$ and $i\sigma_3 = K$. One can redo the above analysis using the isometry $\mathbb{R}^3 \to \mathfrak{su}(2)$ given by

$$\begin{aligned}(x^1, x^2, x^3) &\mapsto x^1 I + x^2 J + x^3 K \\ \vec{x} &\mapsto \widetilde{x}\end{aligned}$$

where this time $\langle \widetilde{x}, \widetilde{y} \rangle := \frac{1}{2} tr(\widetilde{x}\widetilde{y}^*) = -\frac{1}{2} tr(\widetilde{x}\widetilde{y})$. Notice that $\mathfrak{su}(2) = span\{I, J, K\}$ is the Lie algebra of $SU(2)$ the action $(g, \widehat{x}) \mapsto g\widehat{x}g^{-1} = g\widehat{x}g^*$ is just the adjoint action to be defined in a more general setting below. Anticipating this, let us write $Ad(g) : \widehat{x} \mapsto g\widehat{x}g^*$. This gives the map $g \mapsto Ad(g)$; a Lie group homomorphism $SU(2) \to SO(\mathfrak{su}(2), \langle , \rangle)$. Once again we get the same map $SU(2) \to O(3)$ which is a Lie group homomorphism and has kernel $\{\pm I\} \cong \mathbb{Z}_2$. In fact, we have the following commutative diagram:

$$\begin{array}{ccc} SU(2) & = & SU(2) \\ \rho \downarrow & & Ad \downarrow \\ SO(3) & \cong & SO(\mathfrak{su}(2), \langle , \rangle) \end{array}$$

Exercise 8.3 *Check the details here!*

What is the differential of the map $\rho : SU(2) \to O(3)$ at the identity? Let $g(t)$ be a curve in $SU(2)$ with $\frac{d}{dt}\big|_{t=0} g = g'$. We have $\frac{d}{dt}(g(t)Ag^*(t)) = (\frac{d}{dt}g(t))Ag^*(t) + g(t)A(\frac{d}{dt}g(t))^*$ and so the map $ad : g' \mapsto g'A + Ag'^* = [g', A]$

$$\begin{aligned}
\frac{d}{dt}\langle g\widehat{x}, g\widehat{y}\rangle &= \frac{d}{dt}\frac{1}{2}tr(g\widetilde{x}(g\widetilde{y})^*) \\
&\frac{1}{2}tr([g', \widetilde{x}], (\widetilde{y})^*) + \frac{1}{2}tr(\widetilde{x}, ([g', \widetilde{y}])^*) \\
&== \frac{1}{2}tr([g', \widetilde{x}], (\widetilde{y})^*) - \frac{1}{2}tr(\widetilde{x}, [g', \widetilde{y}]) \\
&= \langle [g', \widetilde{x}], \widetilde{y}\rangle - \langle \widetilde{x}, [g', \widetilde{y}]\rangle \\
&= .\langle ad(g')\widetilde{x}, \widetilde{y}\rangle - \langle \widetilde{x}, ad(g')\widetilde{y}\rangle
\end{aligned}$$

From this is follows that the differential of the map $SU(2) \to O(3)$ takes $\mathfrak{su}(2)$ isomorphically onto the space $\mathfrak{so}(3)$. We have

$$\begin{array}{ccc}
\mathfrak{su}(2) & = & \mathfrak{su}(2) \\
d\rho \downarrow & & ad \downarrow \\
SO(3) & \cong & \mathfrak{so}(\mathfrak{su}(2), \langle , \rangle)
\end{array}$$

where $\mathfrak{so}(\mathfrak{su}(2), \langle , \rangle)$ denotes the linear maps $\mathfrak{su}(2) \to \mathfrak{su}(2)$ skew-symmetric with respect to the inner product $\langle \widetilde{x}, \widetilde{y}\rangle := \frac{1}{2}tr(\widetilde{x}\widetilde{y}^*)$.

Chapter 9

Multilinear Bundles and Tensors Fields

A great many people think they are thinking when they are merely rearranging their prejudices.
-William James (1842-1910)

Synopsis: Multilinear maps, tensors, tensor fields.

9.1 Multilinear Algebra

There a just a few things from multilinear algebra that are most important for differential geometry. Multilinear spaces and operations may be defined starting with in the category vector spaces and linear maps but we are also interested in vector bundles and their sections. At this point we may consider the algebraic structure possessed by the sections of the bundle. What we have then is not (only) a vector space but (also) a *module* over the *ring* of smooth functions. The algebraic operation we perform on this level are very similar to vector space calculations but instead of the scalars being the real (or complex) numbers the scalars are functions. So even though we start by defining things for vector spaces it is more efficient to consider modules over a commutative ring R since vector spaces are also modules whose scalar ring just happens to be a field. We do not want spend too much time with the algebra so that we may return to geometry. However, an appendix has been included that covers the material in a more general and systematic way that we do in this section. The reader may consult the appendix when needed.

The tensor space of type (r, s) is the vector space $T^r|_s(\mathrm{V})$ of multilinear maps of the form

$$\Upsilon : \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \times \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \to \mathrm{W}.$$

There is a natural bilinear operation called the (consolidated[1]) tensor product $T^{r_1}|_{s_1}(\mathrm{V}) \times T^{r_2}|_{s_2}(\mathrm{V}) \to T^{r_1+r_2}|_{s_1+s_2}(\mathrm{V})$ defined for $\Upsilon_1 \in T^{r_1}|_{s_1}(\mathrm{V})$ and $\Upsilon_2 \in T^{r_2}|_{s_2}(\mathrm{V})$ by

$$\begin{aligned}&(\Upsilon_1 \otimes \Upsilon_2)(\alpha_1, ..., \alpha_r, v_1, ..., v_s)\\ &:= \Upsilon_1(\alpha_1, ..., \alpha_r)\Upsilon_2(v_1, ..., v_s)\end{aligned}$$

We may put the various types together into direct sum which is a sort of master tensor space.

$$T^{\Diamond}_{\Diamond}(V) = \sum_{r,s} T^r|_s$$

$T^{\Diamond}_{\Diamond}(V)$ is an algebra over $\mathbb{R}$ with the tensor product. In practice, if one is interested in an algebra that is useful for constructing other algebras then the smaller space $T_{\Diamond}(V) = \sum_s T^0_s\Big|$ will usually do the job. It is this later space that is usually refereed to as the tensor algebra books and simply denoted $T(V)$.

Tensors and the tensor algebra is often defined in another way that is equivalent for finite dimensional vector spaces. The alternate approach is based on the idea of the tensor product of two vector spaces. The tensor product $\mathrm{V} \otimes \mathrm{W}$ of vector spaces V and W is often defined in an abstract manner as the universal object in the category whose objects are bilinear maps $\mathrm{V} \times \mathrm{W} \to \mathrm{E}$ and where a morphism between and object $\mathrm{V} \times \mathrm{W} \to \mathrm{E}_1$ and an object $\mathrm{V} \times \mathrm{W} \to \mathrm{E}_2$ is a map $\mathrm{E}_1 \to \mathrm{E}_2$ which make the following diagram commute:

$$\begin{array}{ccc} \mathrm{V} \times \mathrm{W} & \to & \mathrm{E}_1 \\ & \searrow & \downarrow \\ & & \mathrm{E}_2 \end{array}$$

On the other hand it is quite easy to say what $\mathrm{V} \otimes \mathrm{W}$ is in practical terms: It is just the set of all symbols $\mathrm{v} \otimes \mathrm{w}$ where $\mathrm{v} \in \mathrm{V}$ and $\mathrm{w} \in \mathrm{W}$ and subject to the relations $\mathrm{v} \otimes \mathrm{w}$

Our working definition comes out of the following theorem on the existence of the universal object:

Theorem 9.1 *Given finite dimensional vector spaces* V *and* W *there is a vector space denoted* $\mathrm{V} \otimes \mathrm{W}$ *together with a bilinear map* $u : \mathrm{V} \times \mathrm{W} \to \mathrm{V} \otimes \mathrm{W}$ *has the following property:*

Definition 9.1 ***(Universal property)*** *Given any bilinear map* $: \mathrm{V} \times \mathrm{W} \to \mathrm{E}$, *where is another finite dimensional vector space, there is a unique* linear *map* $\widetilde{b} : \mathrm{V} \otimes \mathrm{W} \to \mathrm{E}$ *such that the following diagram commutes:*

$$\begin{array}{ccc} \mathrm{V} \times \mathrm{W} & \xrightarrow{b} & \mathrm{E} \\ u\downarrow & \nearrow_{\widetilde{b}} & \\ \mathrm{V} \otimes \mathrm{W} & & \end{array}$$

[1] The word "consolidated" refers to the fact that we let the contravariant slots and the covariant slots slide past each other so that a consolidated tensor product would never give a multilinear function like $V \times V^* \times V \to \mathbb{R}$. Some authors feel the need to do it otherwise but we shall see how to avoid involving shuffled spaces like $T^{r}{}_{sk}{}^{kl}(\mathrm{V})$ for example.

There is an obvious generalization of this theorem which gives the existence of a universal object for $k-$multilinear maps $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{E}$. Note that the spaces $\mathrm{V}_1, ..., \mathrm{V}_k$ are fixed but the E may vary. The universal space U (together with the associated map $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to U$ is called the tensor product of $\mathrm{V}_1, ..., \mathrm{V}_k$ and is denoted $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.

Now that we have the existence of the tensor product we make the following practical observations:

1. The image of the map $u : \mathrm{V} \times \mathrm{W} \to \mathrm{V} \otimes \mathrm{W}$ is a spanning set for $\mathrm{V} \otimes \mathrm{W}$ so if we denote the image of a pair (v, w) by $\mathrm{v} \otimes \mathrm{w}$ then every element of $\mathrm{V} \otimes \mathrm{W}$ is a linear combination of elements of that form. These special elements are called **simple** or **indecomposable**.

2. Suppose we want to make a linear map from $\mathrm{V} \otimes \mathrm{W}$ to some vector space E. Suppose also that we know what we would like the linear map to do to the simple elements. If the rule we come up with is bilinear in the factors of the simple elements then this rule extends uniquely as a linear map. For example, consider $\mathrm{V} \otimes \mathrm{Hom}(\mathrm{V}, \mathrm{E})$ and the rule $\mathrm{v} \otimes \mathrm{A} \to \mathrm{Av}$. This rule is enough to determine a unique linear map $\mathrm{V} \otimes \mathrm{Hom}(\mathrm{V}, \mathrm{E}) \to \mathrm{E}$ which agrees with the prescription on simple elements (the set of simple elements which is probably not even a vector space). We will call this the extension principal.

3. If $\{\mathrm{v}_1, ..., \mathrm{v}_m\}$ is a basis of V and $\{\mathrm{w}_1, ..., \mathrm{w}_n\}$ a basis of W, then a basis of $\mathrm{V} \otimes \mathrm{W}$ is the doubly indexed set $\{\mathrm{v}_i \otimes \mathrm{w}_j : 1 \leq i \leq n,\ 1 \leq j \leq m\}$. Thus a typical element $\Upsilon \in \mathrm{V} \otimes \mathrm{W}$ has the expansion $\Upsilon = \sum a^{ij} \mathrm{v}_i \otimes \mathrm{w}_j$ for some real numbers a^{ij}. Similarly, for $\Upsilon \in \mathrm{V} \otimes \mathrm{V} \otimes \mathrm{V}^*$ we have

$$\Upsilon = \sum a^{\ ij}{}_{k}\, \mathrm{v}_i \otimes \mathrm{v}_j \otimes v^j$$

where $\{v^1, ..., v^n\}$ is the basis for V^* which is dual[2] to $\{\mathrm{v}_1, ..., \mathrm{v}_m\}$. In general, a basis for $\mathrm{V} \otimes \cdots \mathrm{V} \otimes \mathrm{V}^*$

Exercise 9.1 *Show that the definition of tensor product implies that the simple elements span all of* $\mathrm{V} \otimes \mathrm{W}$. *The uniqueness part of the definition is key here.*

Exercise 9.2 *Show that the extension principle follows directly from the definition of the tensor product and is essentially just a reformulation of the defining properties.*

So now we need to think about a few things. For instance, what is the relation between say $T^1\big|_1(\mathrm{V}) = L(\mathrm{V}^*, \mathrm{V})$ and $\mathrm{V}^* \otimes \mathrm{V}$ or $\mathrm{V} \otimes \mathrm{V}^*$. What about $\mathrm{V}^* \otimes \mathrm{V}$ and $\mathrm{V} \otimes \mathrm{V}^*$. Are they equal? Isomorphic? There is also the often employed relationship between $\mathrm{W} \otimes \mathrm{V}^*$ and $L(\mathrm{V}, \mathrm{W})$ to wonder about. Let us take a quick tour of some basis relations.

[2] So we have $v^i(\mathrm{v}_j) = \delta^i_j$.

1.
$$\mathrm{V} \cong \mathrm{V}^{**}$$

This one is a natural inclusion given by $\mathrm{v} \mapsto (\widetilde{\mathrm{v}} : V^* \to \mathbb{R})$ where $\widetilde{\mathrm{v}} : \alpha {\mapsto} \alpha(\mathrm{v})$. For finite dimensional vector spaces this is an isomorphism. This is also true by definition for reflexive Banach spaces and only in this case will we be able to safely say that $T^1\big|_0(\mathrm{V}) = \mathrm{V}$ unless we simply modify definitions (some authors do just that).

2.
$$L(\mathrm{V},\mathrm{W};\mathbb{R}) = L(\mathrm{V}\otimes\mathrm{W},\mathbb{R}) = (V \otimes W)^*$$

This one is true by the definition of tensor product $\mathrm{V}\otimes\mathrm{W}$ given above. It says that the bilinear maps $\mathrm{V}\times\mathrm{W} \to \mathbb{R}$ are in 1-1 correspondence with linear maps on a special space called the tensor product of V and W. The obvious generalization of this one is $L(\mathrm{V}_1, ..., \mathrm{V}_r; \mathrm{E}) = L(\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_r, \mathrm{E})$. Also, a special case worth keeping in mind is $L(\mathrm{V}^*, \mathrm{V}; \mathbb{R}) = (\mathrm{V}^* \otimes \mathrm{V})^*$.

3.
$$\mathrm{V}\otimes\mathrm{V}^* \cong L(\mathrm{V}^*, \mathrm{V}; \mathbb{R})$$

This one is curious. It is another natural isomorphism that does not generalize well to infinite dimensions. The isomorphism goes like this: Let $\mathrm{v}\otimes\alpha$ be the image of $(\mathrm{v},\alpha) \in \mathrm{V}\times\mathrm{V}^*$ under the universal maps as described above. Elements of this form generate all of $\mathrm{V}\otimes\mathrm{V}^*$ and specifying where these elements should go under a linear map actually determines the linear map as we have seen. In the current case we use $\mathrm{v}\otimes\alpha$ to define an element $\iota(\mathrm{v}\otimes\alpha) \in L(\mathrm{V}^*, \mathrm{V}; \mathbb{R})$:

$$\iota(\mathrm{v}\otimes\alpha)(\beta,\mathrm{w}) := \beta(\mathrm{v})\alpha(\mathrm{w}).$$

This isomorphism is a favorite of differential geometers but if one plans to do differential geometry on general Banach spaces the isomorphism is not available. Sometimes one can recover something like this isomorphism but it takes some work and by then it is realized how small the role of this actually was after all. One can often do quite well without it.
There is something crucial to observe here. Namely, under the identification of $\mathrm{V}^{**}$with V, the isomorphism above just sends $\mathrm{v}\otimes\alpha \in \mathrm{V}\otimes\mathrm{V}^*$ to the tensor product of the two maps $\mathrm{v} : \mathrm{V}^* {\to} \mathbb{R}$ and $\alpha : \mathrm{V} {\to} \mathbb{R}$ which is also written $\mathrm{v}\otimes\alpha$ the result being in $T^1\big|_1(\mathrm{V}) = L(\mathrm{V}^*, \mathrm{V}; \mathbb{R})$. The point is we already had a meaning for the symbol $\otimes$ in the context of multiplying multilinear maps. Here we have just seen that $\mathrm{v}\otimes\alpha \in \mathrm{V}\otimes\mathrm{V}^*$ interpreted properly just is that product. Thus we will actually identify $\mathrm{v}\otimes\alpha$ with $\iota(\mathrm{v}\otimes\alpha)$ and drop the ι completely. Conclusion: $\mathrm{v}\otimes\alpha$ is a multilinear map in either case.

4.
$$\mathrm{W}\otimes\mathrm{V}^* \cong L(\mathrm{V},\mathrm{W})$$

This last natural isomorphism is defined on simple elements using the formula
$$\iota(\mathrm{v}\otimes\alpha)(\beta)\cdot\mathrm{w} := \alpha(\mathrm{w})\mathrm{v}$$
which then extends to all elements of $\mathrm{V}\otimes\mathrm{W}^*$ by the extension principal described above. For reasons similar to before, the map $\iota(\mathrm{v}\otimes\alpha)$ will be identified with $\mathrm{v}\otimes\alpha$ and now $\mathrm{v}\otimes\alpha$ has yet another interpretation. Tensors generally have several interpretations as multilinear maps. For example, if $\Upsilon : \mathrm{V}\times\mathrm{W}\times\mathrm{W}^* \to \mathbb{R}$ is a multilinear map into $\mathbb{R}$ then $\Upsilon : \mathrm{v}\mapsto\Upsilon(\mathrm{v},.,.)$ defined a multilinear map $\mathrm{V}\to L(\mathrm{W},\mathrm{W}^*)$.

Exercise 9.3 *Let* V *and* W *be finite dimensional. Prove that* $(\mathrm{V}\otimes\mathrm{W})^*$ *is naturally isomorphic to* $\mathrm{V}^*\otimes\mathrm{W}^*$*. Note that what makes this work is the fact that* V^{**} *may be identified with* V*.*

Exercise 9.4 *Prove that for finite dimensional vector spaces* $\mathrm{V}\otimes\mathrm{V}^* \hookrightarrow L(\mathrm{V}^*,\mathrm{V};\mathbb{R})$ *described above is an isomorphism.*

9.1.1 Contraction of tensors

Consider a tensor of the form $\mathrm{v}_1\otimes\mathrm{v}_2\otimes\eta^1\otimes\eta^2 \in T^2_2(\mathrm{V})$ we can define the $1,1$ contraction of $\mathrm{v}_1\otimes\mathrm{v}_2\otimes\eta^1\otimes\eta^2$ as the tensor obtained as
$$C^1_1(\mathrm{v}_1\otimes\mathrm{v}_2\otimes\eta^1\otimes\eta^2) = \eta^1(\mathrm{v}_1)\mathrm{v}_2\otimes\eta^2.$$
Similarly we can define
$$C^1_2(\mathrm{v}_1\otimes\mathrm{v}_2\otimes\eta^1\otimes\eta^2) = \eta^2(\mathrm{v}_1)\mathrm{v}_2\otimes\eta^1.$$
In general, we could define C^i_j on "monomials" $\mathrm{v}_1\otimes\mathrm{v}_2 ...\otimes\mathrm{v}_r\otimes\eta^1\otimes\eta^2\otimes...\otimes\eta^s$ and then extend linearly to all of $T^r_s(\mathrm{V})$. This works fine for V finite dimensional and turns out to give a notion of contraction which is the same a described in the next definition.

Definition 9.2 *Let* $\{e_1,...,e_n\}\subset\mathrm{V}$ *be a basis for* V *and* $\{e^1,...,e^n\}\subset\mathrm{V}^*$ *the dual basis. If* $\tau\in T^r_s(\mathrm{V})$ *we define* $C^i_j\tau\in T^{r-1}_{s-1}(\mathrm{V})$
$$\begin{aligned} &C^i_j\tau(\theta^1,...,\theta^{r-1},\mathrm{w}_1,..,\mathrm{w}_{s-1}) \\ &=\sum_{k=1}^n \tau(\theta^1,...,\underset{i-th}{e^k},\dots,\theta^{r-1},\mathrm{w}_1,\dots,\underset{j-th}{e_k},\dots,\mathrm{w}_{s-1}).\end{aligned}$$

It is easily checked that this definition is independent of the basis chosen. In the infinite dimensional case the sum contraction cannot be defined in general to apply to all tensors. However, we can still define contractions on linear combinations of tensors of the form $\mathrm{v}_1\otimes\mathrm{v}_2...\otimes\mathrm{v}_r\otimes\eta^1\otimes\eta^2\otimes...\otimes\eta^s$ as we did above. Returning to the finite dimensional case, suppose that
$$\tau = \sum \tau^{i_1,\dots,i_r}_{j_1,\dots,j_s} e_{i_1}\otimes...\otimes e_{i_r}\otimes e^{j_1}\otimes...\otimes e^{j_s}.$$

Then it is easy to check that if we define

$$\tau^{i_1,\ldots,i_{r-1}}_{j_1,\ldots,j_{s-1}} = \sum_{k=1}^{n} \tau^{i_1,\ldots,k,\ldots,i_{r-1}}_{j_1,\ldots,k,\ldots,j_{s-1}}$$

where the upper repeated index k is in the i-th position and the lower occurrence of k is in the j-th position then

$$C^i_j \tau = \sum \tau^{i_1,\ldots,i_{r-1}}_{j_1,\ldots,j_{s-1}} e_{i_1} \otimes \ldots \otimes e_{i_{r-1}} e^{j_1} \otimes \ldots \otimes e^{j_{s-1}}.$$

Even in the infinite dimensional case the following definition makes sense. The contraction process can be repeated until we arrive at a function.

Definition 9.3 *Given $\tau \in T^r_s$ and $\sigma = \mathrm{w}_1 \otimes \ldots \otimes \mathrm{w}_l \otimes \eta^1 \otimes \eta^2 \otimes \ldots \otimes \eta^m \in T^l_m$ a simple tensor with $l \le r$ and $m \le s$, we define the contraction against σ by*

$$\begin{aligned} &\sigma \lrcorner \tau(\alpha_1, \ldots, \alpha_{r-l}, \mathrm{v}_1, \ldots, \mathrm{v}_{r-l}) \\ &:= C(\tau \otimes (\mathrm{w}_1 \otimes \ldots \otimes \mathrm{w}_l \otimes \eta^1 \otimes \eta^2 \otimes \ldots \otimes \eta^m)) \\ &:= \tau(\eta^1, \ldots, \eta^m, \alpha_1, \ldots, \alpha_{r-l}, \mathrm{w}_1, \ldots, \mathrm{w}_l, \mathrm{v}_1, \ldots, \mathrm{v}_{r-l}). \end{aligned}$$

For a given simple tensor σ we thus have a linear map $\sigma \lrcorner : T^r_s \to T^{r-l}_{s-m}$. For finite dimensional V this can be extended to a bilinear pairing between T^l_m and T^r_s

$$T^l_m(\mathrm{V}) \times T^r_s(\mathrm{V}) \to T^{r-l}_{s-m}(\mathrm{V}).$$

Exercise 9.5 *Show that all the different interpretations of a tensor may be realized by tensoring with a some number of variable tensors and then contracting. Hint: What would the following signify?*

$$\Upsilon : (v_l, a^i) \mapsto v_l \; \Upsilon^{lk}\big|_{ij} \, a^i \text{ (sum over } i \text{ and } l\text{)}$$

Extended Matrix Notation. Notice that if one takes the convention that objects with indices up are "column vectors" and indices down "row vectors" then to get the order of the matrices correctly the repeated indices should read *down then up* going from left to right. So $A^I_J e_I e^J$ should be changed to $e_I A^I_J e^J$ before it can be interpreted as a matrix multiplication.

Remark 9.1 *We can also write $\triangle^{K'}_I A^I_J \triangle^J_{L'} = A^{K'}_{L'}$ where $\triangle^R_{S'} = e^R e'_S$.*

9.1.2 Alternating Multilinear Algebra

In this section we make the simplifying assumption that all of the rings we use will have the following property: The sum of the unity element with itself; $1+1$ is invertible. Thus if we use 2 to denote the element $1+1$ then we assume the existence of a unique element "1/2" such that $2 \cdot 1/2 = 1$. Thus, in the case of fields, the assumption is that the field is not of characteristic 2. The reader need only worry about two cases:

1. The unity "1" is just the number 1 in some subring of $\mathbb{C}$ (e.g. $\mathbb{R}$ or $\mathbb{Z}$) or
2. the unity "1" refers to some sort of function or section with values in a ring like $\mathbb{C}$, $\mathbb{R}$ or $\mathbb{Z}$ which takes on the constant value 1. For example, in the ring $C^\infty(M)$, the unity is just the constant function 1.

Definition 9.4 *A $\mathbb{Z}$-graded algebra is called skew-commutative (or graded commutative) if for $a_i \in A_i$ and $a_j \in A_j$ we have*

$$a_i \cdot a_j = (-1)^{kl} a_j \cdot a_i$$

Definition 9.5 *A morphism of degree n from a graded algebra $A = \bigoplus_{i\in\mathbb{Z}} A_i$ to a graded algebra $B = \bigoplus_{i\in\mathbb{Z}} B_i$ is a algebra homomorphism $h : A \to B$ such that $h(A_i) \subset B_{i+n}$.*

Definition 9.6 *A super algebra is a $\mathbb{Z}_2$-graded algebra $A = A_0 \oplus A_1$ such that $A_i \cdot A_j \subset A_{i+j \ mod2}$ and such that $a_i \cdot a_j = (-1)^{kl} a_j \cdot a_i$ for $i, j \in \mathbb{Z}_2$.*

Alternating tensor maps

Definition 9.7 *A k-multilinear map $\alpha : \mathrm{W} \times \cdots \times \mathrm{W} \to \mathrm{F}$ is called **alternating** if $\alpha(\mathrm{w}_1, ..., \mathrm{w}_k) = 0$ whenever $\mathrm{w}_i = \mathrm{w}_j$ for some $i \neq j$. The space of all **alternating** k-multilinear maps into F will be denoted by $L^k_{alt}(\mathrm{W};\mathrm{F})$ or by $L^k_{alt}(\mathrm{W})$ if the ring is either $\mathbb{R}$ or $\mathbb{C}$ and there is no chance of confusion.*

Remark 9.2 *Notice that we have moved the k up to make room for the Alt thus $L^k_{alt}(\mathrm{W};\mathsf{R}) \subset L^0_k(\mathrm{W};\mathsf{R})$.*

Thus if $\omega \in L^k_{alt}(\mathrm{V})$, then for any permutation σ of the letters $1, 2, ..., k$ we have

$$\omega(\mathrm{w}_1, \mathrm{w}_2, .., \mathrm{w}_k) = sgn(\sigma)\omega(\mathrm{w}_{\sigma_1}, \mathrm{w}_{\sigma_2}, .., \mathrm{w}_{\sigma_k}).$$

Now given $\omega \in L^r_{alt}(\mathrm{V})$ and $\eta \in L^s_{alt}(\mathrm{V})$ we define their **wedge product** or **exterior product** $\omega \wedge \eta \in L^{r+s}_{alt}(\mathrm{V})$ by the formula

$$\omega \wedge \eta(\mathrm{v}_1, ..., \mathrm{v}_r, \mathrm{v}_{r+1}, ..., \mathrm{v}_{r+s}) := \frac{1}{r!s!} \sum_\sigma \mathrm{sgn}(\sigma)\omega(\mathrm{v}_{\sigma_1}, ..., \mathrm{v}_{\sigma_r})\eta(\mathrm{v}_{\sigma_{r+1}}, ..., \mathrm{v}_{\sigma_{r+s}})$$

or by

$$\omega \wedge \eta(\text{“same as above”}) := \sum_{r,s-\text{shuffles } \sigma} \mathrm{sgn}(\sigma)\omega(\mathrm{v}_{\sigma_1}, ..., \mathrm{v}_{\sigma_r})\eta(\mathrm{v}_{\sigma_{r+1}}, ..., \mathrm{v}_{\sigma_{r+s}}).$$

In the latter formula we sum over all permutations such that $\sigma_1 < \sigma_2 < .. < \sigma_r$ and $\sigma_{r+1} < \sigma_{r+2} < .. < \sigma_{r+s}$. This kind of permutation is called an $r, s-$shuffle as indicated in the summation. The most important case is for $\omega, \eta \in L^1_{alt}(\mathrm{V})$ in which case

$$(\omega \wedge \eta)(v, w) = \omega(v)\eta(w) - \omega(w)\eta(v)$$

This is clearly a skew symmetric multi-linear map.

If we use a basis $\varepsilon^1, \varepsilon^2, ..., \varepsilon^n$ for V^* it is easy to show that the set of all elements of the form $\varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$ with $i_1 < i_2 < ... < i_k$ form a basis for . Thus for any $\omega \in A^k(\mathrm{V})$

$$\omega = \sum_{i_1 < i_2 < ... < i_k} a_{i_1 i_2, .., i_k} \varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$$

Remark 9.3 *In order to facilitate notation we will abbreviate as sequence of k-integers $i_1, i_2, ..., i_k$ from the set $\{1, 2, ..., \dim(\mathrm{V})\}$ as I and $\varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$ is written as ε^I. Also, if we require that $i_1 < i_2 < ... < i_k$ we will write $\vec{I}$. We will freely use similar self explanatory notation as we go along with out further comment. For example, the above equation can be written as*

$$\omega = \sum a_{\vec{I}} \varepsilon^{\vec{I}}$$

Lemma 9.1 $L^k_{alt}(\mathrm{V}) = 0$ *if* $k > n = \dim(\mathrm{V})$.

Proof. Easy exercise. ■

If one defines $L^0_{alt}(\mathrm{V})$ to be the scalars $\mathbb{K}$ and recalling that $L^1_{alt}(\mathrm{V}) = \mathrm{V}^*$ then the sum

$$L_{alt}(\mathrm{V}) = \bigoplus_{k=0}^{\dim(M)} L^k_{alt}(\mathrm{V})$$

is made into an algebra via the wedge product just defined.

Proposition 9.1 *For $\omega \in L^r_{alt}(V)$ and $\eta \in L^s_{alt}(V)$ we have $\omega \wedge \eta = (-1)^{rs} \eta \wedge \omega \in L^{r+s}_{alt}(V)$.*

The Abstract Grassmann Algebra

We wish to construct a space that is universal with respect to alternating multilinear maps. To this end, consider the tensor space $T^k(\mathrm{W}) := \mathrm{W}^{k\otimes}$ and let A be the submodule of $T^k(\mathrm{W})$ generated by elements of the form

$$\mathrm{w}_1 \otimes \cdots \mathrm{w}_i \otimes \cdots \otimes \mathrm{w}_i \cdots \otimes \mathrm{w}_k.$$

In other words, A is generated by decomposable tensors with two (or more) equal factors. We define the space of k-vectors to be

$$\mathrm{W} \wedge \cdots \wedge \mathrm{W} := \bigwedge^k \mathrm{W} := T^k(\mathrm{W})/\mathrm{A}.$$

Let $\mathrm{A}_k : \mathrm{W} \times \cdots \times \mathrm{W} \to T^k(\mathrm{W}) \to \not\supseteq^k \mathrm{W}$ be the canonical map composed with projection onto $\bigwedge^k \mathrm{W}$. This map turns out to be an alternating multilinear map. We will denote $\mathrm{A}_k(\mathrm{w}_1, ..., \mathrm{w}_k)$ by $\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_k$. The pair $(\bigwedge^k \mathrm{W}, \mathrm{A}_k)$ is universal with respect to alternating k-multilinear maps: Given any alternating

k-multilinear map $\alpha : \mathrm{W} \times \cdots \times \mathrm{W} \to \mathrm{F}$, there is a unique linear map $\alpha_\wedge : \bigwedge^k \mathrm{W} \to \mathrm{F}$ such that $\alpha = \alpha_\wedge \circ \mathrm{A}_k$; that is$\bigwedge^k$

$$\begin{array}{ccc} \mathrm{W} \times \cdots \times \mathrm{W} & \xrightarrow{\alpha} & \mathrm{F} \\ \mathrm{A}_k \downarrow & \nearrow_{\alpha_\wedge} & \\ \bigwedge^k \mathrm{W} & & \end{array}$$

commutes. Notice that we also have that $\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_k$ is the image of $\mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k$ under the quotient map. Next we define $\bigwedge \mathrm{W} := \sum_{k=0}^{\infty} \bigwedge^k \mathrm{W}$ and impose the multiplication generated by the rule

$$(\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_i) \times (\mathrm{w}'_1 \wedge \cdots \wedge \mathrm{w}'_j) \mapsto \mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_i \wedge \mathrm{w}'_1 \wedge \cdots \wedge \mathrm{w}'_j \in \bigwedge^{i+j} \mathrm{W}.$$

The resulting algebra is called the Grassmann algebra or exterior algebra. If we need to have a $\mathbb{Z}$ grading rather than a $\mathbb{Z}^+$ grading we may define $\bigwedge^k \mathrm{W} := 0$ for $k < 0$ and extend the multiplication in the obvious way.

Notice that since $(\mathrm{w} + \mathrm{v}) \wedge (\mathrm{w} + \mathrm{v}) = 0$, it follows that $\mathrm{w} \wedge \mathrm{v} = -\mathrm{v} \wedge \mathrm{w}$. In fact, any odd permutation of the factors in a decomposable element such as $\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_k$, introduces a change of sign:

$$\begin{aligned} &\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_i \wedge \cdots \wedge \mathrm{w}_j \wedge \cdots \wedge \mathrm{w}_k \\ &= -\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_j \wedge \cdots \wedge \mathrm{w}_i \wedge \cdots \wedge \mathrm{w}_k \end{aligned}$$

Just to make things perfectly clear, we exhibit a short random calculation where the ring is the real numbers $\mathbb{R}$:

$$\begin{aligned} &(2(1 + 2\mathrm{w} + \mathrm{w} \wedge \mathrm{v} + \mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u}) + \mathrm{v} \wedge \mathrm{u}) \wedge (\mathrm{u} \wedge \mathrm{v}) \\ &= (2 + 4\mathrm{w} + 2\mathrm{w} \wedge \mathrm{v} + 2\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} + \mathrm{v} \wedge \mathrm{u}) \wedge (-\mathrm{v} \wedge \mathrm{u}) \\ &= -2\mathrm{v} \wedge \mathrm{u} - 4\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} - 2\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{v} \wedge \mathrm{u} - 2\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} \wedge \mathrm{v} \wedge \mathrm{u} - \mathrm{v} \wedge \mathrm{u} \wedge \mathrm{v} \wedge \mathrm{u} \\ &= -2\mathrm{v} \wedge \mathrm{u} - 4\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} + 0 + 0 = -2\mathrm{v} \wedge \mathrm{u} - 4\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u}. \end{aligned}$$

Lemma 9.2 *If* V *is has rank n, then $\bigwedge^k \mathrm{V} = 0$ for $k \geq n$. If $f_1, ..., f_n$ is a basis for* V *then the set*

$$\{f_{i_1} \wedge \cdots \wedge f_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$$

is a basis for $\bigwedge^k \mathrm{V}$ where we agree that $f_{i_1} \wedge \cdots \wedge f_{i_k} = 1$ if $k = 0$.

The following lemma follows easily from the universal property of $\alpha_\wedge : \bigwedge^k \mathrm{W} \to \mathrm{F}$:

Lemma 9.3 *There is a natural isomorphism*

$$L^k_{alt}(\mathrm{W}; \mathrm{F}) \cong L(\bigwedge^k \mathrm{W}; \mathrm{F})$$

In particular,

$$L^k_{alt}(\mathrm{W}) \cong \left(\bigwedge^k \mathrm{W}\right)^*.$$

Remark 9.4 (Convention) *Let $\alpha \in L^k_{alt}(\mathrm{W};\mathrm{F})$. Because the above isomorphism is so natural it may be taken as an identification and so we sometimes write $\alpha(v_1, ..., v_k)$ as $\alpha(v_1 \wedge \cdots \wedge v_k)$.*

In the finite dimensional case, we have module isomorphisms

$$\bigwedge^k \mathrm{W}^* \cong \left(\bigwedge^k \mathrm{W}\right)^* \cong L^k_{alt}(\mathrm{W})$$

which extends by direct sum to

$$L_{alt}(\mathrm{W}) \cong \bigwedge \mathrm{W}^*$$

which is in fact an algebra isomorphism (we have an exterior product defined for both).

The following table summarizes:

Exterior Products	Isomorphisms that hold in finite dimension	Alternating multilinear maps
$\bigwedge^k \mathrm{W}$	$\downarrow$	
$\bigwedge^k \mathrm{W}^* \cong \left(\bigwedge^k \mathrm{W}\right)^*$	$\cong$	$L^k_{alt}(\mathrm{W})$
$\bigwedge \mathrm{W} = \bigoplus_k \bigwedge^k \mathrm{W}$		
$\bigwedge \mathrm{W}^* = \bigoplus_k \bigwedge^k \mathrm{W}^*$	graded algebra iso. $\cong$	$A(\mathrm{W}) = \bigoplus_k L^k_{alt}(\mathrm{W})$

9.1.3 Orientation on vector spaces

Let V be a finite dimensional vector space. The set of all ordered bases fall into two classes called orientation classes.

Definition 9.8 *Two bases are in the same orientation class if the change of basis matrix from one to the other has positive determinant.*

That is, given two frames (bases) in the same class, say $(f_1, ...f_n)$ and $(\widetilde{f}_1, ...\widetilde{f}_n)$ with

$$\widetilde{f}_i = f_j C^j_i$$

then $\det C > 0$ and we say that the frames determine the same orientation. The relation is easily seen to be an equivalence relation.

Definition 9.9 *A choice of one of the two orientation classes of frames for a finite dimensional vector space* V *is called an* ***orientation*** *on* V*. The vector space in then said to be oriented.*

Exercise 9.6 *Two frames, say $(f_1, ..., f_n)$ and $(\widetilde{f}_1, ..., \widetilde{f}_n)$ determine the same orientation on* V *if and only if $f_1 \wedge ... \wedge f_n = a\ \widetilde{f}_1 \wedge ... \wedge \widetilde{f}_n$ for some positive real number $a > 0$.*

Exercise 9.7 *If σ is a permutation on n letters $\{1, 2, ...n\}$ then $(f_{\sigma 1}, ..., f_{\sigma n})$ determine the same orientation if and only if $sgn(\sigma) = +1$.*

A top form $\omega \in L^n_{alt}(\mathrm{V})$ determines an orientation on V by the rule $(f_1, ..., f_n) \sim (\widetilde{f}_1, ..., \widetilde{f}_n)$ if and only if

$$\omega(f_1, ..., f_n) = \omega(\widetilde{f}_1, ..., \widetilde{f}_n).$$

Furthermore, two top forms $\omega_1, \omega_2 \in L^n_{alt}(\mathrm{V})$ determine the same orientation on V if and only if $\omega_1 = a\omega_2$ for some positive real number $a > 0$.

9.2 Multilinear Bundles

Using the construction of Theorem **??** we can build vector bundles from a given vector bundle in a way that globalizes the usual constructions of linear and multilinear algebra. Let us first consider the case where the bundle has finite rank. Let $E \to M$ be a vector bundle with a transition functions $\Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathbb{R}^n)$. We can construct vector bundles with typical fibers $\mathbb{R}^{n*}$, $L(\mathbb{R}^n, \mathbb{R}^n)$, $T^r_s(\mathbb{R}^n), \bigwedge^k \mathbb{R}^n$, and $\bigwedge^k \mathbb{R}^{n*}$. All we have to do is choose the correct transition functions. They are the following:

1. For $\mathbb{R}^{n*}$ use the maps $\Phi^*_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathbb{R}^{n*})$ where $\Phi^*_{\alpha\beta}(x) = (\Phi_{\alpha\beta}(x)^{-1})^t$.
2. For $L(\mathbb{R}^n, \mathbb{R}^n) = \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ use the maps $\mathrm{Hom}(\Phi_{\alpha\beta}) : U_\alpha \cap U_\beta \to \mathrm{GL}(L(\mathbb{R}^n, \mathbb{R}^n))$ given by
$$(\mathrm{Hom}(\Phi_{\alpha\beta})(x))(A) = \Phi_{\alpha\beta}(x) \circ A \circ \Phi_{\alpha\beta}(x)^{-1}$$
3. For $T^r_s(\mathbb{R}^n)$ use $T^r_s\Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(T^r_s(\mathbb{R}^n))$ given on simple tensor(?)s by
$$T^r_s\Phi_{\alpha\beta}(x)(v_1 \otimes \cdots \otimes \alpha_s) = \Phi_{\alpha\beta}(x)v_1 \otimes \cdots \otimes \Phi^*_{\alpha\beta}(x)\alpha_s.$$
4. For $\bigwedge^k \mathbb{R}^n$ use $\bigwedge^k \Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\bigwedge^k(\mathbb{R}^n))$ given on homogeneous e(?)lements by
$$\bigwedge^k \Phi_{\alpha\beta}(x)(v_1 \wedge \cdots \wedge v_k) = \Phi_{\alpha\beta}(x)v_1 \wedge \cdots \wedge \Phi_{\alpha\beta}(x)v_k.$$

   In this way we construct bundles which we will denote variously by $E^* \to M$, $\mathrm{Hom}(E, E) \to M$, $T^r_s(E) \to M$, $\bigwedge^k E \to M$ and so on.

9.3 Tensor Fields

The space of tensors of type $\binom{r}{\ s}$ at a point $x \in M$ is the just the tensor space $T^r{}_s(T_pM)$ based on T_pM (and T^*_pM) as discussed in the last section. In case

M is finite dimensional we may make the expected identification via the natural isomorphisms:

$$T^r{}_s(T_pM) = (T_pM^{\otimes r}) \otimes (T_p^*M^{\otimes s}) \cong \bigotimes{}^r{}_s(T_pM) = (T_pM^{\otimes r}) \otimes (T_p^*M^{\otimes s})$$

In a local coordinate frame $(x^1, ..., x^n)$ we have an expression at p for elements of $T^r_s(T_pM)$, say A_p, of the form

$$A_p = \sum A^{i_1,...,i_r}{}_{j_1,...,j_r} \left.\frac{\partial}{\partial x^{i_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x^{i_r}}\right|_p \otimes dx^{j_1}\big|_p \otimes \cdots \otimes dx^{j_s}\big|_p$$

which will abbreviated to

$$A(p) = \sum A^I{}_J(p)\partial x_I \otimes dx^J \text{ at } p$$

A assignment of an element $A_p \in T^r_s(T_pM)$ for each $p \in U \subset M$ is called **tensor field** over U and we will normally require that the assignment be made smoothly in the sense determined by the next two definitions.

Definition 9.10 *Let the disjoint union of all the spaces $T^r_s(T_pM)$ over the points in M be denoted by $T^r_s(TM)$. This set can be given structure of a smooth vector bundle in a manner similar to the structure on TM and is called the (r, s)-* ***tensor bundle****. (Recall the discussion in the section 9.2 concerning operations on vector bundles.) In case M is finite dimensional, we also have the bundle $\bigotimes^r_s(TM)$ and the natural bundle isomorphism*

$$\bigotimes{}^r_s(TM) \cong T^r_s(TM).$$

Definition 9.11 *A (smooth)* ***tensor field*** *is a section of the vector bundle $T^r_s(TM) \to M$. The set of all smooth tensor fields is denoted $\mathfrak{T}^r_s(M)$. We can similarly talk of tensor fields defined on an open subset U of M and so we have the space $\mathfrak{T}^r_s(U)$ of tensor fields $\mathfrak{T}^r_s(M)$ over U.*

The assignment $U \mapsto \mathfrak{T}^r_s(U)$ is a sheaf of modules over $\mathcal{C}^\infty$. We also have the alternative notation $\mathfrak{X}_M(U)$ for the case of $r = 0, s = 1$.

Proposition 9.2 *Let M be an n-dimensional smooth manifold. A map $A : U \to T^r_s(M)$ of the form $x \mapsto A(p) = A_p \in T^r_s(T_pM)$ is a smooth tensor field in U if for all coordinate systems $(x^1, ..., x^n)$ defined on open subsets of U the local expression*

$$A = \sum A^I_J \partial x_I \otimes dx^J$$

is such that the functions A^I_J are smooth.

Now one may wonder what would happen if we considered the space $\mathfrak{X}(U)$ as a module over the ring of smooth functions $C^\infty(U)$ and then took tensors products over this ring of r copies of $\mathfrak{X}_M(U)$ and s copies of $\mathfrak{X}_M(U)^*$ to get

$$\mathfrak{X}_M(U) \otimes \cdots \otimes \mathfrak{X}_M(U) \otimes \mathfrak{X}_M(U)^* \otimes \cdots \otimes \mathfrak{X}_M(U)^*. \tag{9.1}$$

Of course, this space is equivalent to a space of multilinear module morphisms but furthermore this turns out to be naturally equivalent to the space of sections $\mathfrak{T}^r_s(U)$ if the manifold is finite dimensional. The proof is an extension of the proof of Theorem 7.2.

Definition 9.12 *We can define contraction of tensor fields by contracting on each tangent space and if the manifold is infinite dimensional then not all tensors can be contracted.*

In any case, we can define contraction against simple tensors. For example, let $\Upsilon \in \mathfrak{T}^2_3(U)$, $\alpha \in \mathfrak{X}_M(U)^*$ and $X \in \mathfrak{X}_M(U)$. Then $X \otimes \alpha \in \mathfrak{T}^1_1(U)$ and we have $(X \otimes \alpha) \lrcorner \Upsilon \in \mathfrak{T}^1_2(U)$ given by

$$((X \otimes \alpha) \lrcorner \Upsilon)(\alpha_1, X_1, X_2) = C(\Upsilon \otimes (X \otimes \alpha)) = \Upsilon(\alpha, \alpha_1, X, X_1, X_2)$$

Definition 9.13 *A $k-$covariant tensor field Υ is called symmetric of for all k-tuples of vector fields $(X_1, ..., X_k)$ and all permutations of k-letters σ we have*

$$\Upsilon(X_{\sigma 1}, ..., X_{\sigma k}) = \Upsilon(X_1, ..., X_k).$$

We can similarly define symmetric contravariant tensors.

Proposition 9.3 *Let $\{U_i, \psi_i\}$ be an atlas for a smooth (or C^r) manifold M with $\psi_i : U_i \to \psi_i(U_i)$. Suppose that we have a smooth (or C^r) tensor field Υ_i of type (r, s) defined on each U_i and that $\{U_i, \rho_i\}$ is a smooth partition of unity. Then the following exists a smooth (or C^r) tensor field Υ of type (r, s) defined on M Furthermore if $\Upsilon_i > 0$ for each i, then we can construct Υ of the same type as Υ_i such that $\Upsilon > 0$.*

Proof. Define a global tensor field for each i by extending $\rho_i \Upsilon_i$ by zero (why is this smooth?). Now define

$$\Upsilon = \sum_j \rho_j \Upsilon_j.$$

This is well defined and smooth at each point of M since the sum is finite in some neighborhood of every point. Also, if $\Upsilon_i > \Omega_i$ for each i then if $x \in U_i$ we have that $\sum_j \rho_j(x) = 1$ (a finite sum) so $\rho_k(x) > 0$ for some k. Thus

$$\Upsilon(x) = \sum_j \rho_j(x)\Upsilon_j(x) \geq \rho_k(x)\Upsilon_k(x) \geq \rho_k(x) > 0.$$

■

9.4 Tensor Derivations

We would like to be able to define derivations of tensor fields . In particular we would like to extend the Lie derivative to tensor fields. For this purpose we introduce the following definition which will be useful not only for extending the Lie derivative but can also be used in several other contexts. Recall the presheaf of tensor fields $U \mapsto \mathfrak{T}^r_s(U)$ on a manifold M.

Definition 9.14 *A differential tensor derivation is a collection of maps* $\mathcal{D}^r_s|_U : \mathfrak{T}^r_s(U) \to \mathfrak{T}^r_s(U)$, *all denoted by* $\mathcal{D}$ *for convenience, such that*

1. $\mathcal{D}$ *is a presheaf map for* $\mathfrak{T}^r_s$ *considered as a presheaf of vector spaces over* $\mathbb{R}$. *In particular, for all open* U *and* V *with* $V \subset U$ *we have*
$$\mathcal{D}\Upsilon|_V = \mathcal{D}\ \Upsilon|_V$$
for all $\Upsilon \in \mathfrak{T}^r_s(U)$.

2. $\mathcal{D}$ *commutes with contractions against simple tensors.*

3. $\mathcal{D}$ *satisfies a derivation law. Specifically, for* $\Upsilon_1 \in \mathfrak{T}^r_s(U)$ *and* $\Upsilon_2 \in \mathfrak{T}^j_k(U)$ *we have*
$$\mathcal{D}(\Upsilon_1 \otimes \Upsilon_2) = \mathcal{D}\Upsilon_1 \otimes \Upsilon_2 + \Upsilon_1 \otimes \mathcal{D}\Upsilon_2.$$

The conditions 2 and 3 imply that for $\Upsilon \in \mathfrak{T}^r_s(U)$, $\alpha_1, ..., \alpha_r \in \mathfrak{X}^*(U)$ and $X_1, ..., X_s \in \mathfrak{X}(U)$ we have

$$\begin{aligned}\mathcal{D}(\Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., X_s)) &= \mathcal{D}\Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., X_s) \\ &+ \sum_i \Upsilon(\alpha_1, ..., \mathcal{D}\alpha_i, ..., \alpha_r, X_1, ..., X_s) \\ &+ \sum_i \Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., ..., \mathcal{D}X_i, ..., X_s).\end{aligned}$$

This is follows by noticing that

$$\Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., X_s) = C(\Upsilon \otimes (\alpha_1 \otimes \cdots \otimes \alpha_r \otimes X_1 \otimes \cdots \otimes X_s))$$

and applying 1 and 2. Also, in the case of finite dimensional manifolds (2) can be replaced by the statement that $\mathcal{D}$ commutes with contractions (why?).

Proposition 9.4 *Let* M *be a finite dimensional manifold and suppose we have a map on global tensors* $\mathcal{D} : \mathfrak{T}^r_s(M) \to \mathfrak{T}^r_s(M)$ *for all* r, s *nonnegative integers such that 2 and 3 above hold for* $U = M$. *Then there is a unique induced tensor derivation which are agrees with* $\mathcal{D}$ *on global sections.*

Proof. We need to define $\mathcal{D} : \mathfrak{T}^r_s(U) \to \mathfrak{T}^r_s(U)$ for arbitrary open U as a derivation. Let δ be a function that vanishes on a neighborhood of V of $p \in U$.

Claim 9.1 *We claim that* $(\mathcal{D}\delta)(p) = 0$.

■

Proof. To see this let β be a bump function equal to 1 on a neighborhood of p and zero outside of V. Then $\delta = (1 - \beta)\delta$ and so

$$\begin{aligned}\mathcal{D}\delta(p) &= \mathcal{D}((1 - \beta)\delta)(p) \\ &= \delta(p)\mathcal{D}(1 - \beta)(p) + (1 - \beta(p))\mathcal{D}\delta(p) = 0\end{aligned}$$

Given $\tau \in \mathfrak{T}^r_s(U)$ let β be a bump function with support in U and equal to 1 on neighborhood of $p \in U$. Then $\beta\tau \in \mathfrak{T}^r_s(M)$ after extending by zero. Now define

$$(\mathcal{D}\tau)(p) = \mathcal{D}(\beta\tau)(p).$$

Now to show this is well defined let β_2 be any other bump function with support in U and equal to 1 on neighborhood of $p_0 \in U$. Then we have

$$\begin{aligned} &\mathcal{D}(\beta\tau)(p_0) - \mathcal{D}(\beta_2\tau)(p_0) \\ &= \mathcal{D}(\beta\tau) - \mathcal{D}(\beta_2\tau))(p_0) = \mathcal{D}((\beta - \beta_2)\tau)(p_0) = 0 \end{aligned}$$

where that last equality follows from our claim above with $\delta = \beta - \beta_2$. Thus $\mathcal{D}$ is well defined on $\mathfrak{T}^r_s(U)$. We now show that $\mathcal{D}\tau$ so defined is an element of $\mathfrak{T}^r_s(U)$. Let ψ_α, U_α be a chart containing p. Let $\psi_\alpha : U_\alpha \to V_\alpha \subset \mathbb{R}^n$. Then we can write $\tau|_{U_\alpha} \in \mathfrak{T}^r_s(U_\alpha)$ as $\tau_{U_\alpha} = \tau^{j_1,\ldots,j_r}_{i_1,\ldots,i_s} dx^{i_1} \otimes \cdots \otimes dx^{i_s} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_r}}$. We can use this to show that $\mathcal{D}\tau$ as defined is equal to a global section in a neighborhood of p and so must be a smooth section itself since the choice of $p \in U$ was arbitrary. To save on notation let us take the case $r = 1$, $s = 1$. Then $\tau_{U_\alpha} = \tau^i_j dx^j \otimes \frac{\partial}{\partial x^i}$. Let β ne a bump function equal to one in a neighborhood of p and zero outside of $U_\alpha \cap U$. Now extend each of the sections $\beta\tau^i_j \in \mathfrak{F}(U)$, $\beta dx^j \in \mathfrak{T}^0_1(U)$ and $\beta\frac{\partial}{\partial x^i} \in \mathfrak{T}^1_0(U)$ to global sections and apply $\mathcal{D}$ to $\beta^3\tau|_{U_\alpha} = \beta\tau^i_j \beta dx^j \otimes \beta\frac{\partial}{\partial x^i}$ to get

$$\begin{aligned} &= \mathcal{D}(\beta^3\tau) = \mathcal{D}(\beta\tau^i_j \beta dx^j \otimes \beta\frac{\partial}{\partial x^i}) \\ &= \mathcal{D}(\beta\tau^i_j)\beta dx^j \otimes \beta\frac{\partial}{\partial x^i} + \beta\tau^i_j \mathcal{D}(\beta dx^j) \otimes \beta\frac{\partial}{\partial x^i} \\ &\quad + \beta\tau^i_j \beta dx^j \otimes \mathcal{D}(\beta\frac{\partial}{\partial x^i}) \end{aligned}$$

Now by assumption $\mathcal{D}$ takes smooth global sections to smooth global sections so both sides of the above equation are smooth. On the other hand, independent of the choice of β we have $\mathcal{D}(\beta^3\tau)(p) = \mathcal{D}(\tau)(p)$ by definition and valid for all p in a neighborhood of p_0. Thus $\mathcal{D}(\tau)$ is smooth and is the restriction of a smooth global section. We leave the proof of the almost obvious fact that this gives a unique derivation $\mathcal{D} : \mathfrak{T}^r_s(U) \to \mathfrak{T}^r_s(U)$ to the reader. ∎

Now one big point that follows from the above considerations is that the action of a tensor derivation on functions, 1-forms and vector fields determines the derivation on the whole tensor algebra. We record this as a theorem.

Theorem 9.2 *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two tensor derivations (so satisfying 1,2, and 3 above) which agree on functions, 1-forms and vector fields. Then $\mathcal{D}_1 = \mathcal{D}_2$. Furthermore, if $\mathcal{D}_U$ can be defined on $\mathfrak{F}(U)$ and $\mathfrak{X}(U)$ for each open $U \subset M$ so that*

1. $\mathcal{D}_U(f \otimes g) = \mathcal{D}_U f \otimes g + f \otimes \mathcal{D}_U g$ *for all* $f, g \in \mathfrak{F}(U)$,

2. *for each* $f \in \mathfrak{F}(M)$ *we have* $(\mathcal{D}_M f)|_U = \mathcal{D}_U \, f|_U$,

3. $\mathcal{D}_U(f \otimes X) = \mathcal{D}_U f \otimes X + f \otimes \mathcal{D}_U X$ *for all* $f \in \mathfrak{F}(U)$ *and* $X \in \mathfrak{X}(U)$,

4. *for each* $X \in \mathfrak{X}(M)$ *we have* $(\mathcal{D}_M X)|_U = \mathcal{D}_U \, X|_U$,

then there is a unique tensor derivation D *on the presheaf of all tensor fields that is equal to* $\mathcal{D}_U$ *on* $\mathfrak{F}(U)$ *and* $\mathfrak{X}(U)$ *for all* U*.*

Sketch of Proof. Define $\mathcal{D}$ on $\mathfrak{X}^*(U)$ by requiring $\mathcal{D}_U(\alpha \otimes X) = \mathcal{D}_U \alpha \otimes X + \alpha \otimes \mathcal{D}_U X$

so that after contraction we see that we must have $(\mathcal{D}_U \alpha)(X) = \mathcal{D}_U(\alpha(X)) - \alpha(\mathcal{D}_U X)$. Now using that $\mathcal{D}$ must satisfy the properties 1,2,3 and 4 and that we wish $\mathcal{D}$ to behave as a derivation we can easily see how $\mathcal{D}$ must act on any simple tensor field and then by linearity on any tensor field. But this prescription can serve as a definition. ■

Corollary 9.1 *The Lie derivative* $\mathcal{L}_X$ *can be extended to a tensor derivation for any* $X \in \mathfrak{X}(M)$*.*

We now present a different way of extending the Lie derivative to tensors that is equivalent to what we have just done. First let $\Upsilon \in \mathfrak{T}^r_s(U)$ If $\phi : U \to \phi(U)$ is a diffeomorphism then we can define $\phi^* \Upsilon \in \mathfrak{T}^r_s(U)$ by

$$
\begin{aligned}
&(\phi^* \Upsilon)(p)(\alpha_1, ..., \alpha_r, v^1, ..., v^s) \\
&= \Upsilon(\phi(p))(T^* \phi^{-1} \cdot \alpha_1, ..., T^* \phi^{-1} \cdot \alpha_r, T\phi \cdot v^1, ..., T\phi \cdot v^s).
\end{aligned}
$$

Now if X is a complete vector field on M we can define

$$
\mathcal{L}_X \Upsilon = \left.\frac{d}{dt}\right|_0 (\mathrm{Fl}_t^{X*} \Upsilon)
$$

just as we did for vector fields. Also, just as before this definition will make sense point by point even if X in not complete.

The Lie derivative on tensor fields is natural in the sense that for any diffeomorphism $\phi : M \to N$ and any vector field X we have

$$
\mathcal{L}_{\phi_* X} \phi_* \tau = \phi_* \mathcal{L}_X \Upsilon.
$$

Exercise 9.8 *Show that the Lie derivative is natural by using the fact that it is natural on functions and vector fields.*

Chapter 10

Differential forms

Let M be a smooth manifold of dimension $n = \dim M$. We form the natural bundle $L^k_{alt}(TM)$ which has as its fiber at p the space $L^k_{alt}(T_pM)$. Let the smooth sections of this bundle be denoted by

$$\Omega^k(M) = \Gamma(M; L^k_{alt}(TM)). \tag{10.1}$$

and sections over $U \subset M$ by $\Omega^k_M(U)$. This space is a module over the ring of smooth functions $C^\infty(M) = \mathcal{F}(U)$. We have the *direct* sum

$$\Omega_M(U) = \textstyle\sum_{n=0}^{\dim M} \Omega^k_M(U) = \Gamma\left(U, \sum_{n=0}^{\dim M} L^k_{alt}(TM)\right).$$

which is a $\mathbb{Z}^+$-graded algebra under the exterior product

$$\begin{aligned}&(\omega \wedge \eta)(p)(v_1, v_2, .., v_r, v_{r+1}, v_{r+2}, .., v_{r+s})\\ &= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma)\omega(p)(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r})\eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})\end{aligned}$$

for $\omega \in \Omega^r_M(U)$ and $\eta \in \Omega^s_M(U)$.

Definition 10.1 *The sections of the bundle $\Omega_M(U)$ are called **differential forms** on U. We identify $\sum_{n=0}^{\dim M} \Omega^k_M(U)$ with the obvious subspace of $\Omega_M(U) = \Omega^k_M(U)$. A differential form in $\Omega^k_M(U)$ is said to be **homogeneous** of **degree** k and is referred to a "k-form".*

Whenever convenient we may extend this to a sum over all $n \in \mathbb{Z}$ by defining (as before) $\Omega^k_M(U) := 0$ for $n < 0$ and $\Omega^k_M(U) := 0$ if $n > \dim(M)$. This is a **$\mathbb{Z}$-graded algebra** under the exterior (wedge) product that is inherited from the exterior product on each fiber $L^k_{alt}(T_pM)$;we have

$$\begin{aligned}&(\omega \wedge \eta)(p)(v_1, v_2, .., v_r, v_{r+1}, v_{r+2}, .., v_{r+s})\\ &= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma)\omega(p)(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r})\eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})\end{aligned}$$

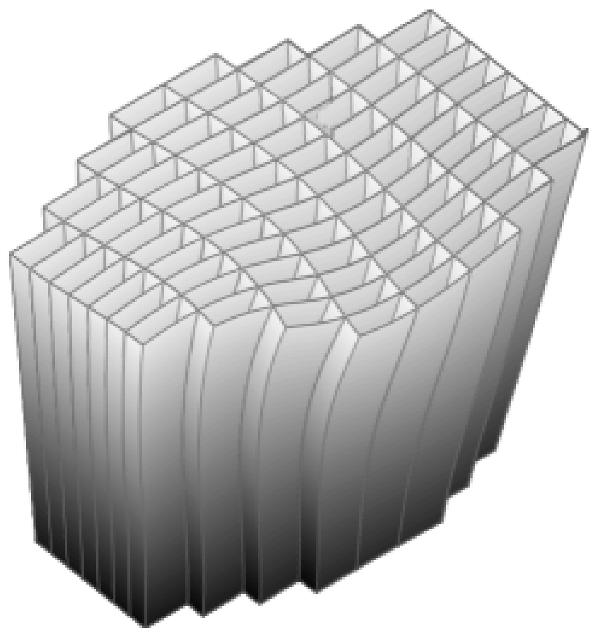


Figure 10.1: 2-form as tubes forming honeycomb.

Of course, we have made the trivial extension of $\wedge$ to the $\mathbb{Z}$-**graded algebra** by declaring that $\omega \wedge \eta = 0$ if either η or ω is homogeneous of negative degree.

The assignment $U \mapsto \Omega_M(U)$ is a presheaf of modules over $\mathcal{C}_M^\infty$. Similar remarks hold for Ω_M^k the (presheaf of) homogeneous forms of degree k. Sections from $\Omega(M)$ are called (global) differential forms or just forms for short.

Just as a tangent vector is the infinitesimal version of a curve through a point so a

10.1 Pullback of a differential form.

Given any smooth map $f : M \to N$ we can define the **pullback** map $f^* : \Omega(N) \to \Omega(M)$ as follows:

Definition 10.2 *Let $\eta \in \Omega^k(N)$. For vectors $v_1, ..., v_k \in T_pM$ define*

$$(f^*\eta)\,(p)(v_1, ..., v_k) = \eta_{f(p)}(T_pv_1, ..., T_pv_k)$$

then the map $f^*\eta : p \to (f^*\eta)\,(p)$ is a differential form on M. $f^*\eta$ is called the ***pullback of*** η ***by*** f.

One has the following easy to prove but important property

Proposition 10.1 *With $f : M \to N$ smooth map and $\eta_1, \eta_2 \in \Omega(N)$ we have*

$$f^*(\eta_1 \wedge \eta_2) = f^*\eta_1 \wedge f^*\eta_2$$

Proof: Exercise

Remark 10.1 *Notice the space $\Omega_M^0(U)$ is just the space of smooth functions $C^\infty(U)$ and so unfortunately we have several notations for the same set: $C^\infty(U) = \mathcal{C}_M^\infty(U) = \mathcal{F}_M(U) = \Omega_M^0(U)$.*

All that follows and much of what we have done so far works well for $\Omega_M(U)$ whether $U = M$ or not and will also respect restriction maps. Thus we will simply write Ω_M instead of $\Omega_M(U)$ or $\Omega(M)$ and $\mathfrak{X}_M$ instead of $\mathfrak{X}(U)$ so forth (recall remark 6.1). In fact, the exterior derivative d commutes with restrictions and so is really a presheaf map.

The algebra of smooth differential forms $\Omega(U)$ is an example of a $\mathbb{Z}$ graded algebra over the ring $C^\infty(U)$ and is also a graded vector space over $\mathbb{R}$. We have for each $U \subset M$

1) a the direct sum decomposition

$$\Omega(U) = \cdots \oplus \Omega^{-1}(U) \oplus \Omega^0(U) \oplus \Omega^1(U) \oplus \Omega^2(U) \cdots$$

where $\Omega^k(U) = 0$ if $k < 0$ or if $k > \dim(U)$;

2) The exterior product is a graded product:

$$\alpha \wedge \beta \in \Omega^{k+l}(U) \text{ for } \alpha \in \Omega^k(U) \text{ and } \beta \in \Omega^l(U)$$

which is

3) graded commutative: $\alpha \wedge \beta = (-1)^{kl}\beta \wedge \alpha$ for $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^l(U)$.

Each of these is natural with respect to restriction and so we have a **presheaf of graded algebras**.

10.2 Exterior Derivative

Here we will define and study the exterior derivative d.

Definition 10.3 *A graded derivation of degree r on $\Omega := \Omega_M$ is a sheaf map $\mathcal{D} : \Omega \to \Omega$ such that for each $U \subset M$,*

$$\mathcal{D} : \Omega^k(U) \to \Omega^{k+r}(U)$$

and such that for $\alpha \in \Omega^k(U)$ and $\beta \in \Omega(U)$ we have

$$\mathcal{D}(\alpha \wedge \beta) = \mathcal{D}\alpha \wedge \beta + (-1)^{kr} \alpha \wedge \mathcal{D}\beta.$$

Along lines similar to our study of tensor derivations one can show that a graded derivation of $\Omega(U)$ is completely determined by, and can be defined by it action on 0-forms (functions) and 1-forms. In fact, since every form can be locally built out of functions and exact one forms, i.e. differentials, we only need 0-forms and exact one forms to determine a graded derivation.

The differential d defined by

$$df(X) = Xf \text{ for } X \in \mathfrak{X}_M \tag{10.2}$$

is a map $\Omega^0_M \to \Omega^1_M$. This map can be extended to a degree one map from the graded space Ω_M to itself. Degree one means that writing $d : \Omega^0_M \to \Omega^1_M$ as $d_0 : \Omega^0_M \to \Omega^1_M$ we can find maps $d_i : \Omega^i_M \to \Omega^{i+1}_M$ that will satisfy our requirements.

Let $\omega_U : U \to L^k_{alt}(\mathsf{M}; \mathsf{M})$. In the following calculation we will identify $L^k_{alt}(\mathsf{M}; \mathsf{M})$ with the $L(\wedge^k \mathsf{M}, \mathsf{M})$. For $\xi_0, ..., \xi_k$ maps $\xi_i : U \to \mathsf{M}$ we have

$$\begin{aligned}
&D\langle \omega_U, \xi_0, ..., \xi_k \rangle(x) \cdot \xi_i \\
&= \left.\frac{d}{dt}\right|_0 \langle \omega_U(x + t\xi_i), \xi_0(x + t\xi_i) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x + t\xi_i) \rangle \\
&= \langle \omega_U(x), \left.\frac{d}{dt}\right|_0 [\xi_0(x + t\xi_i) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x + t\xi_i)] \rangle \\
&+ \langle \left.\frac{d}{dt}\right|_0 \omega_U(x), \xi_0(x + t\xi_i) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x + t\xi_i) \rangle \\
&= \langle \omega_U(x), \sum_{j=0}^{i-1} (-1)^j \xi_j'(x)\xi_i \wedge [\xi_0(x) \wedge ... \wedge \widehat{\xi_j} \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x)] \rangle \\
&+ \langle \omega_U(x), \sum_{j=i+1}^{k} (-1)^{j-1} \xi_j'(x)\xi_i \wedge [\xi_0(x) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \widehat{\xi_j} \wedge ... \wedge \xi_k(x)] \rangle \\
&+ \langle \omega_U'(x)\xi_i, \xi_0(x) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x) \rangle
\end{aligned}$$

Theorem 10.1 *There is a unique graded (sheaf) map $d : \Omega_M \to \Omega_M$, called the* ***exterior derivative****, such that*

1) $d \circ d = 0$

2) d is a graded derivation of degree one, that is

$$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge (d\beta) \tag{10.3}$$

for $\alpha \in \Omega_M^k$.

Furthermore, if $\omega \in \Omega^k(U)$ and $X_0, X_1, ..., X_k \in \mathfrak{X}_M(U)$ then

$$d\omega = \sum_{0 \le i \le k} (-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k))$$
$$+ \sum_{0 \le i < j \le k} (-1)^{i+j} \omega([X_i, X_j], X_0, ..., \widehat{X_i}, ..., \widehat{X_j}, ..., X_k).$$

In particular, we have the following useful formula for $\omega \in \Omega_M^1$ and $X, Y \in \mathfrak{X}_M(U)$:

$$d\omega(X, Y) = X(\omega(Y)) - Y\omega(X) - \omega([X, Y])$$

Proof. First we give a local definition in terms of coordinates and then show that the global formula **??** agree with the local formula. Let U, ψ be a local chart on M. We will first define the exterior derivative on the open set $V = \psi(U) \subset \mathbb{R}^n$. Let $\xi_0, ..., \xi_k$ be local vector fields. The local representation of a form ω is a map $\omega_U : V \to L^k_{skew}(\mathbb{R}^n;\mathbb{R})$ and so has at some $x \in V$ has a derivative $D\omega_U(x) \in L(\mathbb{R}^n, L^k_{skew}(\mathbb{R}^n;\mathbb{R}))$. We define

$$d\omega_U(x)(\xi_0, ..., \xi_k) := \sum_{i=0}^{k} (-1)^i (D\omega_U(x)\xi_i(x))(\xi_0(x), ..., \widehat{\xi_i}, ..., \xi_k(x))$$

where $D\omega_U(x)\xi_i(x) \in L^k_{skew}(\mathbb{R}^n; \mathbb{R}^n)$. This certainly defines a differential form in $\Omega^{k+1}(U)$. Let us call the right hand side of this local formula LOC. We wish to show that the global formula in local coordinates reduces to this local formula. Let us denote the first and second term of the global when expressed in local coordinates $L1$ and $L2$. Using, our calculation 10.2 we have

$$L1 = \sum_{i=0}^{k} (-1)^i \xi_i(\omega(\xi_0, ..., \widehat{\xi_i}, ..., \xi_k)) = \sum_{i=0}^{k} (-1)^i D(\omega(\xi_0, ..., \widehat{\xi_i}, ..., \xi_k))(x)\xi_i(x)$$

$$= \langle \omega_U(x), \sum_{i=0}^{k}(-1)^i \sum_{j=0}^{i-1}(-1)^j \xi_j^{'}(x)\xi_i \wedge [\xi_0(x) \wedge ... \wedge \widehat{\xi_j} \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x)] \rangle$$

$$+ \langle \omega_U(x), \sum_{i=0}^{k}(-1)^i \sum_{j=i+1}^{k}(-1)^{j-1} \xi_j^{'}(x)\xi_i \wedge [\xi_0(x) \wedge ... \wedge \widehat{\xi_i} \wedge ...$$

$$... \wedge \widehat{\xi_j} \wedge ... \wedge \xi_k(x)] \rangle$$

$$+ \sum_{i=0}^{k}(-1)^i \langle \omega_U'(x)\xi_i, \xi_0(x) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x) \rangle$$

$$= \langle \omega_U(x), \sum_{i=0}^{k}\sum_{i<j}^{k}(-1)^{i+j}(\xi_j^{'}(x)\xi_i - \xi_i^{'}(x)\xi_j) \wedge \xi_0(x) \wedge ...$$

$$.... \wedge \widehat{\xi_j} \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x) \rangle$$

$$= LOC + L2.$$

So our global formula reduces to the local one when expressed in local coordinates. ■

Remark 10.2 *As we indicated above, d is a local operator and so commutes with restrictions to open sets. In other words, if U is an open subset of M and d_U denotes the analogous operator on the manifold U then $d_U\alpha\big|_U = (d\alpha)\big|_U$. This operator can thus be expressed locally. In order to save on notation we will use d to denote the exterior derivative on any manifold, forms of any degree and for the restrictions d_U for any open set. It is exactly because d is a natural operator that this will cause no harm.*

If $\psi = (x^1, \ldots x^n)$ is a system of local coordinates on an open set U then all $\alpha \in \Omega_M^k(U)$ are sums of terms of the form $f dx^{i_1} \wedge \cdots \wedge dx^{i_k} = f dx^{\vec{I}}$ where $\vec{I} = (i_1, ..., i_k)$ is a strictly increasing sequence of element from $\{1, ..., n\}$. Then we have

$$d(f dx^{\vec{I}}) = df \wedge dx^{\vec{I}}. \tag{10.4}$$

written out this is

Lemma 10.1 *Given any smooth map $f : M \to N$ we have that d is natural with respect to the pull back:*

$$f^*(d\eta) = d(f^*\eta)$$

Proposition 10.2 $dd = 0$

Proof. This result is an easy but boring exercise in bookkeeping and boils down to the fact that for smooth (C^1) functions mixed partial derivatives are equal. For example,

$$\begin{aligned} d(df) = d(\frac{\partial f}{\partial x^i}dx^i) &= (\frac{\partial}{\partial x^j}\frac{\partial f}{\partial x^i}dx^j) \wedge dx^i \\ &= \frac{\partial^2}{\partial x^j \partial x^i}dx^j \wedge dx^i \\ = \sum_{j<i}\frac{\partial^2 f}{\partial x^j \partial x^i}dx^j \wedge dx^i + &\sum_{j>i}\frac{\partial^2}{\partial x^j \partial x^i}dx^j \wedge dx^i \\ = \sum_{j<i}\left(\frac{\partial^2 f}{\partial x^j \partial x^i} - \frac{\partial^2 f}{\partial x^i \partial x^j}\right) &dx^j \wedge dx^i = 0. \end{aligned}$$

Now the general result follows by using 10.3. ∎

Definition 10.4 *A smooth differential form α is called closed if $d\alpha = 0$ and exact if $\alpha = d\beta$ for some differential form β.*

Corollary 10.1 *Every exact form is closed.*

The converse is not true in general and the extend to which it fails is a topological property of the manifold. This is the point of the De Rham cohomology to be studied in detail in chapter 15. Here we just give the following basic definition:

Definition 10.5 *Since the exterior derivative operator is a graded map of degree one with $d^2 = 0$ we have, for each i, the* ***de Rham cohomology group*** *(actually vector spaces) given by*

$$H^i(M) = \frac{ker(d : \Omega^i(M) \to \Omega^{i+1}(M))}{Im(d : \Omega^{i-1}(M) \to \Omega^i(M))}. \tag{10.5}$$

In other words, we look at closed forms (forms α for which $d\alpha = 0$) and identify any two whose difference is an exact form (a form which is the exterior derivative of some other form).

10.3 Maxwell's equations.

Recall the electromagnetic field tensor

$$(F_{\mu\nu}) = \begin{bmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{bmatrix}.$$

Let us work in units where $c = 1$. Since this matrix is skew symmetric we can form a 2-form called the electromagnetic field 2-form:

$$F = \frac{1}{2}\sum_{\mu,\nu} F_{\mu\nu}dx^\mu \wedge dx^\nu = \sum_{\mu<\nu} F_{\mu\nu}dx^\mu \wedge dx^\nu.$$

Let write $E = E_x dx + E_y dy + E_z dz$ and $B = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy$. One can check that we now have

$$F = E \wedge dt - B.$$

Now we know that F comes from a potential $A = A_\nu dx^\nu$. In fact, we have

$$\begin{aligned} dA = d(A_\nu dx^\nu) &= \sum_{\mu<\nu} (\frac{\partial}{\partial x^\mu} A_\nu - \frac{\partial}{\partial x^\nu} A_\mu) dx^\mu \wedge dx^\nu \\ &= \sum_{\mu<\nu} F_{\mu\nu} dx^\mu \wedge dx^\nu = F. \end{aligned}$$

Thus we automatically have $dF = ddA = 0$. Now what does $dF = 0$ translate into in terms of the E and B? We compute:

$$\begin{aligned} dF &= d(E \wedge dt - B) = dE \wedge dt - dB \\ &= d\left(E_x dx + E_y dy + E_z dz\right) \wedge dt \\ &- (\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}) \\ &= \left[(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}) dy \wedge dz + (\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}) dz \wedge dx + (\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}) dy \wedge dx \right] \wedge dt \\ &+ \frac{\partial B}{\partial t} \wedge dt - (\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}) dx \wedge dy \wedge dz. \end{aligned}$$

From this we conclude that

$$\begin{aligned} \operatorname{div}(\tilde{\mathbf{B}}) &= 0 \\ \operatorname{curl}(\tilde{\mathbf{E}}) + \frac{\partial \tilde{\mathbf{B}}}{\partial t} &= 0 \end{aligned}$$

which is Maxwell's first two equations. Thus Maxwell's first two equations end up being equivalent to just the single equation

$$dF = 0$$

which was true just from the fact that $dd = 0$!

As for the second pair of Maxwell's equations, they too combine to give a single equation

$$*d * F = J.$$

Here J is the differential form constructed from the 4-current $\mathrm{j} = (\rho, \tilde{\mathbf{j}})$ introduced in section 26.4.8 by letting $(j_0, j_1, j_2, j_3) = (\rho, -\tilde{\mathbf{j}})$ and then $J = j_\mu dx^\mu$. The $*$ refers to the Hodge star operator on Minkowski space. The star operator on a general semi-Riemannian manifold will be studied later but we can give a formula for this special case.

Definition 10.6 *Define $\epsilon(\mu)$ to be entries of the diagonal matrix $\Lambda = diag(1,-1,-1,-1)$. Let $*$ be defined on $\Omega^k(\mathbb{R}^4)$ by letting $*(dx^{i_1}\wedge\cdots\wedge dx^{i_k}) = \pm\epsilon(j_1)\epsilon(j_2)\cdots\epsilon(j_k)dx^{j_1}\wedge\cdots\wedge dx^{j_{n-k}}$ where $dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_{n-k}} = \pm dx^0\wedge dx^1\wedge dx^2\wedge dx^3$. (Choose the sign to that makes the last equation true and then the first is true by definition). Extend $*$ linearly to a map $\Omega^k(\mathbb{R}^4)\to\Omega^{4-k}(\mathbb{R}^4)$.*

Exercise 10.1 *Show that $*\circ*$ acts on $\Omega^k(\mathbb{R}^4)$ by $(-1)^{k(4-k)+1}$. How would $*$ be different*

Exercise 10.2 *Show that $*d*F = J$ is equivalent to Maxwell's second two equations*

$$\begin{aligned}\operatorname{curl}(\tilde{\mathbf{B}}) &= \frac{\partial\tilde{\mathbf{E}}}{\partial t} + \tilde{\mathbf{j}}\\ \operatorname{div}(\tilde{\mathbf{E}}) &= \rho.\end{aligned}$$

10.4 Lie derivative, interior product and exterior derivative.

The Lie derivative acts on differential forms since the latter are, from one viewpoint, tensors. When we apply the Lie derivative to a differential form we get a differential form so we should think about the Lie derivative in the context of differential forms.

Lemma 10.2 *For any $X\in\mathfrak{X}(M)$ and any $f\in\Omega^0(M)$ we have $\mathcal{L}_X df = d\mathcal{L}_X f$.*

Proof. For a function f we compute as

$$\begin{aligned}&(\mathcal{L}_X df)(Y)\\ &= (\frac{d}{dt}(Fl_t^X)^* df)(Y) = \frac{d}{dt}df(TFl_t^X\cdot Y)\\ &= \frac{d}{dt}Y((Fl_t^X)^* f) = Y(\frac{d}{dt}(Fl_t^X)^* f)\\ &= Y(\mathcal{L}_X f) = d(\mathcal{L}_X f)(Y)\end{aligned}$$

where $Y\in\mathfrak{X}(M)$ is arbitrary. ∎

Exercise 10.3 *Show that $\mathcal{L}_X(\alpha\wedge\beta) = \mathcal{L}_X\alpha\wedge\beta + \alpha\wedge\mathcal{L}_X\beta$.*

We now have two ways to differentiate sections in $\Omega(M)$. Once again we write Ω_M instead of $\Omega(U)$ or $\Omega(M)$ since every thing works equally well in either case. In other words we are thinking of the presheaf $\Omega_M : U\mapsto\Omega(U)$. First, there is the Lie derivative which turns out to be a graded derivation of degree zero;

$$\mathcal{L}_X : \Omega_M^i \to \Omega_M^i. \tag{10.6}$$

Second, there is the exterior derivative that we just introduced which is a graded derivation of degree 1. In order to relate the two operations we need a third map which, like the Lie derivative, is taken with respect to a given field $X \in \Gamma(U; TM)$. This map is a degree -1 graded derivation and is defined by

$$\iota_X \omega(X_1, \ldots, X_{i-1}) = \omega(X, X_1, \ldots, X_{i-1}) \tag{10.7}$$

where we view $\omega \in \Omega^i_M$ as a skew-symmetric multi-linear map from $\mathfrak{X}_M \times \cdots \times \mathfrak{X}_M$ to $\mathcal{C}^\infty_M$. We could also define ι_X as that unique operator that satisfies

$$\iota_X \theta = \theta(X) \text{ for } \theta \in \Omega^1_M \text{ and } X \in \mathfrak{X}_M$$

$$\iota_X(\alpha \wedge \beta) = (\iota_X \alpha) \wedge \beta + (-1)^k \wedge \alpha \wedge (\iota_X \beta) \text{ for } \alpha \in \Omega^k_M.$$

In other word, ι_X is the graded derivation of Ω_M of degree -1 determined by the above formulas.

In any case, we will call this operator the **interior product** or **contraction operator**.

Notation 10.1 *Other notations for $\iota_X \omega$ include $X \lrcorner \omega = \langle X, \omega \rangle$. These notations make the following theorem look more natural:*

Theorem 10.2 *The Lie derivative is a derivation with respect to the pairing $\langle X, \omega \rangle$. That is*

$$\mathcal{L}_X \langle X, \omega \rangle = \langle \mathcal{L}_X X, \omega \rangle + \langle X, \mathcal{L}_X \omega \rangle$$

or

$$\mathcal{L}_X (X \lrcorner \omega) = (\mathcal{L}_X X) \lrcorner \omega + X \lrcorner (\mathcal{L}_X \omega)$$

Using the "ι_X" notation: $\mathcal{L}_X(\iota_X \omega) = \iota_{\mathcal{L}_X X} \omega + \iota_X \mathcal{L}_X \omega$ (not as pretty).

Proof. Exercise. ■

Now we can relate the Lie derivative, the exterior derivative and the contraction operator.

Theorem 10.3 *Let $X \in \mathfrak{X}_M$. Then we have Cartan's homotopy formula;*

$$\mathcal{L}_X = d \circ \iota_X + \iota_X \circ d \tag{10.8}$$

Proof. One can check that both sides define derivations and so we just have to check that they agree on functions and exact 1-forms. On functions we have $\iota_X f = 0$ and $\iota_X df = Xf = \mathcal{L}_X f$ so formula holds. On differentials of functions we have

$$(d \circ \iota_X + \iota_X \circ d) df = (d \circ \iota_X) df = d \mathcal{L}_X f = \mathcal{L}_X df$$

where we have used lemma 10.2 in the last step. ■

As a corollary can now extend lemma 10.2:

Corollary 10.2 $d \circ \mathcal{L}_X = \mathcal{L}_X \circ d$

Proof.

$$\begin{aligned} d\mathcal{L}_X\alpha &= d(d\iota_X + \iota_X d)(\alpha) \\ &= d\iota_X d\alpha = d\iota_X d\alpha + \iota_X dd\alpha = \mathcal{L}_X \circ d \end{aligned}$$

■

Corollary 10.3 *We have the following formulas:*
1) $\iota_{[X,Y]} = \mathcal{L}_X \circ \iota_Y + \iota_Y \circ \mathcal{L}_X$
2) $\mathcal{L}_{fX}\omega = f\mathcal{L}_X\omega + df \wedge \iota_X\omega$ *for all* $\omega \in \Omega(M)$.

Proof. Exercise. ■

10.5 Time Dependent Fields (Part II)

If we have a time parameterized family of p-forms α_t on M such that $\alpha_t(x) := \alpha(t,x)$ is jointly smooth in t and x, i.e. a time dependent p-form, then we can view it as a p-form on the manifold $\mathbb{R} \times M$. Of course, $\alpha(t,x)$ might only be defined on some open neighborhood of $\{0\} \times M \subset \mathbb{R} \times M$ but we shall assume that $\alpha(t,x)$ is defined on all of $\mathbb{R} \times M$. This is only for simplicity in notation and does not effect the results in any essential way.

10.6 Vector valued and algebra valued forms.

Given vector spaces V and W, one can also define the space $T_k^0(\mathrm{V};\mathrm{W}) = L^k_{skew}$ $(\mathrm{V};\mathrm{W})$ of all (bounded) skew-symmetric maps

$$\underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{k-\text{times}} \to \mathrm{W}$$

We define the wedge product using the same formula as before except that we use the tensor product so that $\alpha \wedge \beta \in L^k_{skew}$ $(\mathrm{V};\mathrm{W}\otimes\mathrm{W})$:

$$\begin{aligned} &(\omega \wedge \eta)(v_1, v_2, .., v_r, v_{r+1}, v_{r+2}, .., v_{r+s}) \\ &= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma)\omega(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r}) \otimes \eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}}) \end{aligned}$$

Globalizing this algebra as usual we get a vector bundle $\mathrm{W}\otimes(\bigwedge^k T^*M)$ which in turn gives rise to a space of sections $\Omega^k(M,\mathrm{W})$ (and a presheaf $U \mapsto \Omega^k(U,\mathrm{W})$) and exterior product $\Omega^k(U,\mathrm{W}) \times \Omega^l(U,\mathrm{W}) \to \Omega^{k+l}(U,\mathrm{W}\otimes\mathrm{W})$. The space $\Omega^k(U,\mathrm{W})$ is a module over $\mathcal{C}^\infty_M(U)$. We still have pullback and a natural exterior derivative

$$d : \Omega^k(U,\mathrm{W}) \to \Omega^{k+1}(U,\mathrm{W})$$

defined by the formula

$$
\begin{aligned}
&d\omega(X_0, ..., X_k) \\
&= \sum_{1\leq i\leq k} (-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\leq i<j\leq k} (-1)^{i+j} \omega(X_0, ..., [X_i, X_j], ..., X_k)
\end{aligned}
$$

where now $\omega(X_0, ..., \widehat{X_i}, ..., X_k)$ is a W-valued function so we take

$$
\begin{aligned}
&X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k))(p) \\
&= D\omega(X_0, ..., \widehat{X_i}, ..., X_k)\Big| \cdot X_i(p) \\
&= d(\omega(X_0, ..., \widehat{X_i}, ..., X_k))(X_i(p))
\end{aligned}
$$

which is an element of W under the usual identification of W with any of its tangent spaces.

To give a local formula valid for finite dimensional M, we let $f_1, ..., f_n$ be a basis of W and $(x^1, ..., x^n)$ local coordinates defined on U. For $\omega = \sum a^j_{\vec{I}}, f_j \otimes dx^{\vec{I}}$ we have

$$
\begin{aligned}
d\omega &= d(f_j \otimes \sum a_{\vec{I},j} dx^{\vec{I}}) \\
&= \sum (f_j \otimes da_{\vec{I},j} \wedge dx^{\vec{I}}).
\end{aligned}
$$

The elements $f_j \otimes dx_p^{\vec{I}}$ form a basis for the vector space $\mathrm{W} \otimes (\bigwedge^k T_p^* M)$ for every $p \in U$.

Now if W happens to be an algebra then the algebra product $\mathrm{W} \times \mathrm{W} \to \mathrm{W}$ is bilinear and so gives rise to a linear map $m{:}\mathrm{W} \otimes \mathrm{W} \to \mathrm{W}$. We compose the exterior product with this map to get a wedge product $\overset{m}{\wedge} : \Omega^k(U, \mathrm{W}) \times \Omega^l(U, \mathrm{W}) \to \Omega^{k+l}(U, \mathrm{W})$

$$
\begin{aligned}
&(\omega \overset{m}{\wedge} \eta)(v_1, v_2, .., v_r, v_{r+1}, v_{r+2}, .., v_{r+s}) \\
&= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma) m \left(\omega(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r}) \otimes \eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})\right) \\
&= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma) \omega(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r}) \cdot \eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})
\end{aligned}
$$

A particularly important case is when W is a Lie algebra $\mathfrak{g}$ with bracket $[.,.]$. Then we write the resulting product $\overset{m}{\wedge}$ as $[.,.]_\wedge$ or just $[.,.]$ when there is no risk of confusion. Thus if $\omega, \eta \in \Omega^1(U, \mathfrak{g})$ are Lie algebra valued 1-forms then

$$[\omega, \eta](X) = [\omega(X), \eta(Y)] + [\eta(X), \omega(Y)].$$

In particular, $\frac{1}{2}[\omega, \omega](X, Y) = [\omega(X), \omega(Y)]$ which might not be zero in general!

10.7 Global Orientation

A rank n vector bundle $E \to M$ is called oriented if every fiber E_p is given a smooth choice of orientation. There are several equivalent ways to make a rigorous definition:

1. A vector bundle is orientable iff has an atlas of bundle charts such that the corresponding transition maps take values in $GL^+(n, \mathbb{R})$ the group of positive determinant matrices. If the vector bundle is orientable then this divides the set of all bundle charts into two classes. Two bundle charts are in the same orientation class the transition map takes values in $GL^+(n, \mathbb{R})$. If the bundle is not orientable there is only one class.

2. If there is a smooth global section s on the bundle $\bigwedge^n E \to M$ then we say that this determines an orientation on E. A frame $(f_1, ..., f_n)$ of fiber E_p is positively oriented with respect to s if and only if $f_1 \wedge ... \wedge f_n = as(p)$ for a positive real number $a > 0$.

3. If there is a smooth global section ω on the bundle $\bigwedge^n E^* \cong L^k_{alt}(E) \to M$ then we say that this determines an orientation on E. A frame $(f_1, ..., f_n)$ of fiber E_p is positively oriented with respect to ω if and only if $\omega(p)(f_1, ..., f_n) > 0$.

Exercise 10.4 *Show that each of these three approaches is equivalent.*

Now let M be an n-dimensional manifold. Let U be some open subset of M which may be all of M. Consider a **top form**, i.e. an n-form $\varpi \in \Omega^n_M(U)$ where $n = \dim(M)$ and assume that ϖ is never zero on U. In this case we will say that ϖ is nonzero or that ϖ is a **volume form**. Every other top form μ is of the form $\mu = f\varpi$ for some smooth function f. This latter fact follows easily from $\dim(\bigwedge^n T_pM) = 1$ for all p. If $\varphi : U \to U$ is a diffeomorphism then we must have that $\varphi^*\varpi = \delta\varpi$ for some $\delta \in C^\infty(U)$ which we will call the Jacobian determinate of φ with respect to the volume element ϖ:

$$\varphi^*\varpi = J_\varpi(\varphi)\varpi$$

Proposition 10.3 *The sign of $J_\varpi(\varphi)$ is independent of the choice of volume form ϖ.*

Proof. Let $\varpi' \in \Omega^n_M(U)$. We have

$$\varpi = a\varpi'$$

for some function a which is never zero on U. We have

$$\begin{aligned} J(\varphi)\varpi &= (\varphi^*\varpi) = (a \circ \varphi)(\varphi^*\varpi') \\ &= (a \circ \varphi)J_{\varpi'}(\varphi)\varpi' = \frac{a \circ \varphi}{a}\varpi \end{aligned}$$

and since $\frac{a \circ \varphi}{a} > 0$ and ϖ is nonzero the conclusion follows. ■

Let us consider a very important special case of this: Suppose that $\varphi : U \to U$ is a diffeomorphism and $U \subset \mathbb{R}^n$. Then letting $\varpi_0 = du^1 \wedge \cdots \wedge du^n$ we have

$$\begin{aligned}
\varphi^* \varpi_0(x) &= \varphi^* du^1 \wedge \varphi^* \cdots \wedge \varphi^* du^n(x) \\
&= \left(\sum \frac{\partial(u^1 \circ \varphi)}{\partial u^{i_1}} \bigg|_x du^{i_1} \right) \wedge \cdots \wedge \left(\sum \frac{\partial(u^n \circ \varphi)}{\partial u^{i_n}} \bigg|_x du^{i_n} \right) \\
&= \det \left(\frac{\partial(u^i \circ \varphi)}{\partial u^j}(x) \right) = J\varphi(x).
\end{aligned}$$

so in this case $J_{\varpi_0}(\varphi)$ is just the usual Jacobian determinant of φ.

Definition 10.7 *A diffeomorphism $\varphi : U \to U \subset \mathbb{R}^n$ is said to be positive or orientation preserving if* $\det(T\varphi) > 0$.

More generally, let a nonzero top form ϖ be defined on $U \subset M$ and let ϖ' be another defined on $U' \subset N$. Then we say that a diffeomorphism $\varphi : U \to U'$ is orientation preserving (or positive) with respect to the pair ϖ, ϖ' if the unique function $J_{\varpi,\varpi'}$ such that $\varphi^* \varpi' = J_{\varpi,\varpi'} \varpi$ is strictly positive on U.

Definition 10.8 *A differentiable manifold M is said to be* ***orientable*** *iff there is an atlas of admissible charts such that for any pair of charts ψ_α, U_α and ψ_β, U_β from the atlas with $U_\alpha \cap U_\beta \neq \emptyset$, the transition map $\psi_\beta \circ \psi_\alpha^{-1}$ is orientation preserving. Such an atlas is called an orienting atlas.*

Exercise 10.5 *The tangent bundle is a vector bundle. Show that this last definition agrees with our definition of an orientable vector bundle in that M is an orientable manifold in the current sense if and only if TM is an orientable vector bundle.*

Let $\mathcal{A}_M$ be the maximal atlas for a orientable differentiable manifold M. Then there are two sub-atlas $\mathcal{A}$ and $\mathcal{A}'$ with $\mathcal{A} \cup \mathcal{A}' = \mathcal{A}_M$, $\mathcal{A} \cap \mathcal{A}' = \emptyset$ and such that the transition maps for charts from $\mathcal{A}$ are all positive and similarly the transition maps of $\mathcal{A}'$ are all positive.. Furthermore if $\psi_\alpha, U_\alpha \in \mathcal{A}$ and $\psi_\beta, U_\beta \in \mathcal{A}'$ then $\psi_\beta \circ \psi_\alpha^{-1}$ is negative (orientation reversing). A choice of one these two atlases is called an **orientation on** M. Every orienting atlas is a subatlas of exactly one of $\mathcal{A}$ or $\mathcal{A}'$. If such a choice is made then we say that M is **oriented**. Alternatively, we can use the following proposition to specify an orientation on M:

Proposition 10.4 *Let $\varpi \in \Omega^n(M)$ be a volume form on M, i.e. ϖ is a nonzero top form. Then ϖ determines an orientation by determining an (orienting) atlas $\mathcal{A}$ by the rule*

$$\psi_\alpha, U_\alpha \in \mathcal{A} \Longleftrightarrow \psi_\alpha \text{ is orientation preserving resp. } \varpi, \varpi_0$$

where ϖ_0 is the standard volume form on $\mathbb{R}^n$ introduced above.

Exercise 10.6 *Prove the last proposition and then prove that we can use an orienting atlas to construct a volume form on an orientable manifold that gives the same orientation as the orienting atlas.*

We now construct a two fold covering manifold $\mathrm{Or}(M)$ for any finite dimensional manifold called the orientation cover. The orientation cover will itself always be orientable. Consider the vector bundle $\bigwedge^n T^*M$ and remove the zero section to obtain

$$(\bigwedge^n T^*M)^\times := \bigwedge^n T^*M - \{\text{zero section}\}$$

Define an equivalence relation on $(\bigwedge^n T^*M)^\times$ by declaring $\nu_1 \sim \nu_2$ iff ν_1 and ν_2 are in the same fiber and if $\nu_1 = a\nu_2$ with $a > 0$. The space of equivalence classes is denoted $\mathrm{Or}(M)$. There is a unique map π_{Or} making the following diagram commute:

$$\begin{array}{ccc} \mathrm{Or}(M) & \rightarrow & (\bigwedge^n T^*M)^\times \\ & & \downarrow \\ \pi_{Or} & \searrow & \\ & & M \end{array}$$

Now $\mathrm{Or}(M) \to M$ is a covering space with the quotient topology and in fact is a differentiable manifold.

10.8 Orientation of manifolds with boundary

Recall that a **half space chart** ψ_α for a manifold with boundary M is a bijection (actually diffeomorphism) of an open subset U_α of M onto an open subset of $\mathbb{R}^{n-}_\lambda$. A C^r **half space atlas** is a collection ψ_α, U_α of such charts such that for any two; ψ_α, U_α and ψ_β, U_β, the map $\psi_\alpha \circ \psi_\beta^{-1}$ is a C^r diffeomorphism on its natural domain (if non-empty). Note: "Diffeomorphism" means in the extended the sense of a being homeomorphism and such that both $\psi_\alpha \circ \psi_\beta^{-1} :: \mathbb{R}^{n-}_\lambda \to \mathbb{R}^n$ and its inverse are C^r in the sense of definition 2.5.

Let us consider the case of finite dimensional manifolds. Then letting $\mathbb{R}^n = \mathbb{R}^n$ and $\lambda = pr_1 : \mathbb{R}^n \to \mathbb{R}$ we have the half space $\mathbb{R}^{n-}_\lambda = \mathbb{R}^n_{u^1 \le 0}$. The funny choice of sign is to make $\mathbb{R}^{n-}_\lambda = \mathbb{R}^n_{u^1 \le 0}$ rather than $\mathbb{R}^n_{u^1 \ge 0}$. The reason we do this is to be able to get the right induced orientation on ∂M without introducing a minus sign into our Stoke's formula proved below. The reader may wish to re-read remark 2.6 at this time.

Now, imitating our previous definition we define an oriented (or orienting) atlas for a finite dimensional manifold with boundary to be an atlas *with ranges all in* $\mathbb{R}^n_{u^1 \le 0}$ and such that the overlap maps $\psi_\alpha \circ \psi_\beta^{-1} :: \mathbb{R}^n_{u^1 \le 0} \to \mathbb{R}^n_{u^1 \le 0}$ are orientation preserving. A manifold with boundary with a choice of (maximal) oriented atlas is called an **oriented manifold with boundary**. If there exists an orienting atlas for M then we say that M is **orientable** just as the case of a manifold without boundary.

Now if $\mathcal{A} = \{(\psi_\alpha, U_\alpha)\}_{\alpha\in A}$ is an orienting atlas for M as above with domains in $\mathbb{R}^n_{u^1\leq 0}$ then the induced atlas $\{(\psi_\alpha|_{U_\alpha\cap\partial M}, U_\alpha\cap\partial M)\}_{\alpha\in A}$ is an orienting atlas for the manifold ∂M and the resulting choice of orientation is called the induced orientation on ∂M. If M is oriented we will always assume that ∂M is given this induced orientation.

Definition 10.9 *A basis $f_1, f_2, ..., f_n$ for the tangent space at a point p on an oriented manifold (with or without boundary) is called positive if whenever $\psi_\alpha = (x^1, ..., x^n)$ is an oriented chart on a neighborhood of p then $(dx^1 \wedge ... \wedge dx^n)(f_1, f_2, ..., f_n) > 0$.*

Definition 10.10 *A vector v in T_pM for a point p on the boundary ∂M is called outward pointing if $T_p\psi_\alpha \cdot v \in \mathbb{R}^{n-}_\lambda$ is outward pointing in the sense that $\lambda(T_p\psi_\alpha \cdot v) > 0$.*

Since we have chosen $\lambda = pr_1$ and hence $\mathbb{R}^{n-}_\lambda = \mathbb{R}^n_{u^1\leq 0}$ for our definition in choosing the orientation on the boundary we have that in this case v is outward pointing iff $T_p\psi_\alpha \cdot v \in \mathbb{R}^n_{u^1\leq 0}$.

Definition 10.11 *A* ***nice chart*** *on a smooth manifold (possibly with boundary) is a chart ψ_α, U_α where ψ_α is a diffeomorphism onto $\mathbb{R}^n_{u^1\leq 0}$ if $U_\alpha\cap\partial M \neq \emptyset$ and a diffeomorphism onto the interior $\mathbb{R}^n_{u^1<0}$ if $U_\alpha\cap\partial M = \emptyset$.*

Lemma 10.3 *Every (oriented) smooth manifold has an (oriented) atlas consisting of nice charts.*

Proof. If ψ_α, U_α is an oriented chart with range in the interior of the left half space $\mathbb{R}^n_{u^1\leq 0}$ then we can find a ball B inside $\psi_\alpha(U_\alpha)$ in $\mathbb{R}^n_{u^1<0}$ and then we form a new chart on $\psi_\alpha^{-1}(B)$ with range B. But a ball is diffeomorphic to $\mathbb{R}^n_{u^1<0}$. So composing with such a diffeomorphism we obtain the nice chart. If ψ_α, U_α is an oriented chart with range meeting the boundary of the left half space $\mathbb{R}^n_{u^1\leq 0}$ then we can find a half ball B_- in $\mathbb{R}^n_{u^1\leq 0}$ with center on $\mathbb{R}^n_{u^1=0}$. Reduce the chart domain as before to have range equal to this half ball. But every half ball is diffeomorphic to the half space $\mathbb{R}^n_{u^1\leq 0}$ so we can proceed by composition as before. ∎

Proposition 10.5 *If $\psi_\alpha = (x^1, ..., x^n)$ is an oriented chart on a neighborhood of p on the boundary of an oriented manifold with boundary then the vectors $\frac{\partial}{\partial x^2}, ..., \frac{\partial}{\partial x^n}$ form a positive basis for $T_p\partial M$ with respect to the induced orientation on ∂M. More generally, if f_1 is outward pointing and $f_1, f_2, ..., f_n$ is positive on M at p, then $f_2, ..., f_n$ will be positive for ∂M at p.*

10.9 Integration of Differential Forms.

Let M be a smooth n-manifold possibly with boundary ∂M and assume that M is oriented and that ∂M has the induced orientation. From our discussion on orientation of manifolds with boundary and by general principles it should be clear that we may assume that all the charts in our orienting atlas have range in the left half space $\mathbb{R}^n_{u^1\leq 0}$. If $\partial M = \emptyset$ then the ranges will just be in the interior $\mathbb{R}^n_{u^1<0} \subset \mathbb{R}^n_{u^1\leq 0}$.

Every k-form $\alpha^{(k)}$ on an open subset U of $\mathbb{R}^n_{u^1\leq 0}$ is of the form $\alpha^{(k)} = a(.)du^1 \wedge \cdots \wedge du^k$ for some smooth function $a(.) \in C^\infty(U)$. If $a(.)$ has compact support $\mathrm{supp}(a)$ then we say that $\alpha^{(k)}$ has compact support $\mathrm{supp}(\alpha^{(k)}) := \mathrm{supp}(a)$. In general, we have the following

Definition 10.12 *A the* ***support*** *of a differential form $\alpha \in \Omega(M)$ is the* closure *of the set $\{p \in M : \alpha(p) \neq 0\}$ and is denoted by* $\mathrm{supp}(\alpha)$. *The set of all k-forms $\alpha^{(k)}$ which have compact support contained in $U \subset M$ is denoted by $\Omega^k_c(U)$.*

Let us return to the case of a form $\alpha^{(k)}$ on an open subset U of $\mathbb{R}^k$. If $\alpha^{(k)}$ has compact support in U we may define the integral $\int_U \alpha^{(k)}$ by

$$\begin{aligned}\int_U \alpha^{(k)} &= \int_U a(u)du^1 \wedge \cdots \wedge du^k \\ &:= \int_U a(u) \left|du^1 \cdots du^k\right|\end{aligned}$$

where this latter integral is the Lebesgue integral of $a(u)$. We have written $\left|du^1 \cdots du^k\right|$ instead of $du^1 \cdots du^k$ to emphasize that the order of the du^i does not matter as it does for $du^1 \wedge \cdots \wedge du^k$.

Now consider an oriented $n-$dimensional manifold M and let $\alpha \in \Omega^n_M$. If α has compact support inside U_α for some chart ψ_α, U_α compatible with the orientation then $\psi_\alpha^{-1} : \psi_\alpha(U_\alpha) \to U_\alpha$ and $(\psi_\alpha^{-1})^*\alpha$ has compact support in $\psi_\alpha(U_\alpha) \subset \mathbb{R}^n_{u^1\leq 0}$. We define

$$\int \alpha := \int_{\psi_\alpha(U_\alpha)} (\psi_\alpha^{-1})^*\alpha.$$

The standard change of variables formula show that this definition is independent of the oriented chart chosen. Now if $\alpha \in \Omega^n(M)$ does not have support contained in some chart domain then we choose a locally finite cover of M by oriented charts ψ_i, U_i and a smooth partition of unity ρ_i, U_i , $\mathrm{supp}(\rho_i) \subset U_i$. Then we define

$$\int \alpha := \sum_i \int_{\psi_i(U_i)} (\psi_i^{-1})^*(\rho_i \alpha)$$

Proposition 10.6 *The above definition is independent of the choice of the charts ψ_i, U_i and smooth partition of unity ρ_i, U_i.*

Proof. Let ϕ_i, V_i, and $\overline{\rho}_i$ be another such choice. Then we have

$$\begin{aligned}
\int \alpha &:= \sum_i \int_{\psi_i(U_i)} (\psi_i^{-1})^*(\rho_i \alpha) \\
&= \sum_i \int_{\psi_i(U_i)} (\psi_i^{-1})^*(\rho_i \sum_j \overline{\rho}_j \alpha) \\
&\sum_i \sum_j \int_{\psi_i(U_i \cap U_j)} (\psi_i^{-1})^*(\rho_i \overline{\rho}_j \alpha) \\
&= \sum_i \sum_j \int_{\phi_j(U_i \cap U_j)} (\phi_j^{-1})^*(\rho_i \overline{\rho}_j \alpha) \\
&= \sum_j \int_{\psi_i(U_i)} (\phi_j^{-1})^*(\overline{\rho}_j \alpha)
\end{aligned}$$

∎

10.10 Stokes' Theorem

Let us start with a couple special cases .

Case 10.1 (1) *Let $\omega_j = f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n$ be a smooth $n-1$ form with compact support contained in the interior of $\mathbb{R}^n_{u^1 \le 0}$ where the hat symbol over the du^j means this j-th factor is omitted. All $n-1$ forms on $\mathbb{R}^n_{u^1 \le 0}$ are sums of forms of this type. Then we have*

$$\begin{aligned}
\int_{\mathbb{R}^n_{u^1 \le 0}} d\omega_j &= \int_{\mathbb{R}^n_{u^1 \le 0}} d(f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^n_{u^1 \le 0}} (df \wedge du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^n_{u^1 \le 0}} (\sum_k \frac{\partial f}{\partial u^k} du^k \wedge du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^n_{u^1 \le 0}} (-1)^{j-1} \frac{\partial f}{\partial u^j} du^1 \wedge \cdots \wedge du^n = \int_{\mathbb{R}^n} (-1)^{j-1} \frac{\partial f}{\partial u^j} du^1 \cdots du^n \\
&= 0
\end{aligned}$$

by the fundamental theorem of calculus and the fact that f has compact support.

Case 10.2 (2) *Let $\omega_j = f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n$ be a smooth $n-1$ form* *with compact support meeting $\partial\mathbb{R}^n_{u^1\le 0} = \mathbb{R}^n_{u^1=0} = 0\times\mathbb{R}^{n-1}$ then*

$$
\begin{aligned}
\int_{\mathbb{R}^n_{u^1\le 0}} d\omega_j &= \int_{\mathbb{R}^n_{u^1\le 0}} d(f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^{n-1}} (-1)^{j-1} \left(\int_{-\infty}^{\infty} \frac{\partial f}{\partial u^j} du^j \right) du^1 \cdots \widehat{du^j} \cdots du^n = \\
&= 0 \text{ if } j \neq 1 \text{ and if } j = 1 \text{ we have} \int_{\mathbb{R}^n_{u^1\le 0}} d\omega_1 = \\
&= \int_{\mathbb{R}^{n-1}} (-1)^{j-1} \left(\int_{-\infty}^{0} \frac{\partial f}{\partial u^1} du^1 \right) du^2 \wedge \cdots \wedge du^n \\
&= \int_{\mathbb{R}^{n-1}} f(0, u^2, ..., u^n) du^2 \cdots du^n \\
&= \int_{\partial\mathbb{R}^n_{u^1\le 0}} f(0, u^2, ..., u^n) du^2 \wedge \cdots \wedge du^n = \int_{\partial\mathbb{R}^n_{u^1\le 0}} \omega_1
\end{aligned}
$$

Now since clearly $\int_{\partial\mathbb{R}^n_{u^1\le 0}} \omega_j = 0$ if $j \neq 1$ or if ω_j has support that doesn't meet $\partial\mathbb{R}^n_{u^1\le 0}$ we see that in any case $\int_{\mathbb{R}^n_{u^1\le 0}} d\omega_j = \int_{\partial\mathbb{R}^n_{u^1\le 0}} \omega_j$. Now as we said all $n-1$ forms on $\mathbb{R}^n_{u^1\le 0}$ are sums of forms of this type and so summing such we have for any smooth $n-1$ form on $\mathbb{R}^n_{u^1\le 0}$.

$$\int_{\mathbb{R}^n_{u^1\le 0}} d\omega = \int_{\partial\mathbb{R}^n_{u^1\le 0}} \omega.$$

Now we define integration on a manifold (possibly with boundary). Let $\mathcal{A}_M = (\psi_\alpha, U_\alpha)_{\alpha\in A}$ be an oriented atlas for a smooth orientable n-manifold M consisting of nice charts so either $\psi_\alpha : U_\alpha \cong \mathbb{R}^n$ or $\psi_\alpha : U_\alpha \cong \mathbb{R}^n_{u^1\le 0}$. Now let $\{\rho_\alpha\}$ be a smooth partition of unity subordinate to $\{U_\alpha\}$. Notice that $\{\rho_\alpha|_{U_\alpha\cap\partial M}\}$ is a partition of unity for the cover $\{U_\alpha \cap \partial M\}$ of ∂M. Then for $\omega \in \Omega^{n-1}(M)$ we have that

$$
\begin{aligned}
\int_M d\omega &= \int_{U_\alpha} \sum_\alpha d(\rho_\alpha \omega) = \sum_\alpha \int_{U_\alpha} d(\rho_\alpha \omega) \\
&= \sum_\alpha \int_{\psi_\alpha(U_\alpha)} \psi_\alpha^* d(\rho_\alpha \omega) = \sum_\alpha \int_{\psi_\alpha(U_\alpha)} d(\psi_\alpha^* \rho_\alpha \omega) \\
&= \sum_\alpha \int_{\psi_\alpha(U_\alpha)} d(\psi_\alpha^* \rho_\alpha \omega) = \sum_\alpha \int_{\partial\psi_\alpha(U_\alpha)} (\psi_\alpha^* \rho_\alpha \omega) \\
&= \sum_\alpha \int_{\partial U_\alpha} \rho_\alpha \omega = \int_{\partial M} \omega
\end{aligned}
$$

so we have proved

Theorem 10.4 (Stokes' Theorem) *Let M be an oriented manifold with boundary (possibly empty) and give ∂M the induced orientation. Then for any*

$\omega \in \Omega^{n-1}(M)$ we have

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

10.11 Vector Bundle Valued Forms.

It will occur in several contexts to have on hand the notion of a differential form with values in a vector bundle.

Definition 10.13 *Let $\xi = (\mathbb{F} \hookrightarrow E \to M)$ be a smooth vector bundle. A differential $p-$form with values in ξ (or values in E) is a smooth section of the bundle $E \otimes \wedge^p T^*M$. These are denoted by $\Omega^p(\xi)$ (or informally $\Omega^p(E)$ if the context is sufficient to avoid confusion with $\Gamma(M, \wedge^p E)$).*

In order to get a grip on the meaning of the bundle let use exhibit transition functions. We know that for a vector bundle knowing the transition functions is tantamount to knowing how local expressions with respect to a frame transform as we change frame(?did I explain this?). A frame for $E \otimes \wedge^p T^*M$ is given by combining a local frame for E with a local frame for $\wedge^p TM$. Of course we must choose an common refinement of the VB-charts to do this but this is obviously no problem. Let $(e_1, ..., e_k)$ frame defined on U which we may as well take to also be a chart domain for the manifold M. Then any local section of $\Omega^p(\xi)$ defined on U has the form

$$\sigma = \sum a^j_{\vec{I}} e_j \otimes dx^I$$

for some smooth functions $a^j_{\vec{I}} = a^j_{i_1 \ldots i_p}$ defined in U. Then for a new local set up with frames $(f_1, ..., f_k)$ and $dy^{\vec{I}} = dy^{i_1} \wedge \cdots \wedge dy^{i_p}$ $(i_1 < ... < i_p)$ then

$$\sigma = \sum \acute{a}^j_{\vec{I}} f_j \otimes dy^{\vec{I}}$$

we get the transformation law

$$\acute{a}^j_{\vec{I}} = a^i_{\vec{J}} C^j_i \frac{\partial x^{\vec{J}}}{\partial y^{\vec{I}}}$$

and where C^j_i is defined by $f_s C^s_j = e_j$.

Exercise 10.7 *Derive the above transformation law.*

Solution 10.1 $\sum a^j_{\vec{I}} e_j \otimes dx^I = \sum a^j_{\vec{I}} f_s C^s_j \otimes \frac{\partial x^{\vec{J}}}{\partial y^{\vec{I}}} dy^I$ *etc.*

A more elegant way of describing the transition functions is just to recall that anytime w have two vector bundles over the same base space and respective typical fibers V and W then the respective transition functions $g_{\alpha\beta}$ and $h_{\alpha\beta}$ (on a common cover) combine to give $g_{\alpha\beta} \otimes h_{\alpha\beta}$ where for a given $x \in U_{\alpha\beta}$

$$g_{\alpha\beta}(x) \otimes h_{\alpha\beta}(x) : \mathsf{V} \otimes \mathsf{W} \to \mathsf{V} \otimes \mathsf{W}$$
$$g_{\alpha\beta}(x) \otimes h_{\alpha\beta}(x).(\mathsf{v}, \mathsf{w}) = g_{\alpha\beta}(x)\mathsf{v} \otimes h_{\alpha\beta}(x)\mathsf{w}.$$

At any rate, these transformation laws fade into the background since if all out expressions are manifestly invariant (or invariantly defined in the first place) then we don't have to bring them up. A more important thing to do is to get used to calculating.

If the vector bundle is actually an algebra bundle then (naming the bundle $\mathcal{A} \to M$ now for "algebra") we may turn $\mathcal{A} \otimes \wedge T^*M := \sum_{p=0}^{n} \mathcal{A} \otimes \wedge^p T^*M$ into an algebra bundle by defining

$$(v_1 \otimes \mu^1) \wedge (v_2 \otimes \mu^2) := v_1 v_2 \otimes \mu^1 \wedge \mu^2$$

and then extending linearly:

$$(a^i_j v_i \otimes \mu^j) \wedge (b^k_l v_k \otimes \mu^l) := v_i v_j \otimes \mu^j \wedge \mu^l$$

From this the sections $\Omega(M, \mathcal{A}) = \Gamma(M, \mathcal{A} \otimes \wedge T^*M)$ become an algebra over the ring of smooth functions. For us the most important example is where $\mathcal{A} = \mathrm{End}(E)$. Locally, say on U, sections σ_1 and σ_2 of $\Omega(M, \mathrm{End}(E))$ take the form $\sigma_1 = A_i \otimes \alpha^i$ and $\sigma_2 = B_i \otimes \beta^i$ where A_i and B_i are maps $U \to \mathrm{End}(E)$. Thus for each $x \in U$, the A_i and B_i evaluate to give $A_i(x), B_i(x) \in \mathrm{End}(E_x)$. The multiplication is then

$$(A_i \otimes \alpha^i) \wedge (B_j \otimes \beta^j) = A_i B_j \otimes \alpha^i \wedge \beta^j$$

where the $A_i B_j : U \to \mathrm{End}(E)$ are local sections given by composition:

$$A_i B_j : x \mapsto A_i(x) \circ B_j(x).$$

Exercise 10.8 *Show that $\Omega(M, \mathrm{End}(E))$ acts on $\Omega(M, E)$ making $\Omega(M, E)$ a bundle of modules over the bundle of algebras $\Omega(M, \mathrm{End}(E))$.*

If this seems all to abstract to the newcomer perhaps it would help to think of things this way: We have a cover of a manifold M by open sets $\{U_\alpha\}$ which simultaneously trivialize both E and TM. Then these give also trivializations on these open sets of the bundles $\mathrm{Hom}(E, E)$ and $\wedge TM$. Associated with each is a frame field for $E \to M$ say $(e_1, ..., e_k)$ which allows us to associate with each section $\sigma \in \Omega^p(M, E)$ a $k-$tuple of $p-$forms $\sigma_U = (\sigma^i_U)$ for each U. Similarly, a section $A \in \Omega^q(M, \mathrm{End}(E))$ is equivalent to assigning to each open set $U \in \{U_\alpha\}$ a matrix of $q-$forms A_U. The algebra structure on $\Omega(M, \mathrm{End}(E))$ is then just matrix multiplication were the entries are multiplies using the wedge product $A_U \wedge B_U$ where

$$(A_U \wedge B_U)^i_j = A^i_k \wedge B^k_j$$

The module structure is given locally by $\sigma_U \mapsto A_U \wedge \sigma_U$. Where did the bundle go? The global topology is now encoded in the transformation laws which tell us what the same section look like when we change to a new from on an overlap $U_\alpha \cap U_\beta$? In this sense the bundle is a combinatorial recipe for pasting together local objects.

Chapter 11

Distributions and Frobenius' Theorem

11.1 Definitions

In this section we take M to be a C^∞ manifold modelled on a Banach space M. Roughly speaking, smooth distribution is an assignment $\triangle$of a subspace $\triangle_p \subset T_pM$ to each $p \in M$ such that for each $p \in M$ there is a family of smooth vector fields $X_1, ..., X_k$ defined on some neighborhood U_p of p and such that $\triangle_x = span\{X_1(x), ..., X_k(x)\}$ for each $x \in U_p$. We call the distribution regular iff we can always choose the vector fields to be linearly independent on each tangent space T_xM for $x \in U_p$ and each U_p. It follows that in this case k is locally constant. For a regular distribution k is called the rank of the distribution. A rank k regular distribution is the same think as a rank k subbundle of the tangent bundle. We can also consider regular distributions of infinite rank by simply defining such to be a subbundle of the tangent bundle.

Definition 11.1 *A (smooth) regular distribution on a manifold M is a smooth vector subbundle of the tangent bundle TM.*

11.2 Integrability of Regular Distributions

By definition a regular distribution $\triangle$ is just another name for a subbundle $\triangle \subset TM$ of the tangent bundle and we write $\triangle_p \subset T_pM$ for the fiber of the subbundle at p. So what we have is a smooth assignment of a subspace $\triangle_p$ at every point. The subbundle definition guarantees that the spaces $\triangle_p$ all have the same dimension (if finite) in each connected component of M. This dimension is called the rank of the distribution. There is a more general notion of distribution which we call a singular distribution which is defined in the same way except for the requirement of constancy of dimension. We shall study singular distributions later.

Definition 11.2 *Let X locally defined vector field. We say that X **lies in** the distribution $\triangle$ if $X(p) \in \triangle_p$ for each p in the domain of X. In this case, we write $X \in \triangle$ (a slight abuse of notation).*

Note that in the case of a regular distribution we can say that for X to lie in the distribution $\triangle$ means that X takes values in the subbundle $\triangle \subset TM$.

Definition 11.3 *We say that a locally defined differential j-form ω **vanishes on** $\triangle$ if for every choice of vector fields $X_1, ..., X_j$ defined on the domain of ω that lie in $\triangle$ the function $\omega(X_1, ..., X_j)$ is identically zero.*

For a regular distribution $\triangle$ consider the following two conditions.

Fro1 For every pair of locally defined vector fields X and Y with common domain that lie in the distribution $\triangle$ the bracket $[X, Y]$ also lies in the distribution.

Fro2 For each locally defined smooth 1-form ω that vanishes on $\triangle$ the 2-form $d\omega$ also vanishes on $\triangle$.

Lemma 11.1 *Conditions (1) and (2) above are equivalent.*

Proof. The proof that these two conditions are equivalent follows easily from the formula

$$d\omega(X, Y) = X(\omega(Y)) - Y\omega(X) - \omega([X, Y]).$$

Suppose that (1) holds. If ω vanishes on $\triangle$ and X, Y lie in $\triangle$ then the above formula becomes

$$d\omega(X, Y) = -\omega([X, Y])$$

which shows that $d\omega$ vanishes on $\triangle$ since $[X, Y] \in \triangle$ by condition (1). Conversely, suppose that (2) holds and that $X, Y \in \triangle$. Then $d\omega(X, Y) = -\omega([X, Y])$ again and a local argument using the Hahn-Banach theorem shows that $[X, Y] = 0$. ■

Definition 11.4 *If either of the two equivalent conditions introduced above holds for a distribution $\triangle$ then we say that $\triangle$ is **involutive**.*

Exercise 11.1 *Suppose that $\mathcal{X}$ is a family of locally defined vector fields of M such that for each $p \in M$ and each local section X of the subbundle $\triangle$ defined in a neighborhood of p, there is a finite set of local fields $\{X_i\} \subset \mathcal{X}$ such that $X = \sum a^i X_i$ on some possible smaller neighborhood of p. Show that if $\mathcal{X}$ is closed under bracketing then $\triangle$ is involutive.*

There is a very natural way for distributions to arise. For instance, consider the punctured 3-space $M = \mathbb{R}^3 - \{0\}$. The level sets of the function $\varepsilon : (x, y, x) \mapsto x^2 + y^2 + x^2$ are spheres whose union is all of $\mathbb{R}^3 - \{0\}$. Now define a distribution by the rule that $\triangle_p$ is the tangent space at p to the sphere containing p. Dually, we can define this distribution to be the given by the rule

$$\triangle_p = \{v \in T_pM : d\varepsilon(v) = 0\}.$$

The main point is that each p contains a submanifold S such that $\triangle_x = T_xS$ for all $x \in S \cap U$ for some sufficiently small open set $U \subset M$. On the other hand, not all distributions arise in this way.

Definition 11.5 *A distribution $\triangle$ on M is called* ***integrable at*** *$p \in M$ there is a submanifold S_p containing p such that $\triangle_x = T_xS_p$ for all $x \in S$. (Warning: S_p is locally closed but not necessarily a closed subset and may only be defined very near p.) We call such submanifold a* ***local integral submanifold*** *of $\triangle$.*

Definition 11.6 *A regular distribution $\triangle$ on M is called* ***(completely) integrable*** *if for every $p \in M$ there is a (local) integral submanifold of $\triangle$ containing p.*

If one considers a distribution on a finite dimensional manifold there is a nice picture of the structure of an integrable distribution. Our analysis will eventually allow us to see that a regular distribution $\triangle$ of rank k on an n-manifold M is (completely) integrable if and only if there is a cover of M by charts ψ_a, U_a such that if $\psi_a = (y^1, ..., y^n)$ then for each $p \in U_a$ the submanifold $S_{\alpha,p}$ defined by $S_{\alpha,p} := \{x \in U_a : y^i(x) = y^i(p) \text{ for } k+1 \leq i \leq n\}$ has

$$\triangle_x = T_xS_{\alpha,p} \text{ for all } x \in S_p.$$

Some authors use this as the definition of integrable distribution but this definition would be inconvenient to generalize to the infinite dimensional case. A main goal of this section is to prove the theorem of Frobenius which says that a regular distribution is integrable if and only if it is involutive.

11.3 The local version Frobenius' theorem

Here we study regular distributions; also known as tangent subbundles. The presentation draws heavily on that given in [L1]. Since in the regular case a distribution is a subbundle of the tangent bundle it will be useful to consider such subbundle a little more carefully. Recall that if $E \to M$ is a subbundle of TM then $E \subset TM$ and there is an atlas of adapted VB-charts for TM; that is, charts $\phi : \tau_M^{-1}(U) \to U \times \mathsf{M} = U \times \mathsf{E} \times \mathsf{F}$ where $\mathsf{E} \times \mathsf{F}$ is a fixed splitting of M. Thus M is modelled on the split space $\mathsf{E} \times \mathsf{F} = \mathsf{M}$. Now for all local questions we may assume that in fact the tangent bundle is a trivial bundle of the form $(U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F})$ where $U_1 \times U_2 \subset \mathsf{E} \times \mathsf{F}$. It is easy to see that our subbundle must now consist of a choice of subspace $\mathsf{E}_1(x, y)$ of $(\mathsf{E} \times \mathsf{F})$ for every $(x, y) \in U_1 \times U_2$. In fact, the inverse of our trivialization gives a map

$$\phi^{-1} : (U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F}) \to (U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F})$$

such that the image under ϕ^{-1} of $\{(x,y)\}\times\mathsf{E}\times\{0\}$ is exactly $\{(x,y)\}\times\mathsf{E}_1(x,y)$. The map ϕ^{-1} must have the form

$$\phi((x,y),v,w) = ((x,y), f_{(x,y)}(v,w), g_{(x,y)}(v,w))$$

for where $f_{(x,y)} : \mathsf{E}\times\mathsf{F}\to\mathsf{E}$ and $g_{(x,y)} : \mathsf{E}\times\mathsf{F}\to\mathsf{E}$ are linear maps depending smoothly on (x,y). Furthermore, for all (x,y) the map $f_{(x,y)}$ takes $\mathsf{E}\times\{0\}$ isomorphically onto $\{(x,y)\}\times\mathsf{E}_1(x,y)$. Now the composition

$$\kappa : (U_1\times U_2)\times\mathsf{E}\hookrightarrow (U_1\times U_2)\times(\mathsf{E}\times\{0\}) \stackrel{\phi^{-1}}{\to} (U_1\times U_2)\times(\mathsf{E}\times\mathsf{F})$$

maps $\{(x,y)\}\times\mathsf{E}$ isomorphically onto $\{(x,y)\}\times\mathsf{E}_1(x,y)$ and must have form

$$\kappa(x,y,v) = (x,y,\lambda(x,y)\cdot v, \ell(x,y)\cdot v)$$

for some smooth maps $(x,y)\mapsto\lambda(x,y)\in L(\mathsf{E},\mathsf{E})$ and $(x,y)\mapsto\ell(x,y)\in L(\mathsf{E},\mathsf{F})$. By a suitable "rotation" of the space $\mathsf{E}\times\mathsf{F}$ for each (x,y) we may assume that $\lambda_{(x,y)} = \mathrm{id}_{\mathsf{E}}$. Now for fixed $v\in\mathsf{E}$ the map $X_v : (x,y)\mapsto(x,y,v,\ell_{(x,y)}v)$ is (a local representation of) a vector field with values in the subbundle E. The principal part is $\mathsf{X}_v(x,y) = (v,\ell_{(x,y)}\cdot v)$.

Now $\ell(x,y)\in L(\mathsf{E},\mathsf{F})$ and so $D\ell(x,y)\in L(\mathsf{E}\times\mathsf{F}, L(\mathsf{E},\mathsf{F}))$. In general for a smooth family of linear maps Λ_u and a smooth map $v : (x,y)\mapsto v(x,y)$ we have

$$D(\Lambda_u\cdot v)(w) = D\Lambda_u(w)\cdot v + \Lambda_u\cdot(Dv)(w)$$

and so in the case at hand

$$\begin{aligned}&D(\ell(x,y)\cdot v)(w_1,w_2)\\ &= (D\ell(x,y)(w_1,w_2))\cdot v + \ell(x,y)\cdot(Dv)(w_1,w_2).\end{aligned}$$

For any two choices of smooth maps v_1 and v_2 as above we have

$$\begin{aligned}[\mathsf{X}_{v_1},\mathsf{X}_{v_2}]_{(x,y)} &= (D\mathsf{X}_{v_2})_{(x,y)}\mathsf{X}_{v_1}(x,y) - (D\mathsf{X}_{v_1})_{(x,y)}\mathsf{X}_{v_2}(x,y)\\ &= ((Dv_2)(v_1,\ell_{(x,y)}v_1) - (Dv_1)(v_2,\ell_{(x,y)}v_2), D\ell(x,y)(v_1,\ell_{(x,y)}v_1)\cdot v_2\\ &+ \ell(x,y)\cdot(Dv_2)(v_1,\ell_{(x,y)}v_1) - D\ell(x,y)(v_2,\ell_{(x,y)}v_2)\cdot v_1\\ &- \ell(x,y)\cdot(Dv_1)(v_2,\ell_{(x,y)}v_2))\\ &= (\xi, D\ell(x,y)(v_1,\ell_{(x,y)}v_1)\cdot v_2 - D\ell(x,y)(v_2,\ell_{(x,y)}v_2)\cdot v_1 + \ell(x,y)\cdot\xi).\end{aligned}$$

where $\xi = (Dv_2)(v_1,\ell_{(x,y)}v_1) - (Dv_1)(v_2,\ell_{(x,y)}v_2)$. Thus $[\mathsf{X}_{v_1},\mathsf{X}_{v_2}]_{(x,y)}$ is in the subbundle iff

$$D\ell(x,y)(v_1,\ell_{(x,y)}v_1)\cdot v_2 - D\ell(x,y)(v_2,\ell_{(x,y)}v_2)\cdot v_1.$$

We thus arrive at the following characterization of involutivity:

Lemma 11.2 *Let Δ be a subbundle of TM. For every $p\in M$ there is a tangent bundle chart containing T_pM of the form described above so that any vector field* *vector field taking values in the subbundle is represented as a map* $\mathsf{X}_v : U_1 \times U_2 \to \mathsf{E} \times \mathsf{F}$ *of the form* $(x, y) \mapsto (v(x, y), \ell_{(x,y)} v(x, y))$*. Then* Δ *is involutive (near* p*) iff for any two smooth maps* $v_1 : U_1 \times U_2 \to \mathsf{E}$ *and* $v_2 : U_1 \times U_2 \to \mathsf{E}$ *we have*

$$D\ell(x, y)(v_1, \ell_{(x,y)} v_1) \cdot v_2 - D\ell(x, y)(v_2, \ell_{(x,y)} v_2) \cdot v_1.$$

Theorem 11.1 *A regular distribution* Δ *on* M *is integrable if and only if it is involutive.*

Proof. First suppose that Δ is integrable. Let X and Y be local vector fields that lie in Δ. Pick a point x in the common domain of X and Y. Our choice of x being arbitrary we just need to show that $[X, Y](x) \in \Delta$. Let $S \subset M$ be a local integral submanifold of Δ containing the point x. The restrictions $X|_S$ and $Y|_S$ are related to X and Y by an inclusion map and so by the result on related vector fields we have that $[X|_S, Y|_S] = [X, Y]|_S$ on some neighborhood of x. Since S is a manifold and $[X|_S, Y|_S]$ a local vector field on S we see that $[X, Y]|_S(x) = [X, Y](x)$ is tangent to S and so $[X, Y](x) \in \Delta$. Suppose now that Δ is involutive. Since this is a local question we may assume that our tangent bundle is a trivial bundle $(U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F})$ and by our previous lemma we know that for any two smooth maps $v_1 : U_1 \times U_2 \to \mathsf{E}$ and $v_2 : U_1 \times U_2 \to \mathsf{E}$ we have

$$D\ell(x, y)(v_1, \ell_{(x,y)} v_1) \cdot v_2 - D\ell(x, y)(v_2, \ell_{(x,y)} v_2) \cdot v_1.$$

Claim 11.1 *For any* $(x_0, y_0) \in U_1 \times U_2$ *there exists possibly smaller open product* $U_1' \times U_2' \subset U_1 \times U_2$ *containing* (x_0, y_0) *and a unique smooth map* $\alpha : U_1' \times U_2' \to U_2$ *such that* $\alpha(x_0, y) = y$ *for all* $y \in U_2'$ *and*

$$D_1\alpha(x, y) = \ell(x, \alpha(x, y))$$

for all $(x, y) \in U_1' \times U_2'$*.*

Before we prove this claim we show how the result follows from it. For any $y \in U_2'$ we have the partial map $\alpha_y(x) := \alpha(x, y)$ and equation **??** above reads $D\alpha_y(x, y) = \ell(x, \alpha_y(x))$. Now if we define the map $\phi : U_1' \times U_2' \to U_1 \times U_2$ by $\phi(x, y) := (x, \alpha_y(x))$ then using this last version of equation **??** and the condition $\alpha(x_0, y) = y$ from the claim we see that

$$\begin{aligned} D_2\alpha(x_0, y_0) &= D\alpha(x_0, .)(y_0) \\ &= D\,\mathrm{id}_{U_2'} = \mathrm{id}\,. \end{aligned}$$

Thus the Jacobian of ϕ at (x_0, y_0) has the block form

$$\begin{pmatrix} \mathrm{id} & 0 \\ * & \mathrm{id} \end{pmatrix}.$$

By the inverse function theorem ϕ is a local diffeomorphism in a neighborhood of (x_0, y_0). We also have that

$$\begin{aligned} (D_1\phi)(x, y) \cdot (v, w) &= (v, D\alpha_y(x) \cdot w) \\ &= (v, \ell(x, \alpha_y(x)) \cdot v). \end{aligned}$$

Which is the form of elements of the subbundle but is also the form of tangents to the submanifolds which are the images of $U_1' \times \{y\}$ under the diffeomorphism ϕ for various choices of $y \in U_2$. This clearly saying that the subbundle is integrable.

Proof of the claim: By translation we may assume that $(x_0, y_0) = (0, 0)$. We use theorem 26.18 from appendix B. With the notation of that theorem we let $f(t, x, y) := \ell(tz, y) \cdot z$ where $y \in U_2$ and z is an element of some ball $B(0, \epsilon)$ in E. Thus the theorem provides us with a smooth map $\beta : J_0 \times B(0, \epsilon) \times U_2$ satisfying $\beta(0, z, y) = y$ and

$$\frac{\partial}{\partial t}\beta(t, z, y) = \ell(tz, \beta(t, z, y)) \cdot z.$$

We will assume that $1 \in J$ since we can always arrange for this by making a change of variables of the type $t = as, z = x/a$ for a sufficiently small positive number a (we may do this at the expense of having to choose a smaller ϵ for the ball $B(0, \epsilon)$. We claim that if we define

$$\alpha(x, y) := \beta(1, x, y)$$

then for sufficiently small $|x|$ we have the required result. In fact we shall show that

$$D_2\beta(t, z, y) = t\ell(tz, \beta(t, z, y))$$

from which it follows that

$$D_1\alpha(x, y) = D_2\beta(1, x, y) = \ell(x, \alpha(x, y))$$

with the correct initial conditions (recall that we translated to (x_0, y_0)). Thus it remains to show that equation **??** holds. From (3) of theorem 26.18 we know that $D_2\beta(t, z, y)$ satisfies the following equation for any $v \in \mathsf{E}$:

$$\begin{aligned}\frac{\partial}{\partial t}D_2\beta(t, z, y) = t\frac{\partial}{\partial t}&\ell(tz, \beta(t, z, y)) \cdot v \cdot z \\ &+ D_2\ell(tz, \beta(t, z, y)) \cdot D_2\beta(t, z, y) \cdot v \cdot z \\ &+ \ell(tz, \beta(t, z, y)) \cdot v.\end{aligned}$$

Now we fix everything but t and define a function of one variable:

$$\Phi(t) := D_2\beta(t, z, y) \cdot v - t\ell(tz, \beta(t, z, y).$$

Clearly, $\Phi(0) = 0$. Now we use two fixed vectors v,z and construct the fields $\mathsf{X}_v(x, y) = (v, \ell_{(x,y)} \cdot v)$ and $\mathsf{X}_z(x, y) = (z, \ell_{(x,y)} \cdot z)$. In this special case, the equation of lemma 11.2 becomes

$$D\ell(x, y)(v, \ell_{(x,y)}v) \cdot z - D\ell(x, y)(z, \ell_{(x,y)}z) \cdot v.$$

Now with this in mind we compute $\frac{d}{dt}\Phi(t)$:

$$\begin{aligned}
\frac{d}{dt}\Phi(t) &= \frac{\partial}{\partial t}(D_2\beta(t,z,y)\cdot v - t\ell(tz,\beta(t,z,y)) \\
&= \frac{\partial}{\partial t}D_2\beta(t,z,y)\cdot v - t\frac{d}{dt}\ell(tz,\beta(t,z,y)) - \ell(tz,\beta(t,z,y)) \\
&= \frac{\partial}{\partial t}D_2\beta(t,z,y)\cdot v - t\{D_1\ell(tz,\beta(t,z,y))\cdot z \\
&+ D_2\ell(tz,\beta(t,z,y))\cdot \frac{\partial}{\partial t}\beta(t,z,y) - \ell(tz,\beta(t,z,y)) \\
&= D_2\ell(tz,\beta(t,z,y))\cdot\{D_2\beta(t,z,y)\cdot v - t\ell(tz,\beta(t,z,y)\}\cdot z \text{ (use 11.3)} \\
&= D_2\ell(tz,\beta(t,z,y))\cdot\Phi(t)\cdot z.
\end{aligned}$$

So we arrive at $\frac{d}{dt}\Phi(t) = D_2\ell(tz,\beta(t,z,y))\cdot\Phi(t)\cdot z$ with initial condition $\Phi(0)=0$ which implies that $\Phi(t)\equiv 0$. This latter identity is none other than $D_2\beta(t,z,y)\cdot v = t\ell(tz,\beta(t,z,y)$. ∎

It will be useful to introduce the notion of a co-distribution and then explore the dual relationship existing between distributions and co-distributions.

Definition 11.7 *A (regular)* ***co-distribution*** Ω *on a manifold* M *is a subbundle of the cotangent bundle. Thus a smooth assignment of a subspace* $\Omega_x \subset T_x^*M$ *for every* $x \in M$*. If* $\dim \Omega_x = l < \infty$ *we call this a rank* l *co-distribution.*

Using the definition of vector bundle chart adapted to a subbundle it is not hard to show , as indicated in the first paragraph of this section, that a (smooth) distribution of rank $k < \infty$ can be described in the following way:

Claim 11.2 *For a smooth distribution* $\triangle$ *of rank on* M *we have that for every* $p \in M$ *there exists a family of smooth vector fields* $X_1, ..., X_k$ *defined near* p *such that* $\triangle_x = \operatorname{span}\{X_1(x), ..., X_k(x)\}$ *for all* x *near* p*.*

Similarly, we have

Claim 11.3 *For a smooth co-distribution* Ω *of rank* k *on* M *we have that for every* $p \in M$ *there exists a family of smooth 1-forms fields* $\omega_1, ..., \omega_k$ *defined near* p *such that* $\Omega_x = \operatorname{span}\{\omega_1(x), ..., \omega_k(x)\}$ *for all* x *near* p*.*

On the other hand we can use a co-distribution to define a distribution and visa-versa. For example, for a regular co-distribution Ω on M we can define a distribution $\triangle^{\perp\Omega}$ by

$$\triangle_x^{\perp\Omega} := \{v \in T_xM : \omega_x(v) = 0 \text{ for all } \omega_x \in \Omega_x\}.$$

Similarly, if $\triangle$ is a regular distribution on M then we can define a co-distribution $\Omega^{\perp\triangle}$ by

$$\Omega_x^{\perp\triangle} := \{\omega_x \in T_x^*M : \omega_x(v) = 0 \text{ for all } v \in \triangle_x\}.$$

Notice that if $\triangle_1 \subset \triangle_2$ then $\triangle_2^{\perp\Omega} \subset \triangle_1^{\perp\Omega}$ and $(\triangle_1 \cap \triangle_2)^{\perp\Omega} = \triangle_1^{\perp\Omega} + \triangle_2^{\perp\Omega}$ etc.

11.4 Foliations

Definition 11.8 *Let M be a smooth manifold modelled on M and assume that $\mathsf{M} = \mathsf{E} \times \mathsf{F}$. A **foliation** $\mathcal{F}_M$ of M (or on M) is a partition of M into a family of disjoint subsets connected $\{\mathcal{L}_\alpha\}_{\alpha \in A}$ such that for every $p \in M$, there is a chart centered at p of the form $\varphi : U \to V \times W \subset \mathsf{E} \times \mathsf{F}$ with the property that for each $\mathcal{L}_\alpha$ the connected components $(U \cap \mathcal{L}_\alpha)_\beta$ of $U \cap \mathcal{L}_\alpha$ are given by*

$$\varphi((U \cap \mathcal{L}_\alpha)_\beta) = V \times \{c_{\alpha,\beta}\}$$

*where $c_{\alpha,\beta} \in W \subset \mathsf{F}$ are constants. These charts are called distinguished charts for the foliation or **foliation charts**. The connected sets $\mathcal{L}_\alpha$ are called the **leaves** of the foliation while for a given chart as above the connected components $(U \cap \mathcal{L}_\alpha)_\beta$ are called **plaques**.*

Recall that the connected components $(U \cap \mathcal{L}_\alpha)_\beta$ of $U \cap \mathcal{L}_\alpha$ are of the form $C_x(U \cap \mathcal{L}_\alpha)$ for some $x \in \mathcal{L}_\alpha$. An important point is that a fixed leaf $\mathcal{L}_\alpha$ may intersect a given chart domain U in many, even an infinite number of disjoint connected pieces no matter how small U is taken to be. In fact, it may be that $C_x(U \cap \mathcal{L}_\alpha)$ is dense in U. On the other hand, each $\mathcal{L}_\alpha$ is connected by definition. The usual first example of this behavior is given by the irrationally foliated torus. Here we take $M = T^2 := S^1 \times S^1$ and let the leaves be given as the image of the immersions $\iota_a : t \mapsto (e^{\mathrm{i}at}, e^{\mathrm{i}t})$ where a is a real numbers. If a is irrational then the image $\iota_a(\mathbb{R})$ is a (connected) dense subset of $S^1 \times S^1$. On the other hand, even in this case there are an infinite number of distinct leaves.

It may seem that a foliated manifold is just a special manifold but from one point of view a foliation is a generalization of a manifold. For instance, we can think of a manifold M as foliation where the points are the leaves. This is called the **discrete foliation** on M. At the other extreme a manifold may be thought of as a foliation with a single leaf $\mathcal{L} = M$ (the **trivial foliation**). We also have handy many examples of 1-dimensional foliations since given any global flow the orbits (maximal integral curves) are the leaves of a foliation. We also have the following special cases:

Example 11.1 *On a product manifold say $M \times N$ we have two complementary foliations:*

$$\{\{p\} \times N\}_{p \in M}$$

and

$$\{M \times \{q\}\}_{q \in N}$$

Example 11.2 *Given any submersion $f : M \to N$ the level sets $\{f^{-1}(q)\}_{q \in N}$ form the leaves of a foliation. The reader will soon realize that any foliation is given locally by submersions. The global picture for a general foliation can be very different from what can occur with a single submersion.*

Example 11.3 *The fibers of any vector bundle foliate the total space.*

Example 11.4 *(Reeb foliation) Consider the strip in the plane given by* $\{(x, y) : |x| \leq 1\}$. *For* $a \in \mathbb{R} \cup \{\pm\infty\}$ *we form leaves* $\mathcal{L}_a$ *as follows:*

$$\mathcal{L}_a := \{(x, a + f(x)) : |x| \leq 1\} \text{ for } a \in \mathbb{R}$$
$$\mathcal{L}_{\pm\infty} := \{(\pm 1, y) : |y| \leq 1\}$$

where $f(x) := \exp\left(\frac{x^2}{1-x^2}\right) - 1$. *By rotating this symmetric foliation about the* y *axis we obtain a foliation of the solid cylinder. This foliation is such that translation of the solid cylinder* C *in the* y *direction maps leaves diffeomorphically onto leaves and so we may let* $\mathbb{Z}$ *act on* C *by* $(x, y, z) \mapsto (x, y + n, z)$ *and then* $C/\mathbb{Z}$ *is a solid torus with a foliation called the Reeb foliation.*

Example 11.5 *The one point compactification of* $\mathbb{R}^3$ *is homeomorphic to* $S^3 \subset \mathbb{R}^4$. *Thus* $S^3 - \{p\} \cong \mathbb{R}^3$ *and so there is a copy of the Reeb foliated solid torus inside* S^3. *The complement of a solid torus in* S^3 *is another solid torus. It is possible to put another Reeb foliation on this complement and thus foliate all of* S^3. *The only compact leaf is the torus which is the common boundary of the two complementary solid tori.*

Exercise 11.2 *Show that the set of all* $v \in TM$ *such that* $v = T\varphi^{-1}(\mathsf{v}, 0)$ *for some* $\mathsf{v} \in \mathsf{E}$ *and some foliated chart* φ *is a (smooth) subbundle of* TM *which is also equal to* $\{v \in TM : v \text{ is tangent to a leaf}\}$.

Definition 11.9 *The* ***tangent bundle of a foliation*** $\mathcal{F}_M$ *with structure pseudogroup* $\Gamma_{\mathsf{M,F}}$ *is the subbundle* $T\mathcal{F}_M$ *of* TM *defined by*

$$\begin{aligned} T\mathcal{F}_M &:= \{v \in TM : v \text{ is tangent to a leaf}\} \\ &= \{v \in TM : v = T\varphi^{-1}(\mathsf{v}, 0) \text{ for some } \mathsf{v} \in \mathsf{E} \text{ and some foliated chart } \varphi\} \end{aligned}$$

11.5 The Global Frobenius Theorem

The first step is to show that the (local) integral submanifolds of an integrable regular distribution can be glued together to form maximal integral submanifolds. These will form the leaves of a distribution.

Exercise 11.3 *If* Δ *is an integrable regular distribution of* TM, *then for any two local integral submanifolds* S_1 *and* S_2 *of* Δ *that both contain a point* x_0, *there is an open neighborhood* U *of* x_0 *such that*

$$S_1 \cap U = S_2 \cap U$$

Theorem 11.2 *If* Δ *is a subbundle of* TM *(i.e. a regular distribution) then the following are equivalent:*

1) Δ *is involutive.*
2) Δ *is integrable.*
3) There is a foliation $\mathcal{F}_M$ *on* M *such that* $T\mathcal{F}_M = \Delta$.

Proof. The equivalence of (1) and (2) is the local Frobenius theorem already proven. Also, the fact that (3) implies (2) is follows from 11.2. Finally, assume that (2) holds so that Δ is integrable. Recall that each (local) integral submanifold is an immersed submanifold which carries a submanifold topology generated by the connected components of the intersections of the integral submanifolds with chart domains. Any integral submanifold S has a smooth structure given by restricting charts U, ψ on M to connected components of $S \cap U$ (not on all of $S \cap U$!). Recall that a local integral submanifold is a regular submanifold (we are not talking about maximal immersed integral submanifolds!). Thus we may take U small enough that $S \cap U$ is connected. Now if we take two (local) integral submanifolds S_1 and S_2 of Δ and any point $x_0 \in S_1 \cap S_2$ (assuming this is nonempty) then a small enough chart U, ψ with $x_0 \in U$ induces a chart $U \cap S_1, \psi|_{U\cap S_1}$ on S_1 and a chart $C_{x_0}(U \cap S_2), \psi|_{C_{x_0}(U\cap S_2)}$ on S_2. But as we know $S_1 \cap U = S_2 \cap U$ and the overlap is smooth. Thus the union $S_1 \cup S_2$ is a smooth manifold with charts given by $U \cap (S_1 \cup S_2), \psi|_{U\cap(S_1\cup S_2)}$ and the overlap maps are $U \cap (S_1 \cap S_2), \psi|_{U\cap(S_1\cap S_2)}$. We may extend to a maximal connected integral submanifold using Zorn's lemma be we can see the existence more directly. Let $\mathcal{L}_a(x_0)$ be the set of all points that can be connected to x_0 by a smooth path $c : [0, 1] \to M$ with the property that for any $t_0 \in [0, 1]$, the image $c(t)$ is contained inside a (local) integral submanifold for all t sufficiently near t_0. Using what we have seen about gluing intersecting integral submanifold together and the local uniqueness of such integral submanifolds we see that $\mathcal{L}_a(x_0)$ is a smooth connected immersed integral submanifold that must be the maximal connected integral submanifold containing x_0. Now since x_0 was arbitrary there is a maximal connected integral submanifold containing any point of M. By construction we have that the foliation $\mathcal{L}$ given by the union of all these leaves satisfies (3). ∎

There is an analogy between the notion of a foliation on a manifold and a differentiable structure on a manifold. In order to see this more clearly it will be useful to introduce the notion of a pseudogroup. The example to keep in mind while reading the following definition is the family of all *locally defined* diffeomorphisms $\vartheta :: M \to M$.

With a little thought, it should be fairly clear that a foliation is the same thing as a manifold with a $\Gamma_{\mathsf{M,F}}$ structure and the atlas described in the definition just given is a foliated atlas.

From this point of view we think of a foliation as being given by a maximal **foliation atlas** which is defined to be a cover of M by foliated charts. The compatibility condition on such charts is that when the domains of two foliation charts, say $\varphi_1 : U_1 \to V_1 \times W_1$ and $\varphi_2 : U_2 \to V_2 \times W_2$, then the overlap map has the form

$$\varphi_2 \circ \varphi_1^{-1}(\mathsf{x}, \mathsf{y}) = (f(\mathsf{x}, \mathsf{y}), g(\mathsf{y})).$$

A plaque in a chart $\varphi_1 : U_1 \to V_1 \times W_1$ is a connected component of a set of the form $\varphi_1^{-1}\{(\mathsf{x}, \mathsf{y}) : \mathsf{y} = \text{constant}\}$.

11.6 Singular Distributions

Lemma 11.3 *Let $X_1, ..., X_n$ be vector fields defined in a neighborhood of $x \in M$ such that $X_1(x), ..., X_n(x)$ are a basis for T_xM and such that $[X_i, X_j] = 0$ in a neighborhood of x. Then there is an open chart $U, \psi = (y^1, ..., y^n)$ containing x such that $X_i|_U = \frac{\partial}{\partial y^i}$.*

Proof. For a sufficiently small ball $B(0, \epsilon) \subset \mathbb{R}^n$ and $t = (t_1, ..., t_n) \in B(0, \epsilon)$ we define

$$f(t_1, ..., t_n) := Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x).$$

By theorem 7.10 the order that we compose the flows does not change the value of $f(t_1, ..., t_n)$. Thus

$$\begin{aligned}
&\frac{\partial}{\partial t_i} f(t_1, ..., t_n) \\
&= \frac{\partial}{\partial t_i} Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x) \\
&= \frac{\partial}{\partial t_i} Fl_{t_i}^{X_i} \circ Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x) \text{ (put the i-th flow first)} \\
&X_i(Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x)).
\end{aligned}$$

Evaluating at $t = 0$ shows that $T_0 f$ is nonsingular and so $(t_1, ..., t_n) \mapsto f(t_1, ..., t_n)$ is a diffeomorphism on some small open set containing 0. The inverse of this map is the coordinate chart we are looking for (check this!). ∎

Definition 11.10 *Let $\mathfrak{X}_{loc}(M)$ denote the set of all sections of the presheaf $\mathfrak{X}_M$. That is*

$$\mathfrak{X}_{loc}(M) := \bigcup_{open\ U \subset M} \mathfrak{X}_M(U).$$

Also, for a distribution Δ let $\mathfrak{X}_\Delta(M)$ denote the subset of $\mathfrak{X}_{loc}(M)$ consisting of local fields X with the property that $X(x) \in \Delta_x$ for every x in the domain of X.

Definition 11.11 *We say that a subset of local vector fields $\mathcal{X} \subset \mathfrak{X}_\Delta(M)$ **spans** a distribution Δ if for each $x \in M$ the subspace Δ_x is spanned by $\{X(x) : X \in \mathcal{X}\}$.*

If Δ is a smooth distribution (and this is all we shall consider) then $\mathfrak{X}_\Delta(M)$ spans Δ. On the other hand as long as we make the convention that the empty set spans the set $\{0\}$ for what every vector space we are considering then any $\mathcal{X} \subset \mathfrak{X}_\Delta(M)$ spans some smooth distribution which we denote by $\Delta(\mathcal{X})$.

Definition 11.12 *An immersed integral submanifold of a distribution Δ is an injective immersion $\iota : S \to M$ such that $T_s\iota(T_sS) = \Delta_{\iota(s)}$ for all $s \in S$. An immersed integral submanifold is called **maximal** its image is not properly contained in the image of any other immersed integral submanifold.*

Since an immersed integral submanifold is an injective map we can think of S as a subset of M. In fact, it will also turn out that an immersed integral submanifold is automatically smoothly universal so that the image $\iota(S)$ is an initial submanifold. Thus in the end, we may as well assume that $S \subset M$ and that $\iota : S \to M$ is the inclusion map. Let us now specialize to the finite dimensional case. Note however that we do *not* assume that the rank of the distribution is constant.

Now we proceed with our analysis. If $\iota : S \to M$ is an immersed integral submanifold and of a distribution $\triangle$ then if $X \in \mathfrak{X}_\Delta(M)$ we can make sense of $\iota^* X$ as a local vector field on S. To see this let U be the domain of X and take $s \in S$ with $\iota(s) \in U$. Now $X(\iota(s)) \in T_s\iota(T_s S)$ we can define

$$\iota^* X(s) := (T_s\iota)^{-1} X(\iota(s)).$$

$\iota^* X(s)$ is defined on some open set in S and is easily seen to be smooth by considering the local properties of immersions. Also, by construction $\iota^* X$ is ι related to X.

Next we consider what happens if we have two immersed integral submanifolds $\iota_1 : S_1 \to M$ and $\iota_2 : S_2 \to M$ such that $\iota_1(S_1) \cap \iota_2(S_2) \neq \emptyset$. By proposition 7.1 we have

$$\iota_i \circ \mathrm{Fl}_t^{\iota_i^* X} = \mathrm{Fl}_t^X \circ \iota_i \text{ for } i = 1, 2.$$

Now if $x_0 \in \iota_1(S_1) \cap \iota_2(S_2)$ then we choose s_1 and s_2 such that $\iota_1(s_1) = \iota_2(s_2) = x_0$ and pick local vector fields $X_1, ..., X_k$ such that $(X_1(x_0), ..., X_k(x_0))$ is a basis for $\triangle_{x_0}$. For $i = 1$ and 2 we define

$$f_i(t^1, ..., t^k) := (\mathrm{Fl}_{t^1}^{\iota_i^* X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{\iota_i^* X_k})$$

and since $\frac{\partial}{\partial t^j}\big|_0 f_i = \iota_i^* X_j$ for $i = 1, 2$ and $j = 1, ..., k$ we conclude that f_i, $i = 1, 2$ are diffeomorphisms when suitable restricted to a neighborhood of $0 \in \mathbb{R}^k$. Now we compute:

$$\begin{aligned}
(\iota_2^{-1} \circ \iota_1 \circ f_1)(t^1, ..., t^k) &= (\iota_2^{-1} \circ \iota_1 \circ \mathrm{Fl}_{t^1}^{\iota_1^* X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{\iota_1^* X_k})(x_1) \\
&= (\iota_2^{-1} \mathrm{Fl}_{t^1}^{X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k} \circ \iota_1)(x_1) \\
&= (\mathrm{Fl}_{t^1}^{\iota_2^* X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{\iota_2^* X_k} \circ \iota_2^{-1} \circ \iota_1)(x_1) \\
&= f_2(t^1, ..., t^k).
\end{aligned}$$

Now we can see that $\iota_2^{-1} \circ \iota_1$ is a diffeomorphism. This allows us to glue together the all the integral manifolds which pass through a fixed x in M to obtain a unique maximal integral submanifold through x. We have prove the following result:

Proposition 11.1 *For a smooth distribution Δ on M and any $x \in M$ there is a unique maximal integral manifold L_x containing x called the **leaf** through x.*

Definition 11.13 *Let $\mathcal{X} \subset \mathfrak{X}_{loc}(M)$. We call X a* ***stable*** *family of local vector fields if for any $X, Y \in \mathcal{X}$ we have*

$$(\mathrm{Fl}_t^X)^* Y \in \mathcal{X}$$

whenever $(\mathrm{Fl}_t^X)^* Y$ is defined. Given an arbitrary subset of local fields $\mathcal{X} \subset \mathfrak{X}_{loc}(M)$ let $\mathcal{S}(\mathcal{X})$ denote the set of all local fields of the form

$$(\mathrm{Fl}_{t^1}^{X_1} \circ \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^t}^{X_k})^* Y$$

where $X_i, Y \in \mathcal{X}$ and where $t = (t^1, ..., t^k)$ varies over all k-tuples such that the above expression is defined.

Exercise 11.4 *Show that $\mathcal{S}(\mathcal{X})$ is the smallest stable family of local vector fields containing $\mathcal{X}$.*

Definition 11.14 *If a diffeomorphism ϕ of a manifold M with a distribution Δ is such that $T_x\phi(\Delta_x) \subset \Delta_{\phi(x)}$ for all $x \in M$ then we call ϕ an* ***automorphism of*** *Δ. If $\phi : U \to \phi(U)$ is such that $T_x\phi(\Delta_x) \subset \Delta_{\phi(x)}$ for all $x \in U$ we call ϕ a* ***local automorphism of*** *Δ.*

Definition 11.15 *If $X \in \mathfrak{X}_{loc}(M)$ is such that $T_x \mathrm{Fl}_t^X(\Delta_x) \subset \Delta_{\mathrm{Fl}_t^X(x)}$ we call X a (local)* ***infinitesimal automorphism*** *of Δ. The set of all such is denoted $\mathrm{aut}_{loc}(\Delta)$.*

Example 11.6 *Convince yourself that $\mathrm{aut}_{loc}(\Delta)$ is stable.*

For the next theorem recall the definition of $\mathfrak{X}_\Delta$.

Theorem 11.3 *Let Δ be a smooth singular distribution on M. Then the following are equivalent:*

1) Δ is integrable.
2) $\mathfrak{X}_\Delta$ is stable.
3) $\mathrm{aut}_{loc}(\Delta) \cap \mathfrak{X}_\Delta$ spans Δ.
4) There exists a family $\mathcal{X} \subset \mathfrak{X}_{loc}(M)$ such that $\mathcal{S}(\mathcal{X})$ spans Δ.

Proof. Assume (1) and let $X \in \mathfrak{X}_\Delta$. If $\mathcal{L}_x$ is the leaf through $x \in M$ then by proposition 7.1

$$\mathrm{Fl}_{-t}^X \circ \iota = \iota \circ \mathrm{Fl}_{-t}^{\iota^* X}$$

where $\iota : \mathcal{L}_x \hookrightarrow M$ is inclusion. Thus

$$\begin{aligned} T_x(\mathrm{Fl}_{-t}^X)(\Delta_x) &= T(\mathrm{Fl}_{-t}^X) \cdot T_x\iota \cdot (T_x\mathcal{L}_x) \\ &= T(\iota \circ \mathrm{Fl}_{-t}^{\iota^* X}) \cdot (T_x\mathcal{L}_x) \\ &= T\iota T_x(\mathrm{Fl}_{-t}^{\iota^* X}) \cdot (T_x\mathcal{L}_x) \\ &= T\iota T_{\mathrm{Fl}_{-t}^{\iota^* X}(x)}\mathcal{L}_x = \Delta_{\mathrm{Fl}_{-t}^{\iota^* X}(x)}. \end{aligned}$$

Now if Y is in $\mathfrak{X}_\Delta$ then at an arbitrary x we have $Y(x) \in \Delta_x$ and so the above shows that $((\mathrm{Fl}_t^X)^* Y)(x) \in \Delta$ so $(\mathrm{Fl}_t^X)^* Y)$ is in $\mathfrak{X}_\Delta$. We conclude that $\mathfrak{X}_\Delta$ is stable and have shown that (1) $\Rightarrow$ (2).

Next, if (2) hold then $\mathfrak{X}_\Delta \subset \mathrm{aut}_{loc}(\Delta)$ and so we have (3).

If (3) holds then we let $\mathcal{X} := \mathrm{aut}_{loc}(\Delta) \cap \mathfrak{X}_\Delta$. Then for $Y, Y \in \mathcal{X}$ we have $(\mathrm{Fl}_t^X)^* Y \in \mathfrak{X}_\Delta$ and so $\mathcal{X} \subset \mathcal{S}(\mathcal{X}) \subset \mathfrak{X}_\Delta$. from this we see that since $\mathcal{X}$ and $\mathfrak{X}_\Delta$ both span Δ so does $\mathcal{S}(\mathcal{X})$.

Finally, we show that (4) implies (1). Let $x \in M$. Since $\mathcal{S}(\mathcal{X})$ spans the distribution and is also stable by construction we have

$$T(\mathrm{Fl}_t^X)\Delta_x = \Delta_{\mathrm{Fl}_t^X(x)}$$

for all fields X from $\mathcal{S}(\mathcal{X})$. Let the dimension Δ_x be k and choose fields $X_1, ..., X_k \in \mathcal{S}(\mathcal{X})$ such that $X_1(x), ..., X_k(x)$ is a basis for Δ_x. Define a map $f :: \mathbb{R}^k \to M$ by

$$f(t^1, ..., t^n) := (\mathrm{Fl}_{t^1}^{X_1} \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k})(x)$$

which is defined (and smooth) near $0 \in \mathbb{R}^k$. As in lemma 11.3 we know that the rank of f at 0 is k and the image of a small enough open neighborhood of 0 is a submanifold. In fact, this image, say $S = f(U)$ is an integral submanifold of Δ through x. To see this just notice that the $T_x S$ is spanned by $\frac{\partial f}{\partial t^j}(0)$ for $j = 1, 2, ..., k$ and

$$\begin{aligned}
\frac{\partial f}{\partial t^j}(0) &= \left.\frac{\partial}{\partial t^j}\right|_0 (\mathrm{Fl}_{t^1}^{X_1} \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k})(x) \\
&= T(\mathrm{Fl}_{t^1}^{X_1} \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^{j-1}}^{X_{j-1}}) X_j ((\mathrm{Fl}_{t^j}^{X_j} \mathrm{Fl}_{t^{j+1}}^{X_{j+1}} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k})(x)) \\
&= ((\mathrm{Fl}_{-t^1}^{X_1})^* (\mathrm{Fl}_{-t^2}^{X_2})^* \circ \cdots \circ (\mathrm{Fl}_{-t^{j-1}}^{X_{j-1}})^* X_j)(f(t^1, ..., t^n)).
\end{aligned}$$

But $\mathcal{S}(\mathcal{X})$ is stable so each $\frac{\partial f}{\partial t^j}(0)$ lies in $\Delta_{f(t)}$. From the construction of f and remembering **??** we see that $\mathrm{span}\{\frac{\partial f}{\partial t^j}(0)\} = T_{f(t)} S = \Delta_{f(t)}$ and we are done. ∎

Chapter 12

Connections on Vector Bundles

12.1 Definitions

A connection can either be defined as a map $\nabla : \mathfrak{X}(M) \times \Gamma(M, E) \to \Gamma(M, E)$ from which one gets a well defined map $\nabla : TM \times \Gamma(M, E) \to \Gamma(M, E)$ or the other way around. The connection should also be natural with respect to restrictions to open sets and so a sheaf theoretic definition could be given.

Definition 12.1 *A connection on a $C^\infty-$vector bundle $E \to M$ is a map ∇ : $\mathfrak{X}(M) \times \Gamma(M, E) \to \Gamma(M, E)$ (where $\nabla(X, s)$ is written as $\nabla_X s$) satisfying the following four properties for all $f \in C^\infty$, $X, X_1, X_2 \in \mathfrak{X}(M)$, $s, s_1, s_2 \in \Gamma(M, E)$ and*

1. $\nabla_{fX}(s) = f\nabla_X s$ for all $f \in C^\infty$, $X \in \mathfrak{X}(M)$ and $s \in \Gamma(M, E)$
2. $\nabla_{X_1+X_2} s = \nabla_{X_1} s + \nabla_{X_2} s$ for all $X_1, X_2 \in \mathfrak{X}(M)$ and $s \in \Gamma(M, E)$
3. $\nabla_X(s_1 + s_2) = \nabla_X s_1 + \nabla_X s_2$ for all $X \in \mathfrak{X}(M)$ and $s_1, s_2 \in \Gamma(M, E)$
4. $\nabla_X(fs) = (Xf)s + f\nabla_X(s)$ for all $f \in C^\infty$, $X \in \mathfrak{X}(M)$ and $s \in \Gamma(M, E)$

As we will see below, for finite dimensional E and M this definition is enough to imply that ∇ induces maps $\nabla^U : \mathfrak{X}(U) \times \Gamma(U, E) \to \Gamma(U, E)$ which are naturally related in the sense we make precise below and furthermore the value $(\nabla_X s)(p)$ depends only on the value X_p and on the values of s along any smooth curve c representing X_p. The proof of these facts depends on the existence of smooth bump functions and so forth. We have already developed the tools to obtain the proof easily in sections 2.8 and 9.4 and so we leave the trivial verification of this to the reader.

In the infinite dimensional case we are not guaranteed such thing and so we may as well include the extra properties into the definition:

Definition 12.2 ((better)) *A **natural covariant derivative** (or **connection**[1]) ∇ on a smooth vector bundle $E \to M$ is an assignment to each open set $U \subset M$ of a map $\nabla^U : \mathfrak{X}(U) \times \Gamma(U, E) \to \mathfrak{X}(U)$ written $\nabla^U : (X, s) \to \nabla^U_X s$ such that the following hold:*

1. *$\nabla^U_X s$ is $C^\infty(U)$-linear in X,*
2. *$\nabla^U_X s$ is $\mathbb{R}$-linear in Y,*
3. *$\nabla^U_X (fs) = f\nabla^U_X s + (Xf)s$ for all $X, Y \in \mathfrak{X}(U)$, $s \in \Gamma(U, E)$ and all $f \in C^\infty(U)$.*
4. *If $V \subset U$ then $r^U_V(\nabla^U_X s) = \nabla^V_{r^U_V X} r^U_V s$ (naturality with respect to restrictions).*
5. *$(\nabla^U_X s)(p)$ only depends of the value of X at p (infinitesimal locality).*

*$\nabla^U_X s$ is called the **covariant derivative** of s with respect to X. We will denote all of the maps ∇^U by the single symbol ∇ when there is no chance of confusion.*

In the same way that extends a derivation to a tensor derivation one may show that a covariant derivative on a vector bundle induces naturally related connections on all the multilinear bundles. In particular, $\pi_* : E^* \to M$ denotes the dual bundle to $E \to M$ we may define connections on $\pi_* : E^* \to M$ and on $\pi \otimes \pi_* : E \otimes E^* \to M$. We do this in such a way that for $e \in \Gamma(M, E)$ and $e^* \in \Gamma(M, E^*)$ we have

$$\nabla^{E \otimes E^*}_X (e \otimes e^*) = \nabla_X e \otimes e^* + e \otimes \nabla^{E^*}_X e^*$$

and

$$(\nabla^{E^*}_X e^*)(e) = X(e^*(e)) - e^*(\nabla_X e).$$

Of course this last formula follows from our insistence that covariant differentiation commutes with contraction:

$$\begin{aligned} X(e^*(e)) &= \\ (\nabla_X C(e \otimes e^*)) &= C(\nabla^{E \otimes E^*}_X (e \otimes e^*)) \\ &= C\left(\nabla_X e \otimes e^* + e \otimes \nabla^{E^*}_X e^*\right) \\ &= e^*(\nabla_X e) + (\nabla^{E^*}_X e^*)(e) \end{aligned}$$

[1]It would be better if we could avoid the term connection at this point and use covariant derivative instead since later we will encounter another definition of a connection which refers to a certain "horizontal" subbundle of the tangent bundle of the total space of a vector bundle (or also of a principal bundle). Connections in this sense determine a natural covariant derivative. Conversely, a natural covariant derivative determines a horizontal subbundle, i.e. a connection in this second (more proper?) sense. Thus a natural covariant derivative and a connection are mutually determining.

where C denotes the contraction $f \otimes f^* \mapsto f^*(f)$. All this works like the tensor derivation extension procedure which we have already done.

Now the bundle $E \otimes E^* \to M$ is naturally isomorphic to $\mathrm{End}(E)$ and by this isomorphism we get a connection on $\mathrm{End}(E)$.

$$(\nabla_X A)(e) = \nabla_X(A(e)) - A(\nabla_X e)$$

Exercise 12.1 *Prove this last formula.*

Solution: Since $c : e \otimes A \mapsto A(e)$ is a contraction we must have

$$\begin{aligned}
\nabla_X(A(e)) &= c\left(\nabla_X e \otimes A + e \otimes \nabla_X A\right) \\
&= A(\nabla_X e) + (\nabla_X A)(e)
\end{aligned}$$

12.2 Local Frame Fields and Connection Forms

Let $\pi : E \to M$ be a rank k vector bundle with a connection ∇. Take M top be of finite dimension n. Recall that a choice of a local frame field over an open set $U \subset M$ is equivalent to a trivialization of the restriction E_U. We now examine expression for the connection from the view point of such a local frame field $e = (e_1, ..., e_k)$. Recall that we have a vector bundle chart (a local trivialization) on an open set U exactly when there exists a frame field. It is not hard to see that there must be a matrix of 1-forms $A = (A_a^b)_{1 \leq a,b \leq k}$ such that for $X \in \Gamma(U)$ we may write

$$\nabla_X e_a = A_a^b(X) e_b.$$

Here and in what follows we use the Einstein summation convention. Also the dependence on the point of evaluation is suppressed since something like $p \mapsto \left. A_a^b \right|_p (X_p) e_b(p)$ is rather awkward looking. The matrix of 1-forms A may be thought of as a matrix valued 1-form A so that for a fixed vector field defined on U we have that $p \mapsto A_p(X_p)$ is a matrix valued function on U. Now let us assume that U is simultaneously the domain of a chart $(x^1, ..., x^n)$ on M. Then we may write $X = X^i \partial_i$ and then

$$\nabla_{\partial_i} e_a = A_a^b(\partial_i) e_b = A_{ia}^b e_b$$

and so

$$\begin{aligned}
\nabla_X s &= \nabla_X(s^a e_a) \\
&= \nabla_X(s^a e_a) \\
&= (X s^a) e_a + s^a \nabla_X e_a \\
&= (X s^a) e_a + s^a A_a^b(X) e_b \\
&= (X s^a) e_a + s^r A_r^a(X) e_a \\
&= (X s^a + A_r^a(X) s^r) e_a
\end{aligned}$$

So the $a-$component of $\nabla_X s$ is $(\nabla_X s)^a = \nabla_X s^a := X s^a + A^a_r(X)s^r$. Of course the frame are defined only locally say on some open set U. The restriction E_U is trivial. Let us examine the forms A^b_a on this open set. The change of frame

$$f^b = g^b_a e_b$$

which in matrix notation is

$$f = eg.$$

Differentiating both sides

$$\begin{aligned} f &= eg \\ \nabla f &= \nabla(eg) \\ fA' &= (\nabla e)g + edg \\ fA' &= egg^{-1}Ag + egg^{-1}dg \\ fA' &= fg^{-1}Ag + fg^{-1}dg \\ A' &= g^{-1}Ag + g^{-1}dg \end{aligned}$$

Conversely, we have the following theorem:

Theorem 12.1 *Let $\pi : E \to M$ be a smooth $\mathbb{F}-$vector bundle of rank k. Suppose we are given a cover $\{U_\alpha\}$ of the base space M by the domains of frame fields $e^\alpha = (e^\alpha_1, ..., e^\alpha_k)$ and an association of a matrix valued $1-$form to $\overset{\alpha}{A} : U_\alpha \to gl(k, \mathbb{F}) \otimes T^*U_\alpha$ to each. Then there is a unique connection on $\pi : E \to M$ which is given in each U_α by*

$$\nabla_X s = (X s^a + A^a_r(X)s^r)e_a$$

for $s = \sum s^a e_a$.

Sometimes one hears that A is locally an element of $\mathrm{Hom}(E, E)$ but the transformation law just discovered says otherwise. The meaning of the statement can only be the following: If E were trivial then we could choose a distinguished global frame field $e = (e_1, ..., e_n)$ and define a connection by the simple rule $\nabla^0_X(s^a e_a) = (X s^a)e_a$. Now an the above calculation of the transformation law shows that the difference of two connections on a vector bundle is in fact an element of $\mathrm{Hom}(E, E)$. Equivalently, if $\triangle A = A - \widehat{A}$ is the difference between the connection forms for two different connections then under a change of frame we have

$$\begin{aligned} (\triangle A)' &= A' - \widehat{A}' \\ &= g^{-1}Ag + g^{-1}dg - (g^{-1}\widehat{A}g + g^{-1}dg) \\ &= g^{-1}Ag - g^{-1}\widehat{A}g \\ &= g^{-1}(\triangle A)g \end{aligned}$$

so that $\triangle A$ defines a section of the bundle $\mathrm{End}(E)$.

Exercise 12.2 *Show that the set of all connections on E is naturally an affine space $C(E)$ whose vector space of differences is* $\mathrm{End}(E)$. *For any fixed connection ∇^0 we have an affine isomorphism* $\mathrm{End}(E) \to C(E)$ *given by* $\triangle A \mapsto \nabla^0 + \triangle A$.

Now in the special case mentioned above for a trivial bundle the connection form in the defining frame is zero and so in that case $\triangle A = A$. So in this case A determines a section of $\mathrm{Hom}(E, E)$. Now any bundle is locally trivial so in this sense A is locally in $\mathrm{End}(E)$. But this is just confusing and in fact cheating since we have changed (by force so to speak) the transformation law for A among frames defined on the same open set to that of $\triangle A$ rather than A. The point is that even though $\triangle A$ and A are equal in the distinguished frame they are not the same after transformation to a new frame. It seems to the author best to treat A for what it is; a matrix valued $1-$form which depends on the frame chosen.

12.3 Parallel Transport

Suppose we are given a curve $c : I \to M$ together with an $E-$valued section along c; that is a map $\sigma : I \to E$ such that the following diagram commutes:

$$\begin{array}{ccc} & & E \\ & \overset{\sigma}{\nearrow} & \downarrow \\ I & \overset{c}{\to} & M \end{array}$$

We wish to define $\nabla_t \sigma = \nabla_{\partial_t} \sigma$. Lets get some motivation. If c is an integral curve of a field X then we have

$$\begin{aligned}(\nabla_X s)(c(t)) &= (X(c(t)).s^a(c(t)) + (\left. A_r^a \right|_{c(t)} X_{c(t)})s^r(c(t)))e_a(c(t)) \\ &= (\dot{c}(t).s^a(t))e_a(t) + (\left. A_r^a \right|_{c(t)} \dot{c}(t))s^r(t)e_a(t)\end{aligned}$$

where we have abbreviated $s^a(t) := s^a(c(t))$ and $e_a(t) := e_a(c(t))$. This shows that the value of $\nabla_X s$ at $c(t)$ depends only on $\dot{c}(t)$ and $(s \circ c)(t)$. This observation motivates the following definition:

Definition 12.3 *Let $c : I \to M$ be a smooth curve and σ an E values section along c. We define another section along c denoted $\nabla_{\partial_t}\sigma$ by the requirement that with respect to any frame field (e_a) we have*

$$\nabla_{\partial_t}\sigma := (\frac{d}{dt}\sigma^a(t))e_a(t) + (\left. A_r^a \right|_{c(t)} \dot{c}(t))\sigma^r(t)e_a(t)$$

Since c might not be even be an immersion the definition only makes sense because of the fact that it is independent of the frame. To do the calculation which shows this frame independence it will pay to make the following abbreviations

Figure 12.1: Parallel Transport.

1. $\sigma' = (\sigma'_1, ..., \sigma'_k)$ (the components of σ with respect to a new basis $f = eg$)
2. $\frac{d}{dt}\sigma^a(t) = d\sigma'$
3. $A = (\, A^a_r|_{c(t)}\, \dot{c}(t))$

Then using matrix notion we have

$$
\begin{aligned}
& f d\sigma' + f A' \sigma' \\
& = egd(g^{-1}\sigma) + eg(g^{-1}Ag + g^{-1}dg)g^{-1}\sigma \\
& = eg(g^{-1}d\sigma - g^{-1}dgg^{-1}\sigma) + eA\sigma + egdgg^{-1}\sigma \\
& = ed\sigma + eAg.
\end{aligned}
$$

Exercise 12.3 *Flesh out this calculation without the abbreviation.*

Now on to the parallel transport.

Definition 12.4 *Let $c : [a, b] \to M$ be a smooth curve. A section σ along c is said to be* ***parallel along*** *c if*

$$(\nabla_{\partial_t}\sigma)(t) = 0 \text{ for all } t \in [a, b].$$

Similarly, a section $\sigma \in \Gamma(M, E)$ is said to be ***parallel*** *if $\nabla_X \sigma = 0$ for all $X \in \mathfrak{X}(M)$.*

Exercise 12.4 *Show that $\sigma \in \Gamma(M, E)$ is a parallel section iff $X \circ c$ is parallel along c for every curve $c : I \to M$*

Exercise 12.5 *Show that for $f : I \to \mathbb{R}$ and $\sigma : I \to M$ is a section of E along c then $\nabla_{\partial_t}(f\sigma) = \frac{df}{dt}\sigma + f\nabla_{\partial_t}\sigma$.*

Exercise 12.6 *Continuing the last exercise show that if $\sigma : I \to U \subset M$ where U is the domain of a local frame field $\{e_1, ..., e_k\}$ then $\sigma(t) = \sum_{i=1}^{k} \sigma^i(t)e_i(c(t))$.*

Theorem 12.2 *Given a smooth curve $c : [a, b] \to M$ and numbers $t_0 \in [a, b]$ with $c(t_0) = p$ and vector $v \in E_p$ there is a unique parallel section σ_c along c such that $\sigma_c(t_0) = v$.*

Proof. In local coordinates this reduces to a first order initial value problem which may be shown to have a unique smooth solution. Thus if the image of the curve lies completely inside a coordinate chart then we have the result. The general result follows from patching these together. This is exactly what we do below when we generalize to piecewise smooth curves so we will leave this last part of the proof to the skeptical reader. ■

Under the conditions of this last theorem the value $\sigma_c(t)$ is vector in the fiber $E_{c(t)}$ and is called the **parallel transport** of v along c from $c(t_0)$ to $c(t)$. Let us denote this by $P(c)_{t_0}^{t}v$. Next we suppose that $c : [a, b] \to M$ is a (continuous) piecewise smooth curve. Thus we may find a monotonic sequence $t_0, t_1, ...t_j = t$ such that $c_i := c|_{[t_{i-1},t_i]}$ (or $c|_{[t_i,t_{i-1}]}$) is smooth. [2] In this case we define

$$P(c)_{t_0}^{t} := P(c)_{t_{j-1}}^{t} \cdots \circ P(c)_{t_0}^{t_1}$$

Now given $v \in E_{c(t_0)}$ as before, the obvious sequence of initial value problems gives a unique piecewise smooth section σ_c along c such that $\sigma_c(t_0) = v$ and the solution must clearly be $P(c)_{t_0}^{t}v$ (why?).

Exercise 12.7 *$P(c)_{t_0}^{t} : E_{c(t_0)} \to E_{c(t)}$ is a linear isomorphism for all t with inverse $P(c)_{t}^{t_0}$ and the map $t \mapsto \sigma_c(t) = P(c)_{t_0}^{t}v$ is a section along c which is smooth wherever c is smooth.*

We may approach the covariant derivative from the direction of parallel transport. Indeed some authors given an axiomatic definition of parallel transport, prove its existence and then use it to define covariant derivative. For us it will suffice to have the following theorem:

Theorem 12.3 *For any smooth section σ of E defined along a smooth curve $c : I \to M$. Then we have*

$$(\nabla_{\partial_t}\sigma)(t) = \lim_{\epsilon \to 0} \frac{P(c)_{t+\epsilon}^{t}\sigma(t+\epsilon) - \sigma(t)}{\epsilon}$$

[2] It may be that $t < t_0$.

Proof. Let $e_1, ..., e_k$ be a basis of $E_{c(t_0)}$ for some fixed $t_0 \in I$. Let $e_i(t) := P(c)_{t_0}^t$. Then $\nabla_{\partial_t} e_i(t) \equiv 0$ and $\sigma(t) = \sum \sigma^i(t) e_i(t)$. Then

$$\begin{aligned}
&\lim_{\epsilon \to 0} \frac{P(c)_{t+\epsilon}^t \sigma(t+\epsilon) - \sigma(t)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{\sigma^i(t+\epsilon) P(c)_{t+\epsilon}^t e_i(t+\epsilon) - \sigma(t)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{\sigma^i(t+\epsilon) e_i(t) - \sigma^i(t) e_i(t)}{\epsilon} \\
&= \sum \frac{d\sigma^i}{dt}(t) e_i(t).
\end{aligned}$$

On the other hand

$$\begin{aligned}
(\nabla_{\partial_t} \sigma)(t) &= \nabla_{\partial_t}(\sigma^i(t) e_i(t)) \\
&= \sum \frac{d\sigma^i}{dt}(t) e_i(t) + \sum \sigma^i(t) \nabla_{\partial_t} e_i(t) \\
&= \sum \frac{d\sigma^i}{dt}(t) e_i(t).
\end{aligned}$$

■

$1-$form $\theta = \sum e_j \theta^i$ which takes any vector to itself:

$$\begin{aligned}
\theta(v_p) &= \sum e_j(p) \theta^i(v_p) \\
&= \sum v^i e_j(p) = v_p
\end{aligned}$$

Let us write $d^\nabla \theta = \frac{1}{2} \sum e_k \otimes T_{ij}^k \theta^i \wedge \theta^j = \frac{1}{2} \sum e_k \otimes \tau^k$. If ∇ is the Levi Civita connection on M then consider the projection $P^\wedge : E \otimes TM \otimes T^*M$ given by $P^\wedge T(\xi, v) = T(\xi, v) - T(v, \xi)$. We have

$$\begin{aligned}
\nabla e_j &= \omega_j^k e_k = e\omega \\
\nabla \theta^j &= -\omega_k^j \theta^k
\end{aligned}$$

$\nabla_\xi (e_j \otimes \theta^j)$
$P^\wedge(\nabla_\xi \theta^j)(v) = -\omega_k^j(\xi)\theta^k(v) + \omega_k^j(v)\theta^k(\xi) = -\omega_k^j \wedge \theta^k$

Let $T(\xi, v) = \nabla_\xi (e_i \otimes \theta^j)(v)$
$= (\nabla_\xi e_i) \otimes \theta^j(v) + e_i \otimes (\nabla_\xi \theta^j)(v) = \omega_i^k(\xi) e_k \otimes \theta^j(v) + e_i \otimes (-\omega_k^j(\xi)\theta^k(v))$
$= \omega_i^k(\xi) e_k \otimes \theta^j(v) + e_i \otimes (-\omega_j^k(\xi)\theta^j(v)) = e_k \otimes (\omega_i^k(\xi) - \omega_j^k(\xi))\theta^j(v)$
Then

$$
\begin{aligned}
(P^\wedge T)(\xi, v) &= T(\xi, v) - T(v, \xi) \\
&= (\nabla e_j) \wedge \theta^j + e_j \otimes d\theta^j \\
&= d^\nabla (e_j \otimes \theta^j)
\end{aligned}
$$

$$
\begin{aligned}
d^\nabla \theta &= d^\nabla \sum e_j \theta^j \\
&= \sum (\nabla e_j) \wedge \theta^j + \sum e_j \otimes d\theta^j \qquad (12.1)\\
&= \sum (\sum_k e_k \otimes \omega_j^k) \wedge \theta^j + \sum e_k \otimes d\theta^k \\
&= \sum_k e_k \otimes (\sum_j \omega_j^k \wedge \theta^j + d\theta^k)
\end{aligned}
$$

So that $\sum_j \omega_j^k \wedge \theta^j + d\theta^k = \frac{1}{2}\tau^k$. Now let $\sigma = \sum f^j e_j$ be a vector field

$$
\begin{aligned}
d^\nabla d^\nabla \sigma = d^\nabla \left(d^\nabla \sum e_j f^j \right) = d^\nabla \left(\sum (\nabla e_j) f^j + \sum e_j \otimes df^j \right) \\
\left(\sum (\nabla e_j) df^j + \sum (d^\nabla \nabla e_j) f^j + \sum \nabla e_j df^j + \sum e_j \otimes ddf^j \right) \\
\sum f^j (d^\nabla \nabla e_j) = \sum f^j
\end{aligned}
$$

So we seem to have a map $f^j e_j \mapsto \Omega_j^k f^j e_k$.

$$
\begin{aligned}
e_r \Omega_j^r &= d^\nabla \nabla e_j = d^\nabla (e_k \omega_j^k) \\
&= \nabla e_k \wedge \omega_j^k + e_k d\omega_j^k \\
&= e_r \omega_k^r \wedge \omega_j^k + e_k d\omega_j^k \\
&= e_r \omega_k^r \wedge \omega_j^k + e_r d\omega_j^r \\
&= e_r (d\omega_j^r + \omega_k^r \wedge \omega_j^k)
\end{aligned}
$$

$$
d^\nabla \nabla e = d^\nabla (e\omega) = \nabla e \wedge \omega + e d\omega
$$

From this we get $0 = d(A^{-1}A)A^{-1} = (dA^{-1})AA^{-1} + A^{-1}dAA^{-1}$

$$dA^{-1} = A^{-1}dAA^{-1}$$

$$\begin{aligned}
\Omega^r_j &= d\omega^r_j + \omega^r_k \wedge \omega^k_j \\
\Omega &= d\omega + \omega \wedge \omega \\
\Omega' &= d\omega' + \omega' \wedge \omega' \\
\Omega' &= d\left(A^{-1}\omega A + A^{-1}dA\right) + \left(A^{-1}\omega A + A^{-1}dA\right) \wedge \left(A^{-1}\omega A + A^{-1}dA\right) \\
&= d\left(A^{-1}\omega A\right) + d\left(A^{-1}dA\right) + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1} \wedge \omega A + A^{-1}dA \wedge A^{-1}dA \\
&= d\left(A^{-1}\omega A\right) + dA^{-1} \wedge dA + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1}\omega A + A^{-1}dA \wedge A^{-1}dA \\
&= dA^{-1}\omega A + A^{-1}d\omega A - A^{-1}\omega dA + dA^{-1} \wedge dA + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1} \wedge \omega A + A^{-1}dA \wedge A^{-1}dA \\
&= A^{-1}d\omega A + A^{-1}\omega \wedge \omega A \\
\Omega' &= A^{-1}\Omega A
\end{aligned}$$

$$\omega' = A^{-1}\omega A + A^{-1}dA$$

These are interesting equations let us approach things from a more familiar setting so as to interpret what we have.

12.4 Curvature

An important fact about covariant derivatives is that they don't need to commute. If $\sigma : M \to E$ is a section and $X \in \mathfrak{X}(M)$ then $\nabla_X\sigma$ is a section also and so we may take it's covariant derivative $\nabla_Y\nabla_X\sigma$ with respect to some $Y \in \mathfrak{X}(M)$. In general, $\nabla_Y\nabla_X\sigma \neq \nabla_X\nabla_Y\sigma$ and this fact has an underlying geometric interpretation which we will explore later. A measure of this lack of commutativity is the **curvature operator** which is defined for a pair $X, Y \in \mathfrak{X}(M)$ to be the map $F(X,Y) : \Gamma(E) \to \Gamma(E)$ defined by

$$F(X,Y)\sigma := \nabla_X\nabla_Y\sigma - \nabla_Y\nabla_X\sigma - \nabla_{[X,Y]}\sigma.$$

or

$$[\nabla_X, \nabla_Y]\sigma - \nabla_{[X,Y]}\sigma$$

Theorem 12.4 *For fixed σ the map $(X,Y) \mapsto F(X,Y)\sigma$ is $C^\infty(M)$ bilinear and antisymmetric.*

$F(X,Y) : \Gamma(E) \to \Gamma(E)$ is a $C^\infty(M)$ module homomorphism; that is it is linear over the smooth functions:

$$F(X,Y)(f\sigma) = fF(X,Y)(\sigma)$$

Proof. We leave the proof of the first part as an exercise. For the second part we just calculate:

$$\begin{aligned}
F(X,Y)(f\sigma) &= \nabla_X\nabla_Y f\sigma - \nabla_Y\nabla_X f\sigma - \nabla_{[X,Y]}f\sigma \\
&= \nabla_X(f\nabla_Y\sigma + (Yf)\sigma) - \nabla_Y(f\nabla_X\sigma + (Xf)\sigma) \\
&- f\nabla_{[X,Y]}\sigma - ([X,Y]f)\sigma \\
&= f\nabla_X\nabla_Y\sigma + (Xf)\nabla_Y\sigma + (Yf)\nabla_X\sigma + X(Yf) \\
&- f\nabla_Y\nabla_X\sigma - (Yf)\nabla_X\sigma - (Xf)\nabla_Y\sigma - Y(Xf) \\
&- f\nabla_{[X,Y]}\sigma - ([X,Y]f)\sigma \\
&= f[\nabla_X,\nabla_Y] - f\nabla_{[X,Y]}\sigma = fF(X,Y)\sigma
\end{aligned}$$

■

Exercise 12.8 *Prove the first part of theorem 12.4. Now recall that*

$$\begin{aligned}
&\mathrm{End}_{C^\infty}(\Gamma(E)) \\
&\cong \Gamma(M,\mathrm{End}(E)) \\
&\cong \Gamma(M, E\otimes E^*)
\end{aligned}$$

Thus we also have F as a map $F : \mathfrak{X}(M)\times\mathfrak{X}(M) \to \Gamma(M,\mathrm{End}(E))$. But then since F is tensorial in the first two slot and antisymmetric we also may think in the following terms

$$\begin{aligned}
&F \in \Gamma(M, \mathrm{Hom}(E,E)\otimes\wedge^2 M) \\
&\quad \text{or} \\
&F \in \Gamma(M, E\otimes E^*\otimes\wedge^2 M).
\end{aligned}$$

In the current circumstance it is harmless to identify $E\otimes E^*\otimes\wedge^2 M$ with $\wedge^2 M\otimes E\otimes E^*$ the second one seems natural too although when translating into matrix notation the first is more consistent. In any case we have a natural structure of an algebra on each fiber given by

$$(A\otimes\alpha)\wedge(B\otimes\beta) := (A\circ B)\otimes\alpha\wedge\beta$$

and this gives a $C^\infty(M)-$algebra structure on $\Gamma(M,\mathrm{Hom}(E,E)\otimes\wedge^2 M)$.

We will describe the relationship between the curvature F and parallel transport but first lets see another approach to curvature. For a vector bundle $E\to M$ we may construct the

Let $e = (e_1,...,e_k)$ be a frame defined on an open set U and for the restriction of a section to U we write

$$\sigma = \sum_{i=1}^k \sigma_U^i e_i$$

for smooth functions $\sigma_U^i : U \to \mathbb{F}$ (which is $\mathbb{R}$ or $\mathbb{C}$). Then locally,

$$\begin{aligned} F(X,Y)\sigma &= F(X,Y) \cdot \sum_{i=1}^{k} \sigma_U^i e_i \\ &= \sum_{i=1}^{k} \sigma_U^i F(X,Y) e_i \\ &\sum_{i=1}^{k} \sigma_U^i F_i(X,Y) \end{aligned}$$

Chapter 13

Riemannian and semi-Riemannian Manifolds

The most beautiful thing we can experience is the mysterious. It is the source of all true art and science.
-Albert Einstein

13.1 The Linear Theory

13.1.1 Scalar Products

Definition 13.1 *A scalar product on a (real) finite dimensional vector space* V *is a nondegenerate symmetric bilinear form* $\mathrm{g} : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$ *. The scalar product is called*

1. *positive (resp. negative) definite if* $\mathrm{g}(v, v) \geq 0$ *(resp.* $\mathrm{g}(v, v) \leq 0$*) for all* $v \in \mathrm{V}$ *and* $\mathrm{g}(v, v) = 0 \Longrightarrow v = 0$*.*
2. *positive (resp. negative) semidefinite if* $\mathrm{g}(v, v) \geq 0$ *(resp.*$\mathrm{g}(v, v) \leq 0$*) for all* $v \in \mathrm{V}$*.*

Nondegenerate means that the map $\mathrm{g} : \mathrm{V} \to \mathrm{V}^*$ given by $v \mapsto \mathrm{g}(v, .)$ is a linear isomorphism or equivalently, if $\mathrm{g}(v, w) = 0$ for all $w \in V$ implies that $v = 0$.

Definition 13.2 *A* ***scalar product space*** *is a pair* V, g *where* V *is a vector space and* g *is a scalar product.*

Remark 13.1 *We shall reserve the terms* ***inner product*** *and* ***inner product space*** *to the case where* g *is positive definite.*

Definition 13.3 *The* ***index*** *of a symmetric bilinear* g *form on* V *is the largest subspace* $W \subset V$ *such that the restriction* $\mathrm{g}|_W$ *is negative definite.*

Given a basis $\mathcal{B} = (v_1, ..., v_n)$ for V we may form the matrix $[\mathrm{g}]^{\mathcal{B}\mathcal{B}}$ which has as ij-th entry $\mathrm{g}(v_i, v_j)$. This is the matrix that represents g with respect to the basis $\mathcal{B}$. So if $v = \mathcal{B}\,[v]^{\mathcal{B}}$, $w = \mathcal{B}\,[w]^{\mathcal{B}}$ then

$$\mathrm{g}(v, w) = [v]^{\mathcal{B}}\,[\mathrm{g}]_{\mathcal{B}\mathcal{B}}\,[w]^{\mathcal{B}}.$$

It is easy to see that the index ind(g) is zero iff g positive semidefinite. It is a standard fact from linear algebra that if g is a scalar product then there exists a basis $e_1, ..., e_n$ for V such that the matrix representative of g with respect to this basis is a diagonal matrix diag$(-1, ..., 1)$ with ones or minus ones along the diagonal and we may arrange for the minus ones come first. Such a basis is called an orthonormal basis for V, g. The number of minus ones appearing is the index ind(g) and so is independent of the orthonormal basis chosen. Thus if $e_1, ..., e_n$ is an orthonormal basis for V, g then $\mathrm{g}(e_i, e_j) = \epsilon_i \delta_{ij}$ where $\epsilon_i = \mathrm{g}(e_i, e_i) = \pm 1$are the entries of the diagonal matrix the first ind(g) of which are equal to -1 and the remaining are equal to 1. Let us refer to the list of ± 1 given by $(\epsilon_1, ..., \epsilon_n)$ as the **signature**.

Remark 13.2 *The convention of putting the minus signs first is not universal and in fact we reserve the right to change the convention to a positive first convention but ample warning will be given. The negative signs first convention is popular in relativity theory but the reverse is usual in quantum field theory. It makes no physical difference in the final analysis as long as one is consistent but it can be confusing when comparing references from the literature.*

Another difference between the theory of positive definite scalar products and indefinite scalar products is the appearance of the ϵ_i from the signature in formulas which would be familiar in positive definite case. For example we have the following:

Proposition 13.1 *Let $e_1, ..., e_n$ be an orthonormal basis for V, g. For any $v \in V, \mathrm{g}$ we have a unique expansion given by $v = \sum_i \epsilon_i \langle v, e_i \rangle e_i$.*

Proof. The usual proof works. One just has to notice the appearance of the ϵ_i. ∎

Definition 13.4 *Let V, g be a scalar product space. We say that v and w are mutually orthogonal iff $\mathrm{g}(v, w) = 0$. Furthermore, given two subspaces W_1 and W_2 of V we say that W_1 is orthogonal to W_2 and write $\mathrm{W}_1 \perp \mathrm{W}_2$ iff every element of W_1 is orthogonal to every element of W_2.*

Since in general g is not necessarily positive definite or negative finite it may be that there are elements that are orthogonal to themselves.

Definition 13.5 *Given a subspace W of a scaler product space V we may consider the orthogonal subspace $\mathrm{W}^{\perp} = \{v \in V : \mathrm{g}(v, w) = 0$ for all $w \in \mathrm{W}\}$.*

We always have $\dim(\mathrm{W}) + \dim(\mathrm{W}^{\perp}) = \dim(\mathrm{V})$ but unless g is definite we may have $\mathrm{W} \cap \mathrm{W}^{\perp} \neq \emptyset$. Of course by nondegeneracy we will always have $V^{\perp} = 0$.

Definition 13.6 *A subspace* W *of a scaler product space* V, g *is called nondegenerate if* $\mathrm{g}|_{\mathrm{W}}$ *is nondegenerate.*

Lemma 13.1 *A subspace* $\mathrm{W} \subset V, \mathrm{g}$ *is nondegenerate iff* $\mathrm{V} = \mathrm{W} \oplus \mathrm{W}^{\perp}$ *(inner direct sum).*

Proof. Easy exercise in linear algebra. ∎

Just as for inner product spaces we define a linear isomorphism $R : \mathrm{V}_1, \mathrm{g}_1 \to \mathrm{V}_2, \mathrm{g}_2$ from one scalar product space to another to be an **isometry** if $\mathrm{g}_1(v, w) = \mathrm{g}_2(Rv, Rw)$. It is not hard to show that if such an isometry exists then g_1 and g_2 have the same index and signature.

13.1.2 Natural Extensions and the Star Operator

If we have a scalar product g on a finite dimensional vector space V then there is a natural way to induce a scalar product on the various tensor spaces $T^r_s(\mathrm{V})$ and on the Grassmann algebra. The best way to explain is by way of some examples.

First consider V^*. Since g is nondegenerate there is a linear isomorphism map $\mathrm{g}_{\flat} : \mathrm{V} \to \mathrm{V}^*$ defined by

$$\mathrm{g}_{\flat}(v)(w) = \mathrm{g}(v, w).$$

Denote the inverse by $\mathrm{g}^{\sharp} : \mathrm{V}^* \to \mathrm{V}$. We force this to be an isometry by defining the scalar product on V^* to be

$$\mathrm{g}^*(\alpha, \beta) = \mathrm{g}(\mathrm{g}^{\sharp}(\alpha), \mathrm{g}^{\sharp}(\beta)).$$

Under this prescription, the dual basis $e^1, ..., e^n$ to an orthonormal basis $e_1, ..., e_n$ for V will be orthonormal. The signature (and hence the index) of g^* and g are the same.

Next consider $T^1_1(\mathrm{V}) = \mathrm{V} \otimes \mathrm{V}^*$. We define the scalar product of two simple tensors $v_1 \otimes \alpha_1, v_2 \otimes \alpha_2 \in \mathrm{V} \otimes \mathrm{V}^*$ by

$$\mathrm{g}^1_1(v_1 \otimes \alpha_1, v_2 \otimes \alpha_2) = \mathrm{g}(v_1, v_2) B^*(\alpha_1, \alpha_2).$$

One can then see that for orthonormal dual bases $e^1, ..., e^n$ and $e_1, ..., e_n$ we have that

$$\{e_i \otimes e^j\}_{1 \leq i,j \leq n}$$

is an orthonormal basis for $T^1_1(\mathrm{V}), \mathrm{g}^1_1$. In general one defines g^r_s so that the natural basis for $T^r_s(\mathrm{V})$ formed from orthonormal $e^1, ..., e^n$ and $e_1, ..., e_n$ will be orthonormal.

Notation 13.1 *In order to reduce notational clutter let us agree to denote all the scalar products coming from* g *simply by* $\langle .,. \rangle$.

Exercise 13.1 *Show that under the natural identification of* $\mathrm{V}\otimes\mathrm{V}^*$ *with* $L(\mathrm{V},\mathrm{V})$ *the scalar product of a linear transformation* A *with it self is the trace of* A.

Next we see how to extend the maps $\mathrm{g}_\flat$ and $\mathrm{g}^\sharp$ to maps on tensors. We give two ways of defining the extensions. In either case, what we want to define is maps $(\mathrm{g}_\flat)^i_\downarrow : T^r{}_s(\mathrm{V}) \to T^{r-1}{}_{s+1}(\mathrm{V})$ and $(\mathrm{g}^\sharp)^\uparrow_j : T^r{}_s(\mathrm{V}) \to T^{r+1}{}_{s-1}(\mathrm{V})$ where $0 \le i \le r$ and $0 \le j \le s$. Our definitions will be given on simple tensors by

$$
\begin{aligned}
&(\mathrm{g}_\flat)^i_\downarrow(w_1 \otimes \cdots \otimes w_r \otimes \omega^1 \otimes \cdots \otimes \omega^s) \\
&= w_1 \otimes \cdots \otimes \widehat{w_i} \otimes \cdots w_r \otimes \mathrm{g}_\flat(w_i) \otimes \omega^1 \otimes \cdots \otimes \omega^s
\end{aligned}
$$

and

$$
\begin{aligned}
&(\mathrm{g}^\sharp)^\uparrow_j(w_1 \otimes \cdots \otimes w_r \otimes \omega^1 \otimes \cdots \otimes \omega^s) \\
&= w_1 \otimes \cdots \otimes w_r \otimes \mathrm{g}^\sharp(\omega^j) \otimes \omega^1 \otimes \cdots \otimes \widehat{\omega^j} \otimes \cdots \otimes \omega^s.
\end{aligned}
$$

This definition is extended to all of $T^r_s(\mathrm{V})$ by linearity. For our second, equivalent definition let $\Upsilon \in T^r{}_s(\mathrm{V})$. Then

$$
\begin{aligned}
&((\mathrm{g}_\flat)^i_\downarrow \Upsilon)(\alpha^1, ..., \alpha^{r-1}; v_1, ..., v_{s+1}) \\
&:= \Upsilon(\alpha^1, ..., \alpha^{r-1}, \mathrm{g}_\flat(v_1); v_2, ..., v_{s+1})
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&((\mathrm{g}^\sharp)^\uparrow_j \Upsilon)(\alpha^1, ..., \alpha^{r+1}; v_1, ..., v_{s-1}) \\
&:= \Upsilon(\alpha^1, ..., \alpha^r, \mathrm{g}^\sharp(\alpha^{r+1}); v_2, ..., v_{s+1})
\end{aligned}
$$

Lets us see what this looks like by viewing the components. Let $f_1, ..., f_n$ be an arbitrary basis of V and let $f^1, ..., f^n$ be the dual basis for V^*. Let $\mathrm{g}_{ij} := \mathrm{g}(f_i, f_j)$ and $\mathrm{g}^{ij} = \mathrm{g}^*(f^i, f^j)$. The reader should check that $\sum_k \mathrm{g}_{kj}\mathrm{g}^{ik} = \delta^i_j$. Now let $\tau \in T^r_s(\mathrm{V})$ and write

$$
\tau = \tau^{i_1,..,i_r}{}_{j_1,...,j_s} f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f^{j_1} \otimes \cdots \otimes f^{j_s}.
$$

Define

$$
\tau^{i_1,...,i_r j}_{\ \ j_1,...,j_{b-1},\widehat{k},j_b...,j_{s-1}} := \tau^{i_1,...,i_r}{}_{j_1,...,j_{j-1}}{}^{j}{}_{j_{b+1}...,j_{s-1}} := \sum_m b^{km} \tau^{i_1,..,i_r}_{j_1,...,j_{b-1},m,j_b...,j_{s-1}}
$$

Then

$$
(\mathrm{g}^\sharp)^\uparrow_a \tau = \tau^{i_1,...,i_r j_a}_{\ \ j_1,...,j_{a-1},\widehat{j_a},j_{a+1}...,j_{s-1}} f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f_{j_a} \otimes f^{j_1} \otimes \cdots \otimes f^{j_{s-1}}.
$$

Thus the $(\mathrm{g}^\sharp)^\uparrow_a$ visually seems to raise an index out of a-th place and puts it up in the last place above. Similarly, the component version of lowering $(\mathrm{g}_\flat)^a_\downarrow$ takes

$$\tau^{\,i_1,..,i_r}{}_{j_1,\dots,j_s}$$

and produces

$$\tau^{\,i_1,..,}{}_{i_a}{}^{,\dots i_r}{}_{j_1,\dots,j_s} = \tau^{i_1,..,\widehat{i_a},\dots i_r}_{i_a j_1,\dots,j_s}.$$

How and why would one do this so called index raising and lowering? What motivates the choice of the slots? In practice one applies this type changing only in specific well motivated situations and the choice of slot placement is at least partially conventional. We will comment further when we actually apply these operators. The notation is suggestive and the $\mathrm{g}^\sharp$ and $\mathrm{g}_\flat$ and their extensions are referred to as **musical isomorphisms** . One thing that is useful to know is that if we raise all the lower indices and lower all the upper ones on a tensor then we can "completely contract" against another one of the original type with the result being the scalar product. For example, let $\tau = \sum \tau_{ij} f^i \otimes f^j$ and $\chi = \sum \chi_{ij} f^i \otimes f^j$. Then letting the components of $(\mathrm{g}^\sharp)^\uparrow_1 \circ (\mathrm{g}^\sharp)^\uparrow_1(\chi)$ by χ^{ij} we have

$$\chi^{ij} = \mathrm{g}^{ik}\mathrm{g}^{jl}\chi_{kl}$$

and

$$\langle \chi, \tau \rangle = \sum \chi_{ij}\tau^{ij}.$$

In general, unless otherwise indicated, we will preform repeated index raising by raising from the first slot $(\mathrm{g}^\sharp)^\uparrow_1 \circ \cdots \circ (\mathrm{g}^\sharp)^\uparrow_1$ and similarly for repeated lowering $(\mathrm{g}_\flat)^1_\downarrow \circ \cdots \circ (\mathrm{g}_\flat)^1_\downarrow$. For example,

$$A_{ijkl} \mapsto A^i{}_{jkl} = \mathrm{g}^{ia}A_{ajkl} \mapsto A^{ij}{}_{kl} = \mathrm{g}^{ia}\mathrm{g}^{jb}A_{abkl}$$

Exercise 13.2 *Verify the above claim directly from the definition of $\langle \chi, \tau \rangle$.*

Even though elements of $L^k_{alt}(\mathrm{V}) \cong \bigwedge^k(\mathrm{V}^*)$ can be thought of as tensors of type $0, k$ that just happen to be anti-symmetric, it is better in most cases to give a scalar product to this space in such a way that the basis

$$\{e^{i_1} \wedge \cdots \wedge e^{i_k}\}_{i_1 < \dots < i_k} = \{e^{\vec{I}}\}$$

is orthonormal if $e^1, ..., e^n$ is orthonormal. Now given any k-form $\omega = a_{\vec{I}} e^{\vec{I}}$ where $e^{\vec{I}} = e^{i_1} \wedge \cdots \wedge e^{i_k}$ with $i_1 < \dots < i_k$ as explained earlier, we can also write $\omega = \frac{1}{k!} a_I e^I$ and then as a tensor

$$\begin{aligned}\omega &= \frac{1}{k!} a_I e^I \\ &= \frac{1}{k!} a_{i_1 \dots i_k} e^{i_1} \otimes \cdots \otimes e^{i_k}.\end{aligned}$$

Thus as a covariant tensor we have

$$\begin{aligned}\langle \omega, \omega \rangle &= \frac{1}{(k!)^2} a_{i_1 \ldots i_k} a^{i_1 \ldots i_k} \\ &= a_I a^I\end{aligned}$$

and as a k-form we want the scalar product to give

$$\begin{aligned}\langle \omega, \omega \rangle &= a_{\vec{I}} a^{\vec{I}} \\ &= \frac{1}{k!} a_I a^I\end{aligned}$$

so the two definitions are different by a factor of $k!$. The definition for forms can be written succinctly as

$$\begin{aligned}&\langle \alpha^1 \wedge \alpha^2 \wedge \cdots \wedge \alpha^k, \beta^1 \wedge \beta^2 \wedge \cdots \wedge \beta^k \rangle \\ &= \det(\langle \alpha^i, \beta^j \rangle)\end{aligned}$$

where the α^i and β^i are 1-forms.

Definition 13.7 *We define the scalar product on $\bigwedge^k \mathrm{V}^* \cong L^k_{alt}(\mathrm{V})$ by first using the above formula for wedge products of 1-forms and then we extending (bi)linearly to all of $\bigwedge^k \mathrm{V}^*$. We can also extend to the whole Grassmann algebra $\bigwedge \mathrm{V}^* = \bigoplus \bigwedge^k \mathrm{V}^*$ by declaring forms of different degree to be orthogonal. We also have the obvious similar definition for $\bigwedge^k \mathrm{V}$ and $\bigwedge \mathrm{V}$.*

We would now like to exhibit the definition of the very useful star operator. This will be a map from $\bigwedge^k \mathrm{V}^*$ to $\bigwedge^{n-k} \mathrm{V}^*$ for each k, $1 \leq k \leq n$ where $n = \dim(M)$. First of all if we have an orthonormal basis $e^1, ..., e^n$ for V^* then $e^1 \wedge \cdots \wedge e^n \in \bigwedge^n \mathrm{V}^*$. But $\bigwedge^n \mathrm{V}^*$ is one dimensional and if $\ell : \mathrm{V}^* \to \mathrm{V}^*$ is any isometry of V^* then $\ell e^1 \wedge \cdots \wedge \ell e^n = \pm e^1 \wedge \cdots \wedge e^n$. In particular, for any permutation σ of the letters $\{1, 2, ..., n\}$ we have $e^1 \wedge \cdots \wedge e^n = sgn(\sigma) e^{\sigma 1} \wedge \cdots \wedge e^{\sigma n}$.

For a given $\mathcal{E}^* = \{e^1, ..., e^n\}$ for V^* (with dual basis for orthonormal basis $\mathcal{E} = \{e_1, ..., e_n\}$) us denote $e^1 \wedge \cdots \wedge e^n$ by $\varepsilon(\mathcal{E}^*)$. Then we have

$$\langle \varepsilon(\mathcal{E}^*), \varepsilon(\mathcal{E}^*) \rangle = \epsilon_1 \epsilon_2 \cdots \epsilon_n = \pm 1$$

and the only elements ω of $\bigwedge^n \mathrm{V}^*$ with $\langle \omega, \omega \rangle = \pm 1$ are $\varepsilon(\mathcal{E}^*)$ and $-\varepsilon(\mathcal{E}^*)$. Given a fixed orthonormal basis $\mathcal{E}^* = \{e^1, ..., e^n\}$, all other orthonormal bases $\mathcal{B}^*$ bases for V^* fall into two classes. Namely, those for which $\varepsilon(\mathcal{B}^*) = \varepsilon(\mathcal{E}^*)$ and those for which $\varepsilon(\mathcal{B}^*) = -\varepsilon(\mathcal{E}^*)$. Each of these two top forms $\pm\varepsilon(\mathcal{E}^*)$ is called a metric volume element for $\mathrm{V}^*, \mathrm{g}^* = \langle , \rangle$. A choice of orthonormal basis determines one of these two volume elements and we call this a choice of an **orientation** for V^*. On the other hand, we have seen that any nonzero top form ω determines an orientation. If we have an orientation given by a top form ω then $\mathcal{E} = \{e_1, ..., e_n\}$ determines the same orientation if and only if $\omega(e_1, ..., e_n) > 0$.

Definition 13.8 *Let an orientation be chosen on* V^* *and let* $\mathcal{E}^* = \{e^1, ..., e^n\}$ *be an oriented orthonormal frame so that* $vol := \varepsilon(\mathcal{E}^*)$ *is the corresponding volume element. Then if* $\mathcal{F} = \{f_1, ..., f_n\}$ *is a basis for* V *with dual basis* $\mathcal{F}^* = \{f^1, ..., f^n\}$ *then*

$$vol = \sqrt{|\det(\mathrm{g}_{ij})|} f^1 \wedge \cdots \wedge f^n$$

where $\mathrm{g}_{ij} = \langle f_i, f_j \rangle$.

Proof. Let $e^i = a^i_j f^j$ then

$$\begin{aligned} \epsilon_i \delta^{ij} &= \pm\delta^{ij} = \langle e^i, e^j \rangle = \langle a^i_k f^k, a^j_m f^m \rangle \\ &= a^i_k a^j_m \langle f^k, f^m \rangle = a^i_k a^j_m \mathrm{g}^{km} \end{aligned}$$

so that $\pm 1 = \det(a^i_k)^2 \det(\mathrm{g}^{km}) = (\det(a^i_k))^2 (\det(\mathrm{g}_{ij}))^{-1}$ and so

$$\sqrt{|\det(\mathrm{g}_{ij})|} = \det(a^i_k).$$

On the other hand,

$$\begin{aligned} \mathrm{vol} &:= \varepsilon(\mathcal{E}^*) = e^1 \wedge \cdots \wedge e^n \\ &= a^1_{k_1} f^{k_1} \wedge \cdots \wedge a^n_{k_1} f^{k_1} = \det(a^i_k) f^1 \wedge \cdots \wedge f^n \end{aligned}$$

and the result follows. ■

Fix an orientation and let $\mathcal{E}^* = \{e^1, ..., e^n\}$ be an orthonormal basis in that orientation class. Then we have chosen one of the two volume forms, say $\mathrm{vol} = \varepsilon(\mathcal{E}^*)$. Now we define $* : \bigwedge^k \mathrm{V}^* \to \bigwedge^{n-k} \mathrm{V}^*$ by first giving the definition on basis elements and then extending by linearity.

Definition 13.9 *Let* V^* *be oriented and let* $\{e^1, ..., e^n\}$ *be a positively oriented orthonormal basis. Let* σ *be a permutation of* $(1, 2, ..., n)$*. On the basis elements* $e^{\sigma(1)} \wedge \cdots \wedge e^{\sigma(k)}$ *for* $\bigwedge^k \mathrm{V}^*$ *define*

$$*(e^{\sigma(1)} \wedge \cdots \wedge e^{\sigma(k)}) = \epsilon_{\sigma_1}\epsilon_{\sigma_2} \cdots \epsilon_{\sigma_k} sgn(\sigma) e^{\sigma(k+1)} \wedge \cdots \wedge e^{\sigma(n)}.$$

In other words,

$$*(e^{i_1} \wedge \cdots \wedge e^{i_k}) = \pm(\epsilon_{i_1}\epsilon_{i_2} \cdots \epsilon_{i_k}) e^{j_1} \wedge \cdots \wedge e^{j_{n-k}}$$

where we take the $+$ *sign iff* $e^{i_1} \wedge \cdots \wedge e^{i_k} \wedge e^{j_1} \wedge \cdots \wedge e^{j_{n-k}} = e^1 \wedge \cdots \wedge e^n$.

Remark 13.3 *In case the scalar product is positive definite* $\epsilon_1 = \epsilon_2 \cdots = \epsilon_n = 1$ *and so the formulas are a bit let cluttered.*

We may develop a formula for the star operator in terms of an arbitrary basis.

Lemma 13.2 *For* $\alpha, \beta \in \bigwedge^k \mathrm{V}^*$ *we have*

$$\langle \alpha, \beta \rangle vol = \alpha \wedge *\beta$$

Proof. It is enough to check this on typical basis elements $e^{i_1} \wedge \cdots \wedge e^{i_k}$ and $e^{m_1} \wedge \cdots \wedge e^{m_k}$. We have

$$\begin{aligned} &(e^{m_1} \wedge \cdots \wedge e^{m_k}) \wedge *(e^{i_1} \wedge \cdots \wedge e^{i_k}) \\ &= e^{m_1} \wedge \cdots \wedge e^{m_k} \wedge (\pm e^{j_1} \wedge \cdots \wedge e^{j_{n-k}}) \end{aligned} \tag{13.1}$$

This latter expression is zero unless $\{m_1, ..., m_k\} \cup \{j_1, ..., j_{n-k}\} = \{1, 2, ..., n\}$ or in other words, unless $\{i_1, ..., i_k\} = \{m_1, ..., m_k\}$. But this is also true for

$$\langle e^{m_1} \wedge \cdots \wedge e^{m_k}, e^{i_1} \wedge \cdots \wedge e^{i_k} \rangle \operatorname{vol}. \tag{13.2}$$

On the other hand if $\{i_1, ..., i_k\} = \{m_1, ..., m_k\}$ then both 13.1 and 13.2 give $\pm$ vol. So the lemma is proved up to a sign. We leave it to the reader to show that the definitions are such that the signs match. ■

Proposition 13.2 *The following identities hold for the star operator:*
1) $*1 = vol$
2) $*vol = (-1)^{\mathrm{ind(g)}}$
3) $**\alpha = (-1)^{\mathrm{ind(g)}}(-1)^{k(n-k)}\alpha$ *for* $\alpha \in \bigwedge^k \mathrm{V}^*$.

Proof. (1) and (2) follow directly from the definitions. For (3) we must first compute $*(e^{j_{k+1}} \wedge \cdots \wedge e^{j_n})$. We must have $*(e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}) = ce^{j_1} \wedge \cdots \wedge e^{j_k}$ for some constant c. On the other hand,

$$\begin{aligned} c\epsilon_1\epsilon_2 \cdots \epsilon_n \operatorname{vol} &= \langle e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}, ce^{j_1} \wedge \cdots \wedge e^{j_k} \rangle \\ &= \langle e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}, *(e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}) \rangle \\ &= \epsilon_{j_{k+1}} \cdots \epsilon_{j_n} sgn(j_{k+1} \cdots j_n, j_1 \cdots j_k) e^{j_{k+1}} \wedge \cdots \wedge e^{j_n} \wedge e^{j_1} \wedge \cdots \wedge e^{j_k} \\ &\epsilon_{j_{k+1}} \cdots \epsilon_{j_n} sgn(j_{k+1} \cdots j_n, j_1 \cdots j_k) \operatorname{vol} \\ &= (-1)^{k(n-k)} \epsilon_{j_{k+1}} \cdots \epsilon_{j_n} \operatorname{vol} \end{aligned}$$

so that $c = \epsilon_{j_{k+1}} \cdots \epsilon_{j_n} (-1)^{k(n-k)}$. Using this we have, for any permutation $J = (j_1, ..., j_n)$,

$$\begin{aligned} **(e^{j_1} \wedge \cdots \wedge e^{j_k}) &= *\epsilon_{j_1}\epsilon_{j_2} \cdots \epsilon_{j_k} sgn(J) e^{j_{k+1}} \wedge \cdots \wedge e^{j_n} \\ &= \epsilon_{j_1}\epsilon_{j_2} \cdots \epsilon_{j_k}\epsilon_{j_{k+1}} \cdots \epsilon_{j_n} sgn(J) e^{j_1} \wedge \cdots \wedge e^{j_k} \\ &= (-1)^{\mathrm{ind(g)}}(-1)^{k(n-k)} e^{j_1} \wedge \cdots \wedge e^{j_k} \end{aligned}$$

which implies the result. ■

13.2 Surface Theory

Let S be a submanifold of $\mathbb{R}^3$. The inverse of a coordinate map $\psi : V \to U \subset \mathbb{R}^2$ is a parameterization $\mathbf{x} : U \to V \subset S$ of a portion U of our surface. Let (u_1, u_2) the coordinates of points in V. For example, the usual parameterization of the sphere

$$\mathbf{x}(\varphi, \theta) = (\cos\theta \sin\varphi, \sin\theta \sin\varphi, \cos\varphi).$$

A curve on a surface S may be given by first letting $t \mapsto (u_1(t), u_2(t))$ be a smooth curve into U and then composing with $\mathbf{x} : U \to S$. For concreteness let the domain of the curve be the interval $[a, b]$. By the ordinary chain rule

$$\dot{\mathbf{x}} = \dot{u}_1 \partial_1 \mathbf{x} + \dot{u}_2 \partial_2 \mathbf{x}$$

and so the length of such a curve is

$$\begin{aligned} L &= \int_a^b |\dot{\mathbf{x}}(t)|\, dt = \int_a^b |\dot{u}_1 \partial_1 \mathbf{x} + \dot{u}_2 \partial_2 \mathbf{x}|\, dt \\ &= \int_a^b (g_{ij} \dot{u}_1 \dot{u}_2)^{1/2} dt \end{aligned}$$

where $g_{ij} = \partial_i \mathbf{x} \cdot \partial_j \mathbf{x}$. Let $p = \mathbf{x}(u_1, u_2)$ be arbitrary in $V \subset S$. The bilinear form g_p given on each $T_p S \subset T\mathbb{R}^3$ where $p = \mathbf{x}(u_1, u_2)$ given by

$$g_p(v, w) = g_{ij} v^i w^j$$

for $v_p = v^1 \partial_1 \mathbf{x} + v^2 \partial_2 \mathbf{x}$ gives a tensor g is called the first fundamental form or metric tensor. The classical notation is $ds^2 = \sum g_{ij} du_j du_j$ which does, whatever it's shortcomings, succinctly encodes the first fundamental form. For example, if we parameterize the sphere $S^2 \subset \mathbb{R}^3$ using the usual spherical coordinates φ, θ we have

$$ds^2 = d\varphi^2 + \sin^2(\varphi) d\theta^2$$

from which the length of a curve $c(t) = \mathbf{x}(\varphi(t), \theta(t))$ is given by

$$L(c) = \int_{t_0}^{t} \sqrt{(\frac{d\varphi}{dt})^2 + \sin^2 \varphi(t) (\frac{d\theta}{dt})^2} dt.$$

Now it may seem that we have something valid only in a single parameterization. Indeed the formulas are given using a single chart and so for instance the curve should not stray from the chart domain V. On the other hand, the expression $g_p(v, w) = g_{ij} v^i w^j$ is an invariant since it is just the length of the vector v as it sits in $\mathbb{R}^3$. So, as the reader has no doubt anticipated, $g_{ij} v^i w^j$ would give the same answer now matter what chart we used. By breaking up a curve into segments each of which lies in some chart domain we may compute it's length using a sequence of integrals of the form $\int (g_{ij} \dot{u}_1 \dot{u}_2)^{1/2} dt$. It is a simple consequence of the chain rule that the result is independent of parameter changes. We also have a well defined notion of surface area of on S. This is given by

$$Area(S) := \int_S dS$$

and where dS is given locally by $\sqrt{g(u^1, u^2)} du^1 du^2$ where $g := \det(g_{ij})$.

We will need to be able to produce normal fields on S. In a coordinate patch we may define

$$\begin{aligned} N &= \partial_1 \mathbf{x}(u_1, u_2) \times \partial_2 \mathbf{x}(u_1, u_2) \\ &= \det \begin{bmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^1}{\partial u^2} & \mathbf{i} \\ \frac{\partial x^2}{\partial u^1} & \frac{\partial x^2}{\partial u^2} & \mathbf{j} \\ \frac{\partial x^3}{\partial u^1} & \frac{\partial x^3}{\partial u^2} & \mathbf{k} \end{bmatrix}. \end{aligned}$$

The unit normal field is then $\mathbf{n} = N/|N|$. Of course, $\mathbf{n}$ is defined independent of coordinates up to sign because there are only two possibilities for a normal direction on a surface in $\mathbb{R}^3$. The reader can easily prove that if the surface is orientable then we may choose a global normal field. If the surface is a closed submanifold (no boundary) then the two choices are characterized as inward and outward.

We have two vector bundles associated with S that are of immediate interest. The first one is just the tangent bundle of S which is in this setting embedded into the tangent bundle of $\mathbb{R}^3$. The other is the normal bundle NS which has as its fiber at $p \in S$ the span of either normal vector $\pm\mathbf{n}$ at p. The fiber is denoted N_pS. Our plan now is to take the obvious connection on $T\mathbb{R}^3$, restrict it to S and then decompose into tangent and normal parts. Restricting to the tangent and normal bundles appropriately,what we end up with is three connections. The obvious connection on $\mathbb{R}^3$ is simply $\overline{\nabla}_\xi(\sum_{i=1}^3 Y^i \frac{\partial}{\partial x^i}) := dY^i(\xi)\frac{\partial}{\partial x^i}$ which exist simply because we have a global distinguished coordinate frame $\{\frac{\partial}{\partial x^i}\}$. The fact that this standard frame is orthonormal with respect to the dot product on $\mathbb{R}^3$ is of significance here. We have both of the following:

1. $\overline{\nabla}_\xi(X \cdot Y) = \overline{\nabla}_\xi X \cdot Y + X \cdot \overline{\nabla}_\xi Y$ for any vector fields X and Y on $\mathbb{R}^3$ and any tangent vector ξ.

2. $\overline{\nabla}_\xi \circ \overline{\nabla}_v = \overline{\nabla}_v \circ \overline{\nabla}_\xi$ (This means the connection has not "torsion" as we define the term later).

Now the connection on the tangent bundle of the surface is defined by projection. Let ξ be tangent to the surface at p and Y a tangent vector field on the surface. Then by definition

$$\nabla_\xi Y = (\overline{\nabla}_\xi Y)^\top$$

where $(\overline{\nabla}_\xi Y)^\top(p)$ is the projection of $\overline{\nabla}_\xi Y$ onto the tangent planes to the surface at p. This gives us a map $\nabla : TS \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ which is easily seen to be a connection. Now there is the left over part $(\overline{\nabla}_\xi Y)^\perp$ but as a map $(\xi, Y) \mapsto (\overline{\nabla}_\xi Y)^\perp$ this does *not* give a connection. On the other hand, if η is a normal field, that is, a section of the normal bundle NS we define $\nabla^\perp_\xi \eta := (\overline{\nabla}_\xi \eta)^\perp$. The resulting map $\nabla : TS \times \Gamma(S, NS) \to \Gamma(S, NS)$ given by $(\xi, \eta) \mapsto \nabla^\perp_\xi \eta$ is indeed a connection on the normal bundle. Here again there is a left over part $(\overline{\nabla}_\xi \eta)^\top$. What about these two left over pieces $(\overline{\nabla}_\xi \eta)^\top$ and $(\overline{\nabla}_\xi Y)^\perp$? These pieces measure the way the surface bends in $\mathbb{R}^3$. We define the shape operator at a point $p \in S$ with respect to a unit normal direction in the following way. First choose the unit normal field $\mathbf{n}$ in the chosen direction as we did above (lets say "outward" for concreteness). Now define $S(p) : T_pS \to T_pS$ by

$$S(p)\xi = \nabla_\xi \mathbf{n}.$$

To see that the result is really tangent to the sphere just notice that $\mathbf{n} \cdot \mathbf{n} = 1$ and so $\nabla_\xi^\perp \mathbf{n}$

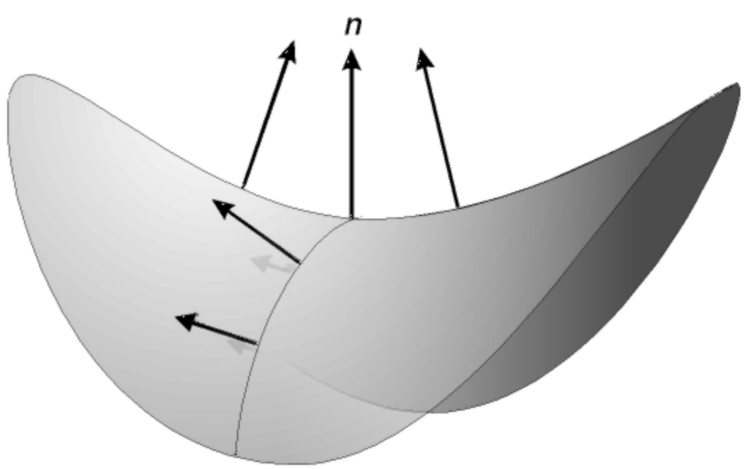
$$\begin{aligned} 0 &= \xi 1 = \xi(\mathbf{n} \cdot \mathbf{n}) \\ &= 2\overline{\nabla}_\xi \mathbf{n} \cdot \mathbf{n} \end{aligned}$$

which means that $\overline{\nabla}_\xi \mathbf{n} \in T_pS$. Thus the fact, that $\mathbf{n}$ had constant length gave us $\overline{\nabla}_\xi \mathbf{n} =(\overline{\nabla}_\xi \mathbf{n})^\top$ and we have made contact with one of the two extra pieces. For a general normal section η we write $\eta = f\mathbf{n}$ for some smooth function on the surface and then

$$\begin{aligned} (\overline{\nabla}_\xi \eta)^\top &= (\overline{\nabla}_\xi f\mathbf{n})^\top \\ &= \left(df(\xi)\mathbf{n} + f\overline{\nabla}_\xi \mathbf{n}\right)^\top \\ &= fS(p)\xi. \end{aligned}$$

so we obtain

Lemma 13.3 $S(p)\xi = f^{-1}(\overline{\nabla}_\xi f\mathbf{n})^\top$



The next result tell us that $S(p) : T_pS \to T_pS$ is symmetric with respect to the first fundamental form.

Lemma 13.4 *Let* $v, w \in T_pS$. *Then we have* $g_p(S(p)v, w) = g_p(v, S(p)w)$.

Proof. The way we have stated the result hide something simple. Namely, tangent vector to the surface are also vectors in $\mathbb{R}^3$ under the usual identification of $T\mathbb{R}^3$ with $\mathbb{R}^3$. With this in mind the result is just $S(p)v \cdot w = v \cdot S(p)w$. Now this is easy to prove. Note that $\mathbf{n} \cdot w = 0$ and so $0 = v(\mathbf{n} \cdot w) = \overline{\nabla}_v \mathbf{n} \cdot w + \mathbf{n} \cdot \overline{\nabla}_v w$. But the same equation holds with v and w interchanged. Subtracting the two expressions gives

$$\begin{aligned} 0 &= \overline{\nabla}_v \mathbf{n} \cdot w + \mathbf{n} \cdot \overline{\nabla}_v w \\ &- \left(\overline{\nabla}_w \mathbf{n} \cdot v + \mathbf{n} \cdot \overline{\nabla}_w v\right) \\ &= \overline{\nabla}_v \mathbf{n} \cdot w - \overline{\nabla}_w \mathbf{n} \cdot v + \mathbf{n} \cdot (\overline{\nabla}_v w - \overline{\nabla}_w v) \\ &= \overline{\nabla}_v \mathbf{n} \cdot w - \overline{\nabla}_w \mathbf{n} \cdot v \end{aligned}$$

from which the result follows. ■

Since $S(p)$ is symmetric with respect to the dot product there are eigenvalues κ_1, κ_2 and eigenvectors $v_{\kappa_1}, v_{\kappa_2}$ such that $v_{\kappa_i} \cdot S(p)v_{\kappa_j} = \delta_{ij}\kappa_i$. Let us calculate in a special coordinate system containing our point p obtained by projecting onto the tangent plane there. Equivalently, we rigidly move the surface until p is at the origin of $\mathbb{R}^3$ and is tangent to the x, y plane. Then the surface is parameterized near p by $(u^1, u^2) \mapsto (u^1, u^2, f(u^1, u^2))$ for some smooth function f with $\frac{\partial f}{\partial u^1}(0) = \frac{\partial f}{\partial u^2}(0) = 0$. At the point p which is now the origin we have $g_{ij}(0) = \delta_{ij}$. Since S is now the graph of the function f the tangent space T_pS is identified with the x, y plane. A normal field is given by $\operatorname{grad} F = \operatorname{grad}(f(x,y) - z) = (\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, -1)$ and the unit normal is

$$\mathbf{n}(u^1, u^2) = \frac{1}{\sqrt{(\frac{\partial f}{\partial u^1})^2 + (\frac{\partial f}{\partial u^2})^2 + 1}}(\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, -1)$$

Letting $r(u^1, u^2) := ((\frac{\partial f}{\partial u^1})^2 + (\frac{\partial f}{\partial u^2})^2 + 1)^{1/2}$ and using lemma 13.3 we have $S(p)\xi = r^{-1}(\overline{\nabla}_\xi r\mathbf{n})^\top = r^{-1}(\overline{\nabla}_\xi N)^\top$ where $N := (\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, -1)$. Now at the origin $r = 1$ and so $\xi \cdot S(p)\xi = \overline{\nabla}_\xi N \cdot \xi = \frac{\partial}{\partial u^k}\frac{\partial F}{\partial u^i}\xi^k\xi^i = \frac{\partial^2 F}{\partial u^k \partial u^i}\xi^k\xi^i$ from which we get the following:

$$\xi \cdot S(p)v = \sum_{ij} \xi^i v^j \frac{\partial f}{\partial u^i \partial u^j}(0)$$

valid for these special type of coordinates and only at the central point p. Notice that this means that once we have the surface positioned as a graph over the x, y-plane and parameterized as above then

$$\xi \cdot S(p)v = D^2 f(\xi, v) \text{ at } 0.$$

Here we must interpret ξ and v on the right hand side to be (ξ^1, ξ^2) and (v^1, v^2) where as on the left hand side $\xi = \xi^1 \frac{\partial \mathbf{x}}{\partial u^1} + \xi^2 \frac{\partial \mathbf{x}}{\partial u^2}$, $v = v^1 \frac{\partial \mathbf{x}}{\partial u^1} + v^2 \frac{\partial \mathbf{x}}{\partial u^2}$.

Exercise 13.3 *Position S to be tangent to the x, y plane as above. Let the x, z plane intersect S in a curve c_1 and the y, z plane intersect S in a curve c_2. Show that by rotating we can make the coordinate vectors $\frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2}$ be eigenvectors for $S(p)$ and that the curvatures of the two curves at the origin are κ_1 and κ_2.*

We have two important invariants at any point p. The first is the Gauss curvature $K := \det(S) = \kappa_1\kappa_2$ and the second is the mean curvature $H = \frac{1}{2} trace(S) = \frac{1}{2}(\kappa_1 + \kappa_2)$.

The sign of H depends on which of the two normal directions we have chosen while the sign of κ does not. In fact, the Gauss curvature turns out to be "intrinsic" to the surface in the sense that it remains constant under any deformation of the surface the preserves lengths of curves. More on this below but first let us establish a geometric meaning for H. First of all, we may vary the point p and then S becomes a function of p and the same for H (and K).

Theorem 13.1 *Let S_t be a family of surfaces given as the image of maps $h_t : S \to \mathbb{R}^3$ and given by $p \mapsto p + t\mathbf{v}$ where $\mathbf{v}$ is a section of $T\mathbb{R}^3\big|_S$ with $\mathbf{v}(0) = 1$ and compact support. Then*

$$\left.\frac{d}{dt}\right|_{t=0} area(S_t) = -\int_S (\mathbf{v} \cdot H\mathbf{n}) dS$$

More generally, the formula is true if $h : (-\varepsilon, \varepsilon) \times S \to \mathbb{R}^3$ is a smooth map and $\mathbf{v}(p) := \left.\frac{d}{dt}\right|_{t=0} h(t, p)$.

Exercise 13.4 *Prove the above theorem by first assuming that $\mathbf{v}$ has support inside a chart domain and then use a partition of unity argument to get the general case.*

Surface S is called a **minimal surface** if $H \equiv 0$ on S. It follows from theorem 13.1 that if S_t is a family of surfaces given as in the theorem that if S_0 is a minimal surface then 0 is a critical point of the function $a(t) := area(S_t)$. Conversely, if 0 is a critical point for all such variations of S then S is a minimal surface.

Exercise 13.5 *Show that Sherk's surface, which is given by $e^z \cos(y) = \cos x$, is a minimal surface. If you haven't seen this surface do a plot of it using Maple or some other graphing software. Do the same for The helicoid $y \tan z = x$.*

Now we move on to the Gauss curvature K. Here the most important fact is that K may be written in terms of the first fundamental form. The significance of this is that if S_1 and S_2 are two surfaces and if there is a map $\phi : S_1 \to S_2$ which preserves the length of curves, then κ^{S_1} and κ^{S_2} are the same in the sense that $K^{S_1} = K^{S_2} \circ \phi$. In the following theorem, "$g_{ij} = \delta_{ij}$ to first order" means that $g_{ij}(0) = \delta_{ij}$ and $\frac{\partial g_{ij}}{\partial u}(0) = \frac{\partial g_{ij}}{\partial v}(0) = 0$.

Theorem 13.2 (Gauss's Theorema Egregium) *Let $p \in S$. There always exist coordinates u, v centered at p (so $u(p) = 0, v(p) = 0$) such that $g_{ij} = \delta_{ij}$ to first order at 0 and for which we have*

$$K(p) = \frac{\partial^2 g_{12}}{\partial u \partial v}(0) - \frac{1}{2}\frac{\partial^2 g_{22}}{\partial u^2}(0) - \frac{1}{2}\frac{\partial^2 g_{11}}{\partial v^2}(0).$$

Proof. In the coordinates described above which give the parameterization $(u, v) \mapsto (u, v, f(u, v))$ where p is the origin of $\mathbb{R}^3$ we have

$$\begin{bmatrix} g_{11}(u,v) & g_{12}(u,v) \\ g_{21}(u,v) & g_{22}(u,v) \end{bmatrix} = \begin{bmatrix} 1 + (\frac{\partial f}{\partial x})^2 & \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} & 1 + (\frac{\partial f}{\partial y})^2 \end{bmatrix}$$

from which we find after a bit of straight forward calculation

$$\begin{aligned} &\frac{\partial^2 g_{12}}{\partial u \partial v}(0) - \frac{1}{2}\frac{\partial^2 g_{22}}{\partial u^2}(0) - \frac{1}{2}\frac{\partial^2 g_{11}}{\partial v^2}(0) \\ &= \frac{\partial^2 f}{\partial u^2}\frac{\partial^2 f}{\partial v^2} - \frac{\partial^2 f}{\partial u \partial v} = \det D^2 f(0) \\ &= \det S(p) = K(p). \end{aligned}$$

■

Note that if we have any other coordinate system s, t centered at p then writing $(u, v) = (x^1, x^2)$ and $(s, t) = (\bar{x}^1, \bar{x}^2)$ we have the transformation law

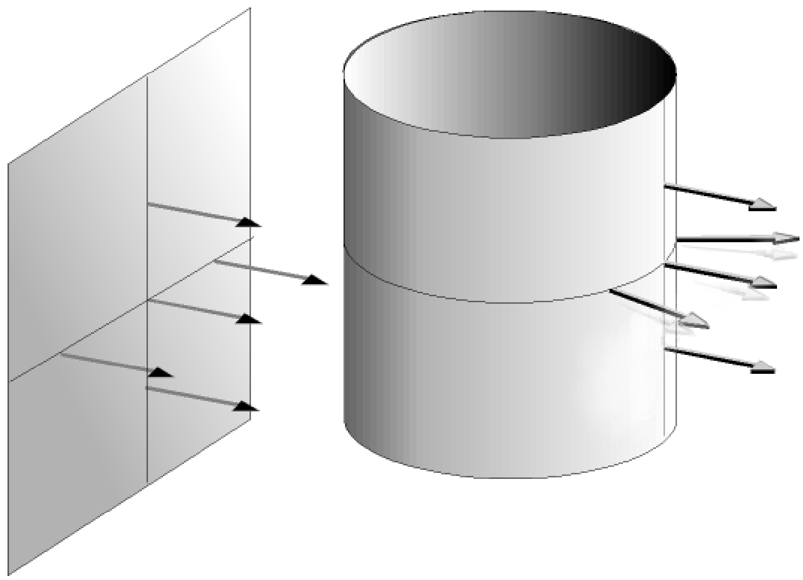
$$\bar{g}_{ij} = g_{kl}\frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^l}{\partial \bar{x}^j}$$

which means that if we know the metric components in any coordinate system then we can get them, and hence $K(p)$, at any point in any coordinate system. The conclusion is the that the metric determines the Gauss curvature. We say that K is an **intrinsic invariant**.

13.3 Riemannian and semi-Riemannian Metrics

Consider a regular submanifold M of a Euclidean space, say $\mathbb{R}^n$. Since we identify T_pM as a subspace of $T_p\mathbb{R}^n \cong \mathbb{R}^n$ and the notion of length of tangent vectors makes sense on $\mathbb{R}^n$ it also makes sense for vectors in T_pM. In fact, if $X_p, Y_p \in T_pM$ and c_1, c_2 are some curves with $\dot{c}_1(0) = X_p$, $\dot{c}_2(0) = Y_p$ then c_1 and c_2 are also a curves in $\mathbb{R}^n$. Thus we have an inner product defined $\mathrm{g}_p(X_p, Y_p) = \langle X_p, Y_p \rangle$. For a manifold that is not given as submanifold of some $\mathbb{R}^n$ we must have an inner product assigned to each tangent space as part of an extra structure. The assignment of a nondegenerate symmetric bilinear form $\mathrm{g}_p \in T_pM$ for every p in a smooth way defines a tensor field $\mathrm{g} \in \mathfrak{X}^0_2(M)$ on M called metric tensor.

Definition 13.10 *If $\mathrm{g} \in \mathfrak{X}^0_2(M)$ is nondegenerate, symmetric and positive definite at every tangent space we call g a **Riemannian metric** (tensor). If g is a Riemannian metric then we call the pair M, g a **Riemannian manifold** .*



The Riemannian manifold as we have defined it is the notion that best generalizes to manifolds the metric notions from surfaces such as arc length of a curve, area (or volume), curvature and so on. But because of the structure of spacetime as expressed by general relativity we need to allow the metric to be indefinite. In this case, some vectors might have negate or zero length.

Recall the index of a bilinear form is the number of negative ones appearing in the signature.

Definition 13.11 *If* $\mathrm{g} \in \mathfrak{X}_2^0(M)$ *is symmetric nondegenerate and has constant index on* M *then we call* g *a semi-Riemannian metric and* M, g *a* ***semi-Riemannian manifold*** *or* ***pseudo-Riemannian manifold.*** *The index is called the* ***index*** *of* M, g *and denoted ind*(M)*. The signature is also constant and so the manifold has a signature also. If the index of a semi-Riemannian manifold (with* $\dim(M) \geq 2$*) is* $(-1, +1, +1 + 1, ...)$ *(or according to some conventions* $(1, -1, -1 - 1, ...)$*) then the manifold is called a* ***Lorentz manifold*** .

The simplest family of semi-Riemannian manifolds are the spaces $\mathbb{R}^n_\nu$ which are the Euclidean spaces $\mathbb{R}^n$ endowed with the scalar products given by

$$\langle x, y\rangle_\nu = -\sum_{i=1}^{\nu} x^i y^i + \sum_{i=\nu+1}^{n} x^i y^i.$$

Since ordinary Euclidean geometry does not use indefinite scalar products we shall call the spaces $\mathbb{R}^n_\nu$ pseudo-Euclidean spaces when the index ν is not zero.

If we write just $\mathbb{R}^n$ then either we are not concerned with a scalar product at all or the scalar product is assumed to be the usual inner product ($\nu = 0$).Thus a Riemannian metric is just the special case of index 0.

Definition 13.12 *Let M, g and N, h be two semi-Riemannian manifolds. A diffeomorphism $\Phi : M \to N$ is called an **isometry** if $\Phi^*\mathrm{h} = \mathrm{g}$. Thus for an isometry $\Phi : M \to N$ we have $\mathrm{g}(v, w) = \mathrm{h}(T\Phi \cdot v, T\Phi \cdot w)$ for all $v, w \in TM$. If $\Phi : M \to N$ is a local diffeomorphism such that $\Phi^*\mathrm{h} = \mathrm{g}$ is called a **local isometry.***

Example 13.1 *We have seen that a regular submanifold of a Euclidean space $\mathbb{R}^n$ is a Riemannian manifold with the metric inherited from $\mathbb{R}^n$. In particular, the sphere $S^{n-1} \subset \mathbb{R}^n$ is a Riemannian manifold. Every isometry of S^{n-1} is the restriction to S^{n-1} of an isometry of $\mathbb{R}^n$ that fixed the origin (and consequently fixes S^{n-1}).*

One way to get a variety of examples of semi-Riemannian manifolds is via a group action by isometries. Let us here consider the case of a discrete group that acts smoothly, properly, freely and by isometries. We have already seen that if we have an action $\rho : G \times M \to M$ satisfying the first three conditions then the quotient space M/G has a unique structure as a smooth manifold such that the projection $\kappa : M \to M/G$ is a covering. Now since G acts by isometries $\rho_g^*\langle ., . \rangle = \langle ., . \rangle$ for all $g \in G$. The tangent map $T\kappa : TM \to T(M/G)$ is onto and so for any $\bar{v}_{\kappa(p)} \in T_{\kappa(p)}(M/G)$ there is a vector $v_p \in T_pM$ with $T_p\kappa.v_p = \bar{v}_{\kappa(p)}$. In fact there is more than one such vector in T_pM (except in the trivial case $G = \{e\}$) but if $T_p\kappa.v_p = T_q\kappa.w_q$ then there is a $g \in G$ such that $\rho_g p = q$ and $T_p\rho_g v_p = w_q$. Conversely, if $\rho_g p = q$ then $T_p\kappa.(T_p\rho_g v_p) = T_q\kappa.w_q$. Now for $\bar{v}_1, \bar{v}_2 \in T_pM$ define $h(\bar{v}_1, \bar{v}_2) = \langle v_1, v_2 \rangle$ where v_1 and v_2 are chosen so that $T\kappa.v_i = \bar{v}_i$. From our observations above this is well defined. Indeed, if $T_p\kappa.v_i = T_q\kappa.w_i = \bar{v}_i$ then there is an isometry ρ_g with $\rho_g p = q$ and $T_p\rho_g v_i = w_i$ and so

$$\langle v_1, v_2 \rangle = \langle T_p\rho_g v_1, T_p\rho_g v_2 \rangle = \langle w_1, w_2 \rangle.$$

It is easy to show that $x \mapsto h_x$ defined a metric on M/G with the same signature as that of $\langle ., . \rangle$ and $\kappa^* h = \langle ., . \rangle$. In fact we will use the same notation for either the metric on M/G or on M which is not such an act of violence since κ is now a local isometry. The simplest example of this construction is $\mathbb{R}^n/\Gamma$ for some lattice Γ and where we use the canonical Riemannian metric on $\mathbb{R}^n$. In case the lattice is isomorphic to $\mathbb{Z}^n$ then $\mathbb{R}^n/\Gamma$ is called a flat torus of dimension n. Now each of these tori are locally isometric but may not be globally so. To be more precise, suppose that $f_1, f_2, ..., f_n$ is a basis for $\mathbb{R}^n$ which is not necessarily orthonormal. Let Γ_f be the lattice consisting of integer linear combinations of $f_1, f_2, ..., f_n$. The question now is what if we have two such lattices Γ_f and $\Gamma_{\bar{f}}$ when is $\mathbb{R}^n/\Gamma_f$ isometric to $\mathbb{R}^n/\Gamma_{\bar{f}}$? Now it may seem that since these are clearly diffeomorphic and since they are locally isometric then they must be (globally) isometric. But this is not the case. We will be able to give a good reason for this shortly but for now we let the reader puzzle over this.

Every smooth manifold that admits partitions of unity also admits at least one (in fact infinitely may) Riemannian metrics. This includes all finite dimensional paracompact manifolds. The reason for this is that the set of all Riemannian metric tensors is, in an appropriate sense, convex. To wit:

Proposition 13.3 *Every smooth manifold admits a Riemannian metric.*

Proof. As in the proof of 13.3 above we can transfer the Euclidean metric onto the domain U_α of any given chart via the chart map ψ_α. The trick is to piece these together in a smooth way. For that we take a smooth partition of unity U_α, ρ_α subordinate to a cover by charts U_α, ψ_α. Let g_α be any metric on U_α and define

$$\mathrm{g}(p) = \sum \rho_\alpha(p) g_\alpha(p).$$

The sum is finite at each $p \in M$ since the partition of unity is locally finite and the functions $\rho_\alpha \mathrm{g}_\alpha$ are extended to be zero outside of the corresponding U_α. The fact that $\rho_\alpha \geq 0$ and $\rho_\alpha > 0$ at p for at least one α easily gives the result that g positive definite is a Riemannian metric on M. ■

The length of a tangent vector $X_p \in T_pM$ in a *Riemannian* manifold is given by $\sqrt{\mathrm{g}(X_p, X_p)} = \sqrt{\langle X_p, X_p \rangle}$. In the case of an indefinite metric ($\nu > 0$) we will need a classification:

Definition 13.13 *A tangent vector $\nu \in T_pM$ to a semi-Riemannian manifold M is called*

1. *spacelike if $\langle \nu, \nu \rangle > 0$*
2. *lightlike or null if $\langle \nu, \nu \rangle = 0$*
3. *timelike if $\langle \nu, \nu \rangle < 0$.*

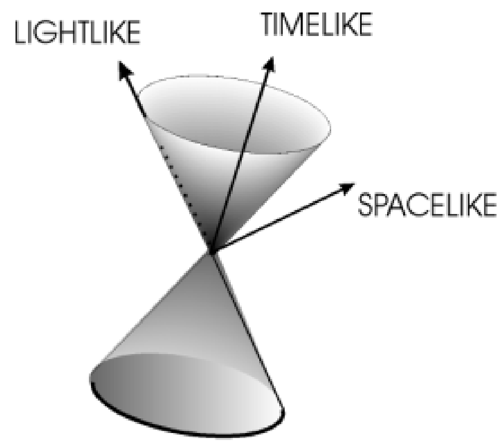
Definition 13.14 *The set of all timelike vectors T_pM in is called the **light cone** at p.*

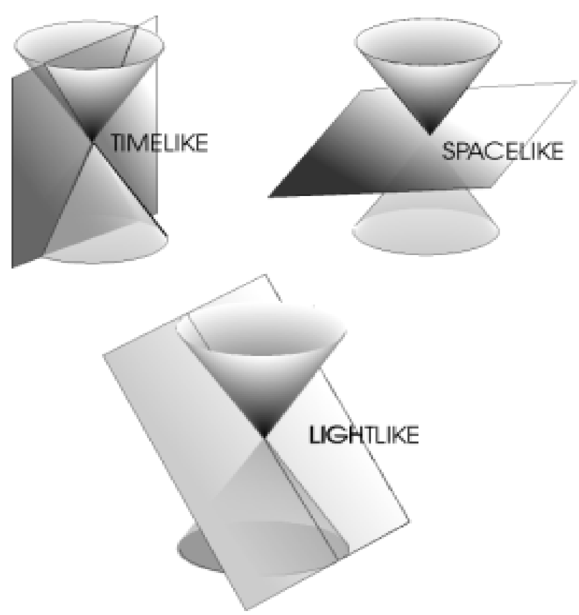
Definition 13.15 *Let $I \subset \mathbb{R}$ be some interval. A curve $c : I \to M, \mathrm{g}$ is called spacelike, lightlike, or timelike according as $\dot{c}(t) \in T_{c(t)}M$ is spacelike, lightlike, or timelike respectively for all $t \in I$.*

For Lorentz spaces, that is for semi-Riemannian manifolds with index equal to 1 and dimension greater than or equal to 2, we may also classify subspaces into three categories:

Definition 13.16 *Let M, g be a Lorentz manifold. A subspace $\mathrm{W} \subset T_pM$ of the tangents space is called*

1. *spacelike if $\mathrm{g}|_{\mathrm{W}}$ is positive definite,*
2. *time like if $\mathrm{g}|_{\mathrm{W}}$ nondegenerate with index 1,*


LIGHTLIKE
TIMELIKE
SPACELIKE


TIMELIKE
SPACELIKE
LIGHTLIKE

3. *lightlike if* $\mathrm{g}|_{\mathrm{W}}$ *is degenerate.*

Remark 13.4 (Notation) *We will usually write* $\langle X_p, Y_p\rangle$ *or* $\mathrm{g}(X_p, Y_p)$ *in place of* $\mathrm{g}(p)(X_p, X_p)$. *Also, just as for any tensor field we define the function* $\langle X, Y\rangle$ *which for a pair of vector fields is given by* $\langle X, Y\rangle(p) = \langle X_p, Y_p\rangle$.

In local coordinates $(x^1, ..., x^n)$ on $U \subset M$ we have that $\mathrm{g}|_U = \sum \mathrm{g}_{ij} dx^i \otimes dx^j$ where $\mathrm{g}_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\rangle$. Thus if $X = \sum X^i \frac{\partial}{\partial x^i}$ and $Y = \sum Y^i \frac{\partial}{\partial x^i}$ on U then

$$\langle X, Y\rangle = \sum \mathrm{g}_{ij} X^i Y^i \tag{13.3}$$

Remark 13.5 *The expression* $\langle X, Y\rangle = \sum \mathrm{g}_{ij} X^i Y^i$ *means that for all* $p \in U$ *we have* $\langle X(p), Y(p)\rangle = \sum \mathrm{g}_{ij}(p) X^i(p) Y^i(p)$ *where as we know that functions* X^i *and* Y^i *are given by* $X^i = dx^i(X)$ *and* $Y^i = dx^i(Y)$.

Recall that a continuous curve $c : [a, b] \to M$ into a smooth manifold is called **piecewise smooth** if there exists a partition $a = t_0 < t_1 < \cdots < t_k = b$ such that c restricted to $[t_i, t_{i+1}]$ is smooth for $0 \leq i \leq k - 1$. Also, a curve $c : [a, b] \to M$ is called **regular** if it has a nonzero tangent for all $t \in [a, b]$.

Definition 13.17 *Let* M, g *be Riemannian. If* $c : [a, b] \to M$ *is a (piecewise smooth) curve then the length of the curve from* $c(a)$ *to* $c(b)$ *is defined by*

$$\operatorname*{length}_{c(a)\to c(b)}(c) = \int_a^t \langle \dot{c}(t), \dot{c}(t)\rangle^{1/2} dt. \tag{13.4}$$

Definition 13.18 *Let* M, g *be semi-Riemannian. If* $c : [a, b] \to M$ *is a (piecewise smooth) timelike or spacelike curve then*

$$\tau_{c(a),c(b)}(c) = \int_a^b |\langle \dot{c}(t), \dot{c}(t)\rangle|^{1/2} dt$$

is called the ***length*** *of the curve.*

In general, if we wish to have a positive real number for a length then in the semi-Riemannian case we need to include absolute value signs in the definition so the proper time is just the timelike special case of a generalized arc length defined for any smooth curve by $\int_a^b |\langle \dot{c}(t), \dot{c}(t)\rangle|^{1/2} dt$ but unless the curve is either timelike or spacelike this arc length can have some properties that are decidedly not like our ordinary notion of length. In particular, curve may connect two different points and the generalized arc length might still be zero! It becomes clear that we are not going to be able to define a metric distance function as we soon will for the Riemannian case.

Definition 13.19 *A* ***positive reparameterization*** *of a piecewise smooth curve* $c : I \to M$ *is a curve defined by composition* $c \circ f^{-1} : J \to M$ *where* $f : I \to J$ *is a piecewise smooth bijection that has* $f' > 0$ *on each subinterval* $[t_{i-1}, t_i] \subset I$ *where* c *is smooth.*

Remark 13.6 (important fact) *The integrals above are well defined since $c'(t)$ is defined except for a finite number of points in $[a,b]$. Also, it is important to notice that by standard change of variable arguments a positive reparameterization $\widetilde{c}(u) = c(f^{-1}(u))$ where $u = f(t)$ does not change the (generalized) length of the curve*

$$\int_a^b |\langle \dot{c}(t), \dot{c}(t)\rangle|^{1/2} dt = \int_{f^{-1}(a)}^{f^{-1}(b)} |\langle \widetilde{c}'(u), \widetilde{c}'(u)\rangle|^{1/2} du.$$

Thus the (generalized) length of a piecewise smooth curve is a geometric property of the curve; i.e. a semi-Riemannian invariant.

13.4 The Riemannian case (positive definite metric)

Once we have a notion of the length of a curve we can then define a distance function (metric in the sense of "metric space") as follow. Let $p, q \in M$. Consider the set path(p,q) of all smooth curves which begin at p and end at q. We define

$$\mathrm{dist}(p,q) = \inf\{l \in \mathbb{R} : l = \mathrm{length}(c) \text{ and } c \in \mathrm{path}(p,q)\}. \tag{13.5}$$

or a general manifold just because $\mathrm{dist}(p,q) = r$ does **not necessarily** mean that there must be a curve connecting p to q having length r. To see this just consider the points $(-1,0)$ and $(1,0)$ on the punctured plane $\mathbb{R}^2 - 0$.

Theorem 13.3 (distance topology) *Given a Riemannian manifold, define the distance function* dist *as above. Then M,* dist *is a metric space and the induced topology coincides with the manifold topology on M.*

Proof. That dist is true distance function (metric) we must show that
(1) dist is symmetric,
(2) dist satisfies the triangle inequality,
(3) $\mathrm{dist}(p,q) \geq 0$ and
(4) $\mathrm{dist}(p,q) = 0$ iff $p = q$.

Now (1) is obvious and (2) and (3) are clear from the properties of the integral and the metric tensor. To prove (4) we need only show that if $p \neq q$ then $\mathrm{dist}(p,\dot{q}) > 0$. Choose a chart ψ_α, U_α containing p but not q (M is Hausdorff). Now since $\psi_\alpha(U_\alpha) \subset \mathbb{R}^n$ we can transfer the Euclidean distance to U_α and define a small Euclidean ball $B_{Euc}(p,r)$ in this chart. Now any path from p to q must hit the boundary sphere $S(r) = \partial B_{Euc}(p,r)$. Now by compactness of $\bar{B}_{Euc}(p,r)$ we see that there are constants C_0 and C_1 such that $C_1\delta_{ij} \geq \mathrm{g}_{ij}(x) \geq C_0\delta_{ij}$ for all $x \in \bar{B}_{Euc}(p,r)$. Now any piecewise smooth curve $c : [a,b] \to M$ from p to q hits $S(r)$ at some parameter value $b_1 \leq b$ where we may assume this is the first hit (i.e. $c(t) \in B_{Euc}(p,r)$ for $a \leq t < b_0$). Now there is a curve that goes directly from p to q with respect to the Euclidean distance; i.e. a radial curve in the given Euclidean coordinates. This curve is given in coordinates as $\delta_{p,q}(t) = \frac{1}{b-1}(b-t)x(p) + \frac{1}{b-a}(t-a)x(q)$. Thus we have

$$\begin{aligned}\text{length}(c) &\geq \int_a^{b_0} \left(\mathrm{g}_{ij} \frac{d(x^i \circ c)}{dt} \frac{d(x^j \circ c)}{dt} \right)^{1/2} dt \geq C_0^{1/2} \int_a^{b_0} \left(\delta_{ij} \frac{d(x^i \circ c)}{dt} \right)^{1/2} dt \\ &= C_0^{1/2} \int_a^{b_0} |c'(t)| \, dt \geq C_0^{1/2} \int_a^{b_0} \left| \delta'_{p,q}(t) \right| dt = C_0^{1/2} r.\end{aligned}$$

Thus we have that $\text{dist}(p,q) = \inf\{\text{length}(c) : c \text{ a curve from } p \text{ to } q\} \geq C_0^{1/2} r > 0$. This last argument also shows that if $\text{dist}(p,x) < C_0^{1/2} r$ then $x \in B_{Euc}(p,r)$. This means that if $B(p, C_0^{1/2} r)$ is a ball with respect to dist then $B(p, C_0^{1/2} r) \subset B_{Euc}(p,r)$. Conversely, if $x \in B_{Euc}(p,r)$ then letting $\delta_{p,x}$ a "direct curve" analogous to the one above that connects p to x we have

$$\begin{aligned}\text{dist}(p,x) &\leq \text{length}(\delta_{p,x}) \\ &= \int_a^{b_0} \left(\mathrm{g}_{ij} \frac{d(x^i \circ \delta)}{dt} \frac{d(x^j \circ \delta)}{dt} \right)^{1/2} dt \\ &\leq C_1^{1/2} \int_a^{b_0} \left| \delta'_{p,x}(t) \right| dt = C_1^{1/2} r\end{aligned}$$

so we conclude that $B_{Euc}(p,r) \subset B(p, C_1^{1/2} r)$. Now we have that inside a chart, every dist-ball contains a Euclidean ball and visa vera. Thus since the manifold topology is generated by open subsets of charts we see that the two topologies coincide as promised. ■

13.5 Levi-Civita Connection

In the case of the semi-Riemannian spaces $\mathbb{R}^n_\nu$ one can identify vector fields with maps $\mathsf{X}{:}\mathbb{R}^n_\nu \to \mathbb{R}^n_\nu$ and thus it makes sense to differentiate a vector field just as we would a function. For instance, if $\mathsf{X}= (f^1, ..., f^n)$ then we can define the directional derivative in the direction of v at $p \in T_p\mathbb{R}^n_\nu \cong \mathbb{R}^n_\nu$ by $\nabla_v \mathsf{X} = (D_p f^1 \cdot v, ..., D_p f^n \cdot v)$ and we get a vector in $T_p\mathbb{R}^n_\nu$ as an answer. Taking the derivative of a vector field seems to require involve the limit of difference quotient of the type

$$\lim_{t \to 0} \frac{X(p+tv) - X(p)}{t}$$

and yet how can we interpret this in a way that makes sense for a vector field on a general manifold? One problem is that $p + tv$ makes no sense if the manifold isn't a vector space. This problem is easily solve by replacing $p + tv$ by $c(t)$ where $\dot{c}(0) = v$ and $c(0) = p$. We still have the more serious problem that $X(c(t)) \in T_{c(t)}M$ while $X(p) = X(c(0)) \in T_pM$. The difficulty is that $T_{c(t)}M$ is not likely to be the same vector space as T_pM and so what sense does $X(c(t)) - X(p)$ make? In the case of a vector space (like $\mathbb{R}^n_\nu$) every tangent space is canonically isomorphic to the vector space itself so there is sense to be made of a difference quotient involving vectors from different tangent spaces. In order to get an idea of how we might define $\nabla_v X$ on a general manifold, let us look again at the case of a submanifold M of $\mathbb{R}^n$. Let $X \in \mathfrak{X}(M)$ and $v \in T_pM$. Form a curve with $\dot{c}(0) = v$ and $c(0) = p$ and consider the composition $X \circ c$. Since every vector tangent to M is also a vector in $\mathbb{R}^n$ we can consider $X \circ c$ to take values in $\mathbb{R}^n$ and then take the derivative

$$\left.\frac{d}{dt}\right|_0 X \circ c.$$

This is well defined but while $X \circ c(t) \in T_{c(t)}M \subset T_{c(t)}\mathbb{R}^n$ we only know that $\left.\frac{d}{dt}\right|_0 X \circ c \in T_p\mathbb{R}^n$. A good answer should have been in T_pM. The simple solution is to take the orthogonal projection of $\left.\frac{d}{dt}\right|_0 X \circ c$ onto $T_{c(0)}M$. Our tentative definition is then

$$\nabla_v X := \left(\left.\frac{d}{dt}\right|_0 X \circ c \right)^{\perp} \in T_pM.$$

This turns out to be a very good definition since it turns out that we have the following nice results:

1. (Smoothness) If X and Y are smooth vector fields then the map

$$p \mapsto \nabla_{X_p} Y$$

   is also a smooth vector field on M. This vector filed is denoted $\nabla_X Y$.

2. (Linearity over $\mathbb{R}$ in second “slot”) For two vector fields X and Y and any $a, b \in \mathbb{R}$ we have

$$\nabla_v(aX_1 + bX_2) = a\nabla_v X_1 + b\nabla_v X_2.$$

3. (Linearity over $C^\infty(M)$ in first “slot”)For any three vector fields X, Y and Z and any $f, g \in C^\infty(M)$ we have

$$\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z.$$

4. (Product rule) For $v \in T_pM$, $X \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$ we have

$$\begin{aligned} \nabla_v fX &= f(p)\nabla_v X + (vf)X(p) \\ &= f(p)\nabla_v X + df(v)X(p). \end{aligned}$$

   Or in terms of two fields X, Y

$$\nabla_X fY = f\nabla_X Y + (Xf)Y.$$

5. $\nabla_v\langle X, Y\rangle = \langle \nabla_v X, Y\rangle + \langle X, \nabla_v Y\rangle$ for all v, X, Y.

Now if one takes the approach of abstracting these properties with the aim of defining a so called covariant derivative it is a bit unclear whether we should define $\nabla_X Y$ for a pair of fields X, Y or define $\nabla_v X$ for a tangent vector v and a field X. It turns out that one can take either approach and when done properly we end up with equivalent notions. We shall make the following our basic definition of a covariant derivative.

Definition 13.20 *A* ***natural covariant derivative*** *(or* ***connection***[1]*)* ∇ *on a smooth manifold* M *is an assignment to each open set* $U \subset M$ *of a map* $\nabla^U : \mathfrak{X}(U) \times \mathfrak{X}(U) \to \mathfrak{X}(U)$ *written* $\nabla^U : (X, Y) \to \nabla^U_X Y$ *such that the following hold:*

1. $\nabla^U_X Y$ *is* $C^\infty(U)$*-linear in* X,
2. $\nabla^U_X Y$ *is* $\mathbb{R}$*-linear in* Y,
3. $\nabla^U_X (fY) = f\nabla^U_X Y + (Xf)Y$ *for all* $X, Y \in \mathfrak{X}(U)$ *and all* $f \in C^\infty(U)$.
4. *If* $V \subset U$ *then* $r^U_V(\nabla^U_X Y) = \nabla^V_{r^U_V X} r^U_V Y$ *(naturality with respect to restrictions).*
5. $(\nabla^U_X Y)(p)$ *only depends of the value of* X *at* p *(infinitesimal locality).*

Here $\nabla^U_X Y$ *is called the covariant derivative of* Y *with respect to* X*. We will denote all of the maps* ∇^U *by the single symbol* ∇ *when there is no chance of confusion.*

Definition 13.21 *If* M *is endowed with a semi-Riemannian metric* $\mathrm{g} = \langle .,. \rangle$ *then a connection* ∇ *on* M *is called a* ***metric covariant derivative*** *(for* g*) iff*

MC $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$
for all $X, Y, Z \in \mathfrak{X}(U)$ *and all* $U \subset M$

We have worked in naturality with respect to restriction and infinitesimal locality in order to avoid a discussion of the technicalities of localization and globalization on infinite dimensional manifolds. If the connection arises from a spray or system of Christoffel symbols as defined below then these conditions would follow automatically from the rest (see the proposition below). We also have the following intermediate result.

[1]It would be better if we could avoid the term connection at this point and use covariant derivative instead since later we will encounter another definition of a connection which refers to a certain "horizontal" subbundle of the tangent bundle of the total space of a vector bundle (or also of a principal bundle). Connections in this sense determine a natural covariant derivative. Conversely, a natural covariant derivative determines a horizontal subbundle, i.e. a connection in this second (more proper?) sense. Thus a natural covariant derivative and a connection are mutually determining.

Lemma 13.5 *Suppose that M admits cut-off functions and $\nabla^M : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is such that (1), (2) and (3) hold (for $U = M$). Then if on some open U either $X = 0$ or $Y = 0$ then*

$$(\nabla_X^M Y)(p) = 0 \text{ for all } p \in U.$$

Proof. We prove the case of $Y|_U = 0$ and leave the case of $X|_U = 0$ to the reader.

Let $q \in U$. Then there is some function f which is identically one on a neighborhood $V \subset U$ of q and which is zero outside of U thus $fY \equiv 0$ on M and so since ∇^M is linear we have $\nabla^M(fY) \equiv 0$ on M. Thus since (3) holds for global fields we have

$$\begin{aligned}\nabla^M(fY)(q) &= f(p)(\nabla_X^M Y)(q) + (X_q f)Y_q \\ &= (\nabla_X^M Y)(q) = 0.\end{aligned}$$

Since $q \in U$ was arbitrary we have the result. ■

In the case of finite dimensional manifolds we have

Proposition 13.4 *Let M be a finite dimensional smooth manifold. Suppose that there exist an operator $\nabla^M : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ such that (1), (2) and (3) hold (for $U = M$). Then if we set $\nabla_X^U Y := r_U^M(\nabla_{\widetilde{X}}^M \widetilde{Y})$ for any extensions of X and $Y \in \mathfrak{X}(U)$ to global fields $\widetilde{X}$ and $\widetilde{Y} \in \mathfrak{X}(M)$ then $U \mapsto \nabla^U$ is a natural covariant derivative.*

Proof. By the previous lemma $\nabla_X^U Y := r_U^M(\nabla_{\widetilde{X}}^M \widetilde{Y})$ is a well defined operator which is easily checked to satisfy (1),(2) , (3) and (4) of definition 13.20. We now prove property (5). Let $\alpha \in T^*M$ and fix $Y \in \mathfrak{X}(U)$. define a map $\mathfrak{X}(U) \to C^\infty(U)$ by $X \mapsto \alpha(\nabla_X^U Y)$. By theorem 7.2 we see that $\alpha(\nabla_X^U Y)$ depend only on the value of X at $p \in U$. ■

Since many authors only consider finite dimensional manifolds they define a covariant derivative to be a map $\nabla^M : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ satisfying (1), (2) and (3). Later we will study connections and we show how these give rise to natural covariant derivatives.

It is common to write expressions like $\nabla_{\frac{\partial}{\partial x^i}} X$ where X is a global field and $\frac{\partial}{\partial x^i}$ is defined only on a coordinate domain U. This still makes sense as a field $p \mapsto \nabla_{\frac{\partial}{\partial x^i}(p)} X$ on U by virtue of (5) or by interpreting $\nabla_{\frac{\partial}{\partial x^i}} X$ as $\nabla_{\frac{\partial}{\partial x^i}} X|_U$ and invoking (4) if necessary. Let us agree to call a map $\nabla^M ::$ $\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ such that (1), (2) and (3) hold for $U = M$ a globally defined covariant derivative on $\mathfrak{X}(M)$. We now introduce the notion of a system of Christoffel symbols. We show below that if a globally defined covariant derivative ∇^M on $\mathfrak{X}(M)$ is induced by a system of Christoffel symbols then defining $\nabla_X^U Y := r_U^M(\nabla_{\widetilde{X}}^M \widetilde{Y})$ as in the finite dimensional case gives a natural covariant derivative.

Definition 13.22 *A **system of Christoffel symbols** on a smooth manifold M (modelled on M) is an assignment of a differentiable map*

$$\Gamma_\alpha : \psi_\alpha(U_\alpha) \to L(\mathsf{M}, \mathsf{M}; \mathsf{M})$$

to every admissible chart U_α, ψ_α such that if U_α, ψ_α and U_β, ψ_β are two such charts with $U_\alpha \cap U_\beta \neq \emptyset$, then for all $p \in U_\alpha \cap U_\beta$

$$\begin{aligned}&D(\psi_\beta \circ \psi_\alpha^{-1}) \cdot \Gamma_\alpha(x) \\ &= D^2(\psi_\beta \circ \psi_\alpha^{-1}) + \Gamma_\beta(y) \circ (D(\psi_\beta \circ \psi_\alpha^{-1}) \times D(\psi_\beta \circ \psi_\alpha^{-1}))\end{aligned}$$

where $y = \psi_\beta(p)$ and $x = \psi_\alpha(p)$. For finite dimensional manifolds with $\psi_\alpha = (x^1, ..., x^n)$ and $\psi_\beta = (y^1, ..., y^n)$ this last condition reads

$$\frac{\partial y^r}{\partial x^k}(x)\Gamma_{ij}^k(x) = \frac{\partial^2 y^r}{\partial x^i \partial x^j}(x) + \bar{\Gamma}_{pq}^r(y(x))\frac{\partial y^p}{\partial x^i}(x)\frac{\partial y^q}{\partial x^j}(x)$$

where $\Gamma_{ij}^k(x)$ are the components of $\Gamma_\alpha(x)$ and $\bar{\Gamma}_{pq}^r$ the components of $\Gamma_\beta(y(x)) = \Gamma_\beta(\psi_\beta \circ \psi_\alpha^{-1}(x))$ with respect to the standard basis of $L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^n)$.

Proposition 13.5 *Given a system of Christoffel symbols on a smooth manifold M there is a unique natural covariant derivative ∇ on M such that the principal part of $\nabla_X Y$ with respect to a chart U_α, ψ_α is given by $D\mathsf{Y}(\mathrm{x}).\mathsf{X}(\mathrm{x}) + \Gamma_\alpha(\mathrm{x})(\mathsf{X}(\mathrm{x}), \mathsf{Y}(\mathrm{x}))$ for $x \in \psi_\alpha(U_\alpha)$. Conversely, a natural covariant derivative determines a system of Christoffel symbols.*

Proof. Let a system of Christoffel symbols be given. Now for any open set $U \subset M$ we may let $\{U_a, \psi_a\}_a$ be any family of charts such that $\bigcup_a U_a = U$. Given vector fields $X, Y \in \mathfrak{X}(U)$ we define

$$s_{X,Y}(U_a) := \nabla^{U_a}_{r^U_{U_a} X} r^U_{U_a} Y$$

to have principal representation

$$\nabla^{U_a}_{\overset{\alpha}{\mathsf{X}}} \overset{\alpha}{\mathsf{Y}} = D\overset{\alpha}{\mathsf{Y}} \cdot \overset{\alpha}{\mathsf{X}} + \Gamma^\alpha(\overset{\alpha}{\mathsf{X}}, \overset{\alpha}{\mathsf{Y}}).$$

It is straight forward to check that the change of chart formula for Christoffel symbols implies that

$$r^{U_a}_{U_a \cap U_b} s_{X,Y}(U_a) = s_{X,Y}(U_a \cap U_b) = r^{U_b}_{U_a \cap U_b} s_{X,Y}(U_a)$$

and so by sheaf theoretic arguments there is a unique section

$$\nabla_X Y \in \mathfrak{X}(U)$$

such that

$$r^U_{U_a} \nabla_X Y = s_{X,Y}(U_a)$$

The verification that this defines a natural covariant derivative is now a straightforward (but tedious) verification of (1)-(5) in the definition of a natural covariant derivative.

For the converse, suppose that ∇ is a natural covariant derivative on M. Define the Christoffel symbol for a chart U_a, ψ_α to be in the following way. For fields

$$\mathsf{x} \to (\mathsf{x}, \overset{\alpha}{\mathsf{X}}(\mathsf{x}))$$

and

$$\mathsf{x} \to (\mathsf{x}, \overset{\alpha}{\mathsf{Y}}(\mathsf{x}))$$

one may define $\Theta(\overset{\alpha}{\mathsf{X}}, \overset{\alpha}{\mathsf{Y}}) := \nabla^{U_a}_{\overset{\alpha}{\mathsf{X}}} \overset{\alpha}{\mathsf{Y}} - D\overset{\alpha}{\mathsf{Y}} \cdot \overset{\alpha}{\mathsf{X}}$ and then use the properties (1)-(5) to show that $\Theta(\overset{\alpha}{\mathsf{X}}, \overset{\alpha}{\mathsf{Y}})(\mathsf{x})$ depends only on the values of $\overset{\alpha}{\mathsf{X}}$ and $\overset{\alpha}{\mathsf{Y}}$ at the point x. Thus there is a function $\Gamma : U_\alpha \to L(\mathsf{M}, \mathsf{M}; \mathsf{M})$ such that $\Theta(\overset{\alpha}{\mathsf{X}}, \overset{\alpha}{\mathsf{Y}})(\mathsf{x}) = \Gamma(\mathsf{x})(\overset{\alpha}{\mathsf{X}}(\mathsf{x}), \overset{\alpha}{\mathsf{Y}}(\mathsf{x}))$. We wish to show that this defines a system of Christoffel symbols. But this is just an application of the chain rule. ■

In finite dimensions and using traditional notation

$$\nabla_X Y = \left(\frac{\partial Y^k}{\partial x^j} X^j + \Gamma^k_{ij} X^i Y^j \right) \frac{\partial}{\partial x^k}.$$

where $X = X^j \frac{\partial}{\partial x^j}$ and $Y = Y^j \frac{\partial}{\partial x^j}$. In particular,

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \Gamma^k_{ij} \frac{\partial}{\partial x^k}.$$

Proof. The proof follows directly from the local definition and may be easily checked by the patient reader. One should check that the transformation law in the definition of a system of Christoffel symbols implies that $D\mathsf{Y}(\mathrm{x}).\mathsf{X}(\mathrm{x}) + \Gamma_\alpha(\mathrm{x})(\mathsf{X}(\mathrm{x}), \mathsf{Y}(\mathrm{x}))$ transforms as the principal local representative of a vector. ■

Remark 13.7 *We will eventually define an extension of the* ***covariant derivative*** *to tensor fields. Let us get a small head start on that by letting* $\nabla_X f := Xf$ *for* $f \in C^\infty(M)$. *We now have the following three expression for the same thing:*

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

Notice that with this definition (4) above reads more like a product rule:

$$\nabla_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle.$$

Definition 13.23 *Define the operator* $T_\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ *by*

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

T_∇ *is called the* ***torsion tensor*** *for the connection* ∇.

Theorem 13.4 *For a given Riemannian manifold M, g, there is a unique metric connection ∇ such that its torsion is zero; $T_\nabla \equiv 0$. This unique connection is called the* ***Levi-Civita derivative*** *for M, g.*

Proof. We will derive a formula that must be satisfied by ∇ which can in fact be used to define ∇. Let X, Y, Z, W be arbitrary vector fields on $U \subset M$. If ∇ exists as stated then on U we must have

$$\begin{aligned} X\langle Y, Z\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle \\ Y\langle Z, X\rangle &= \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle \\ Z\langle X, Y\rangle &= \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle. \end{aligned}$$

where we have written ∇^U simply as ∇. Now add the first two equations to the third one to get

$$\begin{aligned} &X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle \\ &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle + \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle \\ &\quad -\langle \nabla_Z X, Y\rangle - \langle X, \nabla_Z Y\rangle. \end{aligned}$$

Now if we assume the torsion zero hypothesis then this reduces to

$$\begin{aligned} &X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle \\ &= \langle Y, [X, Z]\rangle + \langle X, [Y, Z]\rangle \\ &\quad -\langle Z, [X, Y]\rangle + 2\langle \nabla_X Y, Z\rangle. \end{aligned}$$

Solving we see that $\nabla_X Y$ must satisfy

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle &= X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle \\ &\quad \langle Z, [X, Y]\rangle - \langle Y, [X, Z]\rangle - \langle X, [Y, Z]\rangle. \end{aligned}$$

Now since knowing $\langle \nabla_X Y, Z\rangle$ for all Z is tantamount to knowing $\nabla_X Y$ we conclude that if ∇ exists then it is unique. On the other hand, the patient reader can check that if we actually define $\langle \nabla_X Y, Z\rangle$ and hence $\nabla_X Y$ by this equation then all of the defining properties of a connection are satisfied and furthermore T_∇ will be zero. ∎

It is not difficult to check that we may define a system of Christoffel symbols for the Levi Civita derivative by the formula

$$\Gamma^\alpha(\mathsf{X}, \mathsf{Y}) := \nabla_{\mathsf{X}} \mathsf{Y} - D\mathsf{Y} \cdot \mathsf{X}$$

where X, Y and $\nabla_{\mathsf{X}} \mathsf{Y}$ are the principal representatives of X, Y and $\nabla_X Y$ respectively for a given chart U_α, ψ_α.

Proposition 13.6 *Let M be a semi-Riemannian manifold of dimension n and let $U, \psi = (x^1, ..., x^n)$ be a chart. Then we have the formula*

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl} \left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l} \right).$$

where $g_{jk} g^{ki} = \delta^i_j$.

13.6 Covariant differentiation of vector fields along maps.

Let $F : N \to M$ be a smooth map. A **vector field along** F is a map $Z : N \to TM$ such that the following diagram commutes:

$$\begin{array}{ccc} & & TM \\ Z & \nearrow & \downarrow^{\tau_M} \\ N & \xrightarrow{F} & M \end{array}$$

We denote the set of all smooth vector fields along a map F by $\mathfrak{X}_F$. Let $F : N \to M$ be a smooth map and let the model spaces of M and N be M and N respectively.

We shall say that a pair of charts Let U, ψ be a chart on M and let V, φ be a chart on N such that $\varphi(V) \subset U$. We shall say that such a pair of charts is adapted to the map F.

Assume that there exists a system of Christoffel symbols on M. We may define a covariant derivative $\nabla_X Z$ of a vector field Z along F with respect to a field $X \in \mathfrak{X}(N)$ by giving its principal representation with respect to any pair of charts adapted to F. Then $\nabla_X Z$ will itself be a vector fields along F. The map F has a local representation $F_{V,U} : V \to \psi(U)$ defined by $F_{V,U} := \psi \circ F \circ \varphi^{-1}$. Similarly the principal representation $\mathsf{Y}^* : \varphi(V) \to \mathsf{M}$ of Y is given by $T\psi \circ Z \circ \varphi^{-1}$ followed by projection onto the second factor of $\psi(U) \times \mathsf{M}$. Now given any vector field $X \in \mathfrak{X}(N)$ with principal representation $\mathsf{X} : \psi(U) \to \mathsf{N}$ we define the covariant derivative $\nabla_X Y^*$ of X with respect to Z as that vector field along F whose principal representation with respect to any arbitrary pair of charts adapted to F is

$$D\mathsf{Z}(\mathsf{x}) \cdot \mathsf{X}(\mathsf{x}) + \Gamma(F(\mathsf{x}))(DF_{V,U}(\mathsf{x}) \cdot \mathsf{X}(\mathsf{x}), \mathsf{Z}(\mathsf{x})).$$

The resulting map $\nabla : \mathfrak{X}(N) \times \mathfrak{X}_F \to \mathfrak{X}(N)$ has the following properties:

1. $\nabla : \mathfrak{X}(N) \times \mathfrak{X}_F \to \mathfrak{X}(N)$ is $C^\infty(N)$ linear in the first argument.

2. For the second argument we have

   $$\nabla_X(fZ) = f\nabla_X Z + X(f)Z$$

   for all $f \in C^\infty(N)$.

3. If Z happens to be of the form $Y \circ F$ for some $Y \in \mathfrak{X}(M)$ then we have

   $$\nabla_X(Y \circ F) = (\nabla_{TF \cdot X} Y) \circ F.$$

4. $(\nabla_X Z)(p)$ depends only on the value of X at $p \in N$ and we write $(\nabla_X Z)(p) = \nabla_{X_p} Z$.

For a curve $c :: \mathbb{R} \to M$ and $Z :: \mathbb{R} \to TM$ we define

$$\frac{\nabla Z}{dt} := \nabla_{d/dt} Z \in \mathfrak{X}_F$$

If Z happens to be of the form $Y \circ c$ then we have the following alternative notations with varying degrees of precision:

$$\nabla_{d/dt}(Y \circ c) = \nabla_{\dot{c}(t)} Y = \nabla_{d/dt} Y = \frac{\nabla Y}{dt}$$

13.7 Covariant differentiation of tensor fields

Let ∇ be a natural covariant derivative on M. It is a consequence of proposition 9.4 that for each $X \in \mathfrak{X}(U)$ there is a unique tensor derivation ∇_X on $\mathfrak{T}^r_s(U)$ such that ∇_X commutes with contraction and coincides with the given covariant derivative on $\mathfrak{X}(U)$ (also denoted ∇_X) and with $\mathcal{L}_X f$ on $C^\infty(U)$.

To describe the covariant derivative on tensors more explicitly consider $\Upsilon \in \mathfrak{T}^1_1$ with a 1-form Since we have the contraction $Y \otimes \Upsilon \mapsto C(Y \otimes \Upsilon) = \Upsilon(Y)$ we should have

$$\begin{aligned}
\nabla_X \Upsilon(Y) &= \nabla_X C(Y \otimes \Upsilon) \\
&= C(\nabla_X (Y \otimes \Upsilon)) \\
&= C(\nabla_X Y \otimes \Upsilon + Y \otimes \nabla_X \Upsilon) \\
&= \Upsilon(\nabla_X Y) + (\nabla_X \Upsilon)(Y)
\end{aligned}$$

and so we should define $(\nabla_X \Upsilon)(Y) := \nabla_X(\Upsilon(Y)) - \Upsilon(\nabla_X Y)$. If $\Upsilon \in \mathfrak{T}^1_s$ then

$$(\nabla_X \Upsilon)(Y_1, ..., Y_s) = \nabla_X(\Upsilon(Y_1, ..., Y_s)) - \sum_{i=1}^{s} \Upsilon(..., \nabla_X Y_i, ...)$$

Now if $\Upsilon \in \mathfrak{T}^0_s$ we apply this to $\nabla Z \in \mathfrak{T}^1_1$ and get

$$\begin{aligned}
(\nabla_X \nabla Z)(Y) &= X(\nabla Z(Y)) - \nabla Z(\nabla_X Y) \\
&= \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z
\end{aligned}$$

form which we get the following definition:

Definition 13.24 *The second covariant derivative of a vector field $Z \in \mathfrak{T}^0_s$ is*

$$\nabla^2 Z : (X, Y) \mapsto \nabla^2_{X,Y}(Z) = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z$$

Definition 13.25 *A tensor field Υ is said to be parallel if $\nabla_\xi \Upsilon = 0$ for all ξ. Similarly, if $\sigma : I \to T^r_s(M)$ is a tensor field along a curve $c : I \to M$ satisfies $\nabla_{\partial_t} \sigma = 0$ on I then we say that σ is parallel along c. Just as in the case of a general connection on a vector bundle we then have a parallel transport map $P(c)^t_{t_0} : T^r_s(M)_{c(t_0)} \to T^r_s(M)_{c(t)}$.*

Exercise 13.6 *Prove that*

$$\nabla_{\partial_t}\sigma(t) = \lim_{\epsilon\to 0}\frac{P(c)^t_{t+\epsilon}\sigma(t+\epsilon)-\sigma(t)}{\epsilon}.$$

Also, if $\Upsilon \in \mathfrak{T}^r_s$ *then if* c^X *is the curve* $t \mapsto Fl^X_t(p)$

$$\nabla_X\Upsilon(p) = \lim_{\epsilon\to 0}\frac{P(c^X)^t_{t+\epsilon}(\Upsilon\circ Fl^X_t(p)) - Y\circ Fl^X_t(p)}{\epsilon}.$$

The map $\nabla_X : \mathfrak{T}^r_s M \to \mathfrak{T}^r_s M$ just defined commutes with contraction. This means for instance that

$$\nabla_i(\left.\Upsilon^{jk}\right|_k) = \nabla_i \left.\Upsilon^{jk}\right|_k.$$

Furthermore, if the connection we are extending is the Levi Civita connection for semi-Riemannian manifold M, g then

$$\nabla_\xi \mathrm{g} = 0 \text{ for all } \xi$$

To see this recall that

$$\nabla_\xi(\mathrm{g}\otimes Y\otimes W) = \nabla_\xi \mathrm{g}\otimes X\otimes Y + \mathrm{g}\otimes\nabla_\xi X\otimes Y + \mathrm{g}\otimes X\otimes\nabla_\xi Y$$

which upon contraction yields

$$\begin{aligned}\nabla_\xi(\mathrm{g}(X,Y)) &= (\nabla_\xi \mathrm{g})(X,Y) + \mathrm{g}(\nabla_\xi X, Y) + \mathrm{g}(X,\nabla_\xi Y)\\ \xi\langle X,Y\rangle &= (\nabla_\xi \mathrm{g})(X,Y) + \langle\nabla_\xi X,Y\rangle + \langle X,\nabla_\xi Y\rangle.\end{aligned}$$

We see that $\nabla_\xi \mathrm{g} \equiv 0$ for all ξ if and only if $\langle X,Y\rangle = \langle\nabla_\xi X,Y\rangle + \langle X,\nabla_\xi Y\rangle$ for all ξ, X, Y. In other words the statement that the metric tensor is parallel (constant) with respect to ∇ is the same as saying that the connection is a metric connection. Since we are assuming the connection has torsion zero metric connection means we have a Levi-Civita connection. On the other hand, for a fixed $T \in \mathfrak{T}^r_s(M)$ the map $X \mapsto \nabla_X T$ is $C^\infty(M)-$ linear and so we may view ∇ as being a map from $\mathfrak{T}^r_s(M)$ to $\mathfrak{T}^r_s(M)_{C^\infty}\otimes\Omega^1(M) \cong \Gamma(\mathrm{Hom}(TM, L^1_{alt}M))$. We call $\nabla T \in \mathfrak{T}^r_s(M)_{C^\infty}\otimes\Omega^1(M)$ the **covariant differential** of T.

13.8 Comparing the Differential Operators

On a smooth manifold we have the Lie derivative $\mathcal{L}_X : \mathfrak{T}^r_s(M) \to \mathfrak{T}^r_s(M)$ and the exterior derivative $d : \Omega^k(M) \to \Omega^{k+1}(M)$ and in case we have a torsion free covariant derivative ∇ then that make three differential operators which we would like to compare. To this end we restrict attention to purely covariant tensor fields $\mathfrak{T}^0_s(M)$.

The extended map $\nabla_\xi : \mathfrak{T}^0_s(M) \to \mathfrak{T}^0_s(M)$ respects the subspace consisting of alternating tensors and so we have a map

$$\nabla_\xi : L^k_{alt}(M) \to L^k_{alt}(M)$$

which combine to give a degree preserving map

$$\nabla_\xi : L_{alt}(M) \to L_{alt}(M)$$

or in other notation

$$\nabla_\xi : \Omega(M) \to \Omega(M).$$

It is also easily seen that not only do we have $\nabla_\xi(\alpha \otimes \beta) = \nabla_\xi \alpha \otimes \beta + \alpha \otimes \nabla_\xi \beta$ but also

$$\nabla_\xi(\alpha \wedge \beta) = \nabla_\xi \alpha \wedge \beta + \alpha \wedge \nabla_\xi \beta$$

Now as soon as one realizes that $\nabla\omega \in \Omega^k(M)_{C^\infty} \otimes \Omega^1(M)$ instead of $\Omega^{k+1}(M)$ we search for a way to fix things. By antisymetrizing we get a map $\Omega^k(M) \to \Omega^{k+1}(M)$ which turns out to be none other than our old friend the exterior derivative as will be shown below.

Now recall that $X(\omega(Y)) = (\nabla_X \omega)(Y) + \omega(\nabla_X Y)$ for $\omega \in \Omega^1(M)$ and X and Y vector fields. More generally, for $T \in \mathfrak{T}^0_s(M)$ we have

$$(\nabla_X S)(Y_1, ..., Y_s) = X(S(Y_1, ..., Y_s)) - \sum_{i=1}^{s} S(Y_1, ..., Y_{i-1}, \nabla_X Y_i, Y_{i+1}, ..., Y_s) \tag{13.6}$$

and a similar formula for the Lie derivative:

$$(\mathcal{L}_X S)(Y_1, ..., Y_s) = X(S(Y_1, ..., Y_s)) - \sum_{i=1}^{s} S(Y_1, ..., Y_{i-1}, \mathcal{L}_X Y_i, Y_{i+1}, ..., Y_s). \tag{13.7}$$

On the other hand, ∇ is torsion free and so $\mathcal{L}_X Y_i = [X, Y_i] = \nabla_X Y_i - \nabla_{Y_i} X$ and so we obtain

$$(\mathcal{L}_X S)(Y_1, ..., Y_s) = (\nabla_X S)(Y_1, ..., Y_s) + \sum_{i=1}^{s} S(Y_1, ..., Y_{i-1}, \nabla_{Y_i} X, Y_{i+1}, ..., Y_s). \tag{13.8}$$

When ∇ is the Levi-Civita connection for the Riemannian manifold M, g we get the interesting formula

$$(\mathcal{L}_X \mathrm{g})(Y, Z) = \mathrm{g}(\nabla_X Y, Z) + \mathrm{g}(Y, \nabla_X Z) \tag{13.9}$$

for vector fields $X, Y, Z \in \mathfrak{X}(M)$.

Theorem 13.5 *If ∇ is any natural, torsion free covariant derivative on M then we have*

$$d\omega(X_0, X_1, ..., X_k) = \sum_{i=0}^{k} (-1)^i (\nabla_{X_i} \omega)(X_0, ..., \widehat{X_i}, ..., X_k)$$

Proof. Recall the formula

$$
\begin{aligned}
d\omega(X_0, X_1, ..., X_k) &= \sum_{i=0}^{k}(-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\leq r<s\leq k}(-1)^{r+s}\omega([X_r, X_s], X_0, ..., \widehat{X_r}, ..., \widehat{X_s}, ..., X_k) \\
&= \sum_{i=0}^{k}(-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\leq r<s\leq k}(-1)^{r+s}\omega(\nabla_{X_r}X_s - \nabla_{X_r}X_s, X_0, ..., \widehat{X_r}, ..., \widehat{X_s}, ..., X_k) \\
&\sum_{i=0}^{k}(-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\leq r<s\leq k}(-1)^{r+1}\omega(X_0, ..., \widehat{X_r}, ..., \nabla_{X_r}X_s, ..., X_k) \\
&- \sum_{1\leq r<s\leq k}(-1)^{s}\omega(X_0, ..., \nabla_{X_s}X_r, ..., \widehat{X_s}, ..., X_k) \\
&= \sum_{i=0}^{k}(-1)^i \nabla_{X_i}\omega(X_0, ..., \widehat{X_i}, ..., X_k) \text{ (by using 13.6)}
\end{aligned}
$$

■

Chapter 14

Formalisms for Calculation

14.1 Tensor Calculus

When working with tensors there is no need to feel obliged to use holonomic frames (those coming from coordinate systems: $\frac{\partial}{\partial x^i}, dx^i$). There are always a great variety of frame fields sometimes defined on sets which are larger than possible is possible for coordinate frames. Of course within a coordinate chart any other frame field $(e_1, ..., e_n)$ and dual frame $(e^1, ..., e^n)$ may be written in terms of the coordinate frames as $e_i = a^i \frac{\partial}{\partial x^i}$ and $e^i = b_i dx^i$.

From the point of view of arbitrary frame the tensor calculus is much that same as it is for coordinate frames. One writes any tensor as

$$\begin{aligned} T &= T^{i_1 \ldots i_k}{}_{j_1 \ldots j_l}\, e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e^{j_1} \otimes \cdots \otimes e^{j_l} \\ &= T^I_J\, e_I \otimes e^J. \end{aligned}$$

If two frame fields are related on the overlap of their domains by $f_i = e_j g^j_i$ and $f^i = (g^{-1})^i_j e^j$ then with $g^I_R := g^{i_1 \ldots i_k}_{r_1 \ldots r_k} := g^{i_1}_{r_1} g^{i_2}_{r_2} \cdots g^{i_k}_{r_k}$ and $(g^{-1})^I_R := (g^{-1})^{i_1 \ldots i_k}_{r_1 \ldots r_k} := (g^{-1})^{i_1}_{r_1} \cdots (g^{-1})^{i_l}_{r_l}$ we have

$$\begin{aligned} T'^I_J\, f_I \otimes f^J &= T^R_S\, e_R \otimes e^S \\ T'^I_J\, e_R g^R_I \otimes (g^{-1})^J_S e^S &= T^R_S\, e_R \otimes e^S \\ (T'^I_J\, g^R_I (g^{-1})^J_S) e_R \otimes e^S &= T^R_S\, e_R \otimes e^S \end{aligned}$$

and so the transformation law for the tensor components is $T^R_S = g^R_I T'^I_J (g^{-1})^J_S$ or

$$(g^{-1})^I_R T^R_S (g)^S_J = T'^I_J$$

$(Bv) = v^j \mapsto b^i{}_j\, v^j$
$(Bv, w) = (b^i{}_j\, v^j) g_{is} w^s$
$(v, B^t w) = v^i g_{ik} (b(t)^k_j w^j)$
$(b^i{}_j\, v^j) g_{is} w^s = g_{jk} v^j (b(t)^k_s w^s)$

$b^i{}_j\, g_{is} = g_{jk} b(t)^k_s$
$g^{ek}\, b^i{}_j\, g_{is} = g^{ej} g_{jk} b(t)^k_s$
$g^{ej}\, b^i{}_j\, g_{is} = \delta^e_k b(t)^k_s$
$b_i{}^j = b(t)^j_i$
$v_i \mapsto b_i{}^j\, v_j$

14.2 Covariant Exterior Calculus, Bundle-Valued Forms

A moving frame of an open subset U of an n-dimensional $C^\infty-$ manifold is an n-tuple of vector fields $X_1, ..., X_n \in \mathfrak{X}(U)$ such that $X_1(p), ..., X_n(p)$ is a basis for T_pM for every $p \in M$. It is easy to see that there is a dual moving frame which is an $n-$tuple of $1-$forms $\theta^1, ..., \theta^n \in \mathfrak{X}^*(U)$ so that

$$\theta^i(X_j) = \delta^i_j.$$

Now we can clearly find an open cover $\{U_\alpha\}$ of M with set up like above for each:

$$\begin{gathered} U_\alpha \\ X_{\alpha 1}, ..., X_{\alpha n} \in \mathfrak{X}(U_\alpha) \\ \theta^1_\alpha, ..., \theta^n_\alpha \in \mathfrak{X}^*(U_\alpha) \end{gathered}$$

On possibility is to choose an atlas $\{U_\alpha, \psi_\alpha\}$ for M and let $X_{\alpha 1}, ..., X_{\alpha n}$ be the coordinate vector fields $\{\frac{\partial}{\partial x^1_\alpha}, ..., \frac{\partial}{\partial x^n_\alpha}\}$ so that also $\{\theta^1_\alpha, ..., \theta^n_\alpha\} = \{dx^1_\alpha, ..., dx^n_\alpha\}$ but this is not the only choice and that is where the real power comes from since we can often choose the frames in a way that fits the geometric situation better than coordinate fields could. For example, on a Riemannian manifold we might choose the frames to be orthonormal at each point. This is something that we cannot expect to happen with coordinate vector fields unless the curvature is zero.

Returning to a general C^∞-manifold, let us consider the TM-valued

Chapter 15

Topology

When science finally locates the center of the universe, some people will be surprised to learn they're not it.

-Anonymous

15.1 Attaching Spaces and Quotient Topology

Suppose that we have a topological space X and a surjective set map $f : X \to S$ onto some set S. We may endow S with a natural topology according to the following recipe. A subset $U \subset S$ is defined to be open if and only if $f^{-1}(U)$ is an open subset of X. This is particularly useful when we have some equivalence relation on X which allows us to consider the set of equivalence classes $X/\sim$. In this case we have the canonical map $\varrho : X \to X/\sim$ which takes $x \in X$ to its equivalence class $[x]$. The quotient topology is then given as before by the requirement that $U \subset S$ is open iff and only if $\varrho^{-1}(U)$ is open in X. A common application of this idea is the identification of a subspace to a point. Here we have some subspace $A \subset X$ and the equivalence relation is given by the following two requirements:

$$\begin{array}{rl} \text{If } x \in X\backslash A & \text{then } x \sim y \text{ only if } x = y \\ \text{If } x \in A & \text{then } x \sim y \text{ for any } y \in A \end{array}$$

In other words, every element of A is identified with every other element of A. We often denote this space by X/A.

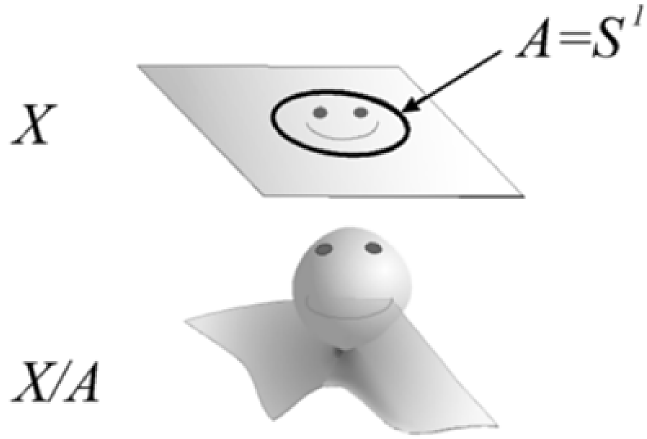
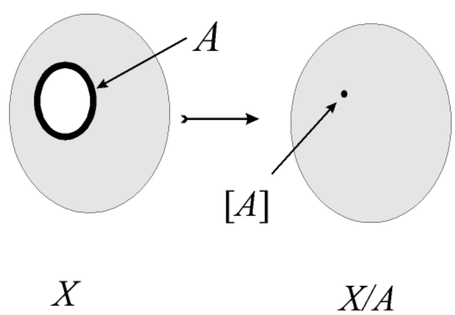


Figure 15.1: creation of a "hole"

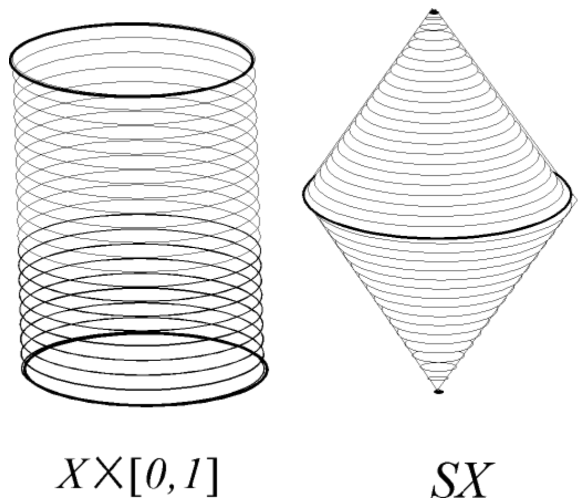


A hole is removed by identification

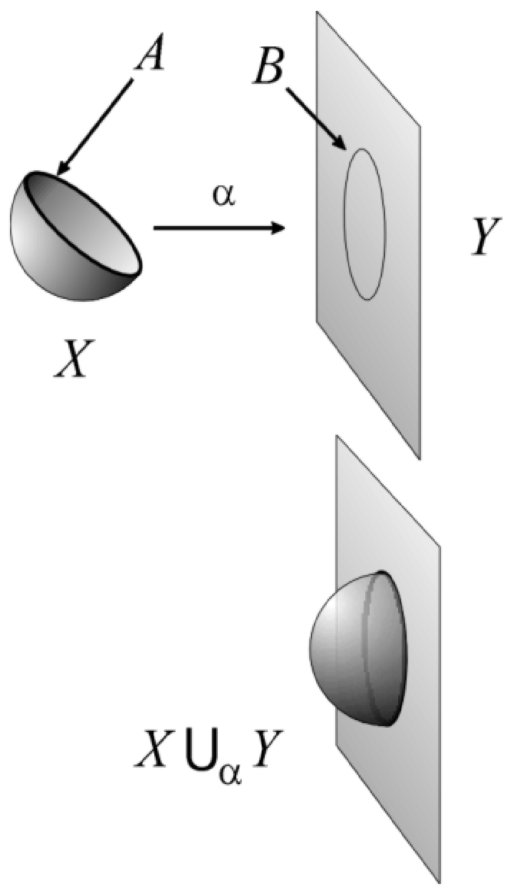
It is not difficult to verify that if X is Hausdorff (resp. normal) and A is closed then X/A is Hausdorff (resp. normal). The identification of a subset to a point need not simplify the topology but may also complicate the topology as shown in the figure.

An important example of this construction is the suspension. If X is a topological space then we define its suspension SX to be $(X \times [0,1])/A$ where $A := (X \times \{0\}) \cup (X \times \{1\})$. For example it is easy to see that $SS^1 \cong S^2$. More generally, $SS^{n-1} \cong S^n$.

Consider two topological spaces X and Y and subset $A \subset X$ a closed subset. Suppose that we have a map $\alpha : A \to B \subset Y$. Using this map we may define an equivalence relation on the disjoint union $X \bigsqcup Y$ which is given by requiring that $x \sim \alpha(x)$ for $x \in A$. The resulting topological space is denoted $X \cup_\alpha Y$.



X×[0,1]
SX



A
B
α
Y
X
X ∪α Y

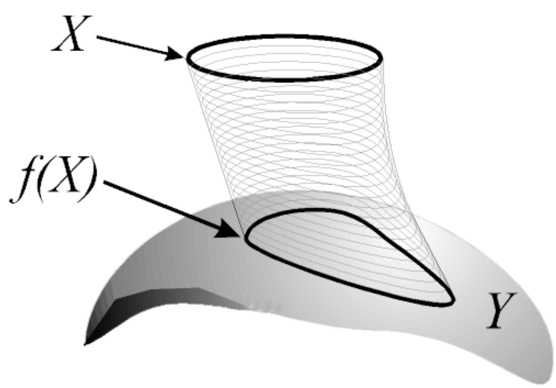


Figure 15.2: Mapping Cylinder

Attaching a 2-cell

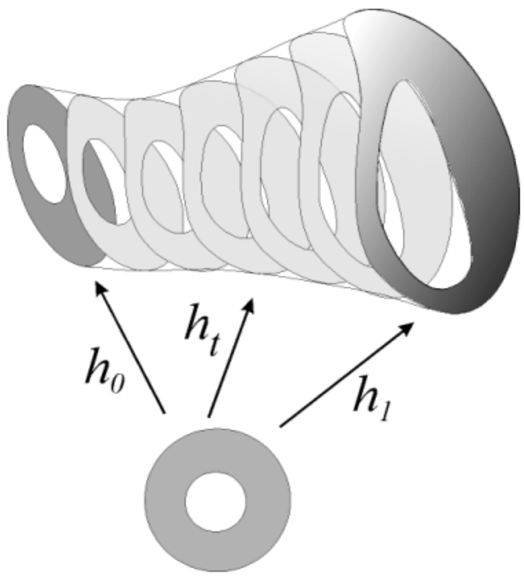
Another useful construction is that of a the mapping cylinder of a map $f : X \to Y$. First we transfer the map to a map on the base $X \times \{0\}$ of the cylinder $X \times I$ by

$$f(x, 0) := f(x)$$

and then we form the quotient $Y \cup_f (X \times I)$. We denote this quotient by M_f and call it the mapping cylinder of f.

15.2 Topological Sum

15.3 Homotopy



Homotopy as a family of maps.

Definition 15.1 *Let $f_0, f_1 : X \to Y$ be maps. A homotopy from f_0 to f_1 is a one parameter family of maps $\{h_t : X \to Y : 0 \leq t \leq 1\}$ such that $h_0 = f_0$, $h_1 = f_1$ and such that $(x, t) \mapsto h_t(x)$ defines a (jointly continuous) map $X \times [0, 1] \to Y$. If there exists such a homotopy we write $f_0 \simeq f_1$ and say that f_0 is homotopic to f_1. If there is a subspace $A \subset X$ such that $h_t| A = f_0| A$ for all $t \in [0, 1]$ then we say that f_0 is homotopic to f_1 relative to A and we write $f_0 \simeq f_1(rel\ A)$.*

It is easy to see that homotopy equivalence is in fact an equivalence relation. The set of homotopy equivalence classes of maps $X \to Y$ is denoted $[X, Y]$ or $\pi(X, Y)$.

Definition 15.2 *Let $f_0, f_1 : (X, A) \to (Y\ , B)$ be maps of topological pairs. A homotopy from f_0 to f_1 is a homotopy h of the underlying maps $f_0, f_1 : X \to Y$ such that $h_t(A) \subset B$ for all $t \in [0, 1]$. If $S \subset X$ then we say that f_0 is homotopic to f_1 relative to S if $h_t| S = f_0| S$ for all $t \in [0, 1]$.*

The set of homotopy equivalence classes of maps $(X, A) \to (Y\ , B)$ is denoted $[(X, A), (Y, B)]$ or $\pi((X, A), (Y, B))$. As a special case we have the notion of a homotopy of **pointed maps** $f_0, f_1 : (X, x_0) \to (Y, y_0)$. The points x_0 and y_0 are called the **base points** and are commonly denoted by the generic symbol $*$. The

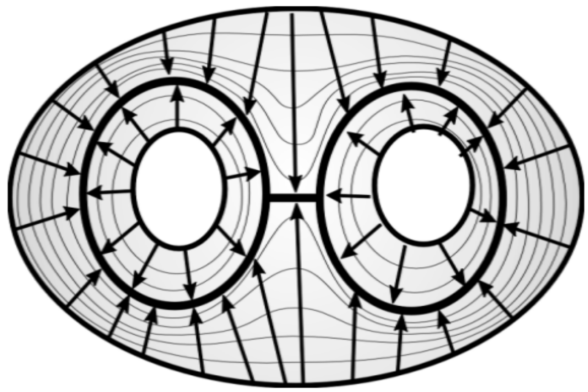


Figure 15.3: Retraction onto “eyeglasses”

set of all homotopy classes of pointed maps between pointed topological spaced is denoted $[(X, x_0), (Y, y_0)]$ or $\pi((X, x_0), (Y, y_0))$ but if the base points are fixed and understood then we denote the space of pointed homotopy classes as $[X, Y]_0$ or $\pi(X, Y)_0$. We may also wish to consider morphisms of **pointed pairs** such as $f : (X, A, a_0) \to (Y, B, b_0)$ which is given by a map $f : (X, A) \to (Y\ , B)$ such that $f(a_0) = b_0$. Here usually have $a_0 \in A$ and $b_0 \in B$. A homotopy between two such morphisms, say f_0 and $f_1 : (X, A, a_0) \to (Y, B, b_0)$ is a homotopy h of the underlying maps $(X, A) \to (Y\ , B)$ such that $h_t(a_0) = b_0$ for all $t \in [0, 1]$. Clearly there are many variations on this theme of restricted homotopy.

Remark 15.1 *Notice that if $f_0, f_1 : (X, A) \to (Y\ , y_0)$ are homotopic as maps of topological pairs then we automatically have $f_0 \simeq f_1(rel\ A)$. However, this is not necessarily the case if $\{y_0\}$ is replaced by a set $B \subset Y$ with more than one element.*

Definition 15.3 *A (strong)* ***deformation retraction*** *of X onto subspace $A \subset X$ is a homotopy f_t from $f_0 = \mathrm{id}_X$ to f_1 such that $f_1(X) \subset A$ and $f_t|\, A = \mathrm{id}_A$ for all $t \in [0, 1]$. If such a retraction exists then we say that A is a (strong)* ***deformation retract*** *of X.*

Example 15.1 *Let $f_t : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}^n \backslash \{0\}$ be defined by*

$$f_t(x) := t\frac{x}{|x|} + (1 - t)x$$

for $0 \le t \le 1$. Then f_t gives a deformation retraction of $\mathbb{R}^n \backslash \{0\}$ onto $S^{n-1} \subset \mathbb{R}^n$.

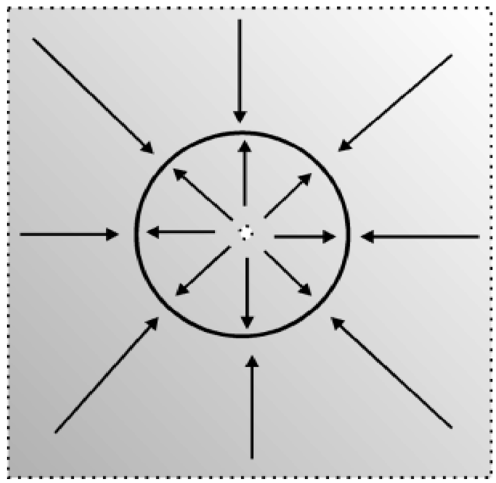


Figure 15.4: Retraction of punctured plane onto S^1

Definition 15.4 *A map $f : X \to Y$ is called a* ***homotopy equivalence*** *if there is a map $g : Y \to X$ such that $f \circ g \simeq \mathrm{id}_Y$ and $g \circ f \simeq \mathrm{id}_X$. The maps are then said to be homotopy inverses of each other. In this case we say that X and Y are* ***homotopy equivalent*** *and are said to be of the same* ***homotopy type****. We denote this relationship by $X \simeq Y$*

Definition 15.5 *A space X is called* ***contractible*** *if it is homotopy equivalent to a one point space.*

Definition 15.6 *A map $f : X \to Y$ is called* ***null-homotopic*** *if it is homotopic to a constant map.*

Equivalently, one can show that X is contractible iff every map $f : X \to Y$ is null-homotopic.

15.4 Cell Complexes

Let I denote the closed unit interval and let $I^n := I \times \cdots \times I$ be the n-fold Cartesian product of I with itself. The boundary of I is $\partial I = \{0, 1\}$ and the boundary of I^2 is $\partial I^2 = (I \times \{0, 1\}) \cup (\{0, 1\} \times I)$. More generally, the boundary of I^n is the union of the sets of the form $I \times \cdots \times \partial I \cdots \times I$. Also, recall that the closed unit n-disk D^n is the subset of $\mathbb{R}^n$ given by $\{|x|^2 \leq 1\}$ and has as boundary the sphere S^{n-1}. From the topological point of view the pair $(I^n, \partial I^n)$ is indistinguishable from the pair (D^n, S^{n-1}). In other words , $(I^n, \partial I^n)$ is homeomorphic to (D^n, S^{n-1}).

There is a generic notation for any homeomorphic copy of $I^n \cong D^n$ which is simply $\overline{e}^n$. Any such homeomorph of D^n is referred to as a closed n-cell. If we wish to distinguish several copies of such a space we might add an index to the notation as in $\overline{e}_1^n, \overline{e}_2^n$...etc. The interior of $\overline{e}^n$ is called an open n-cell and is generically denoted by e^n. The boundary is denoted by $\partial\overline{e}^n$ (or just ∂e^n). Thus we always have $(\overline{e}^n, \partial\overline{e}^n) \cong (D^n, S^{n-1})$.

An important use of the attaching idea is the construction of so called cell complexes . The open unit ball in $\mathbb{R}^n$ or any space homeomorphic to it is referred to as an open n-cell and is denoted by e^n. The closed ball is called a closed n-cell and has as boundary the $n-1$ sphere. A 0-cell is just a point and a 1-cell is a (homeomorph of) the unit interval the boundary of which is a pair of points. We now describe a process by which one can construct a large and interesting class of topological spaces called cell complexes. The steps are as follows:

1. Start with any discrete set of points and regard these as 0-cells.

2. Assume that one has completed the $n-1$ step in the construction with a resulting space X^{n-1}, construct X^n by attaching some number of copies of n-cells $\{e_\alpha^n\}_{\alpha \in A}$ (indexed by some set A) by attaching maps $f_\alpha : \partial e_\alpha^n = S^{n-1} \to X^{n-1}$.

3. Stop the process with a resulting space X^n called a **finite cell complex** or continue indefinitely according to some recipe and let $X = \bigcup_{n \geq 0} X^n$ and define a topology on X as follows: A set $U \subset X$ is defined to be open iff $U \cap X^n$ is open in X^n (with the relative topology). The space X is called a **CW-complex** or just a **cell complex** .

Definition 15.7 *Given a cell complex constructed as above the set X^n constructed at the n-th step is called the n-**skeleton**. If the cell complex is finite then the highest step n reached in the construction is the whole space and the cell complex is said to have dimension n. In other words, a finite cell complex has **dimension** n if it is equal to its own n-skeleton.*

It is important to realize that the stratification of the resulting topological space by the via the skeletons and also the open cells that are homeomorphically embedded are part of the definition of a cell complex and so two different cell complexes may in fact be homeomorphic without being the same cell complex. For example, one may realize the circle S^1 by attaching a 1-cell to a 0-cell or by attaching two 1-cells to two different 0-cells as in figure 15.5.

Another important example is the projective space $P^n(\mathbb{R})$ which can be thought of as a the hemisphere $\overline{S_+^n} = \{x \in \mathbb{R}^{n+1} : x^{n+1} \geq 0\}$ which antipodal points of the boundary $\partial\overline{S_+^n} = S^{n-1}$ identified. But S^{n-1} with antipodal point identified is just $P^{n-1}(\mathbb{R})$ and so we can obtain $P^n(\mathbb{R})$ by attaching an n-cell e^n to $P^{n-1}(\mathbb{R})$ via the attaching map $\partial e^n = S^{n-1} \to P^{n-1}(\mathbb{R})$ which is just the quotient map of S^{n-1} onto $P^{n-1}(\mathbb{R})$. By repeating this analysis inductively we

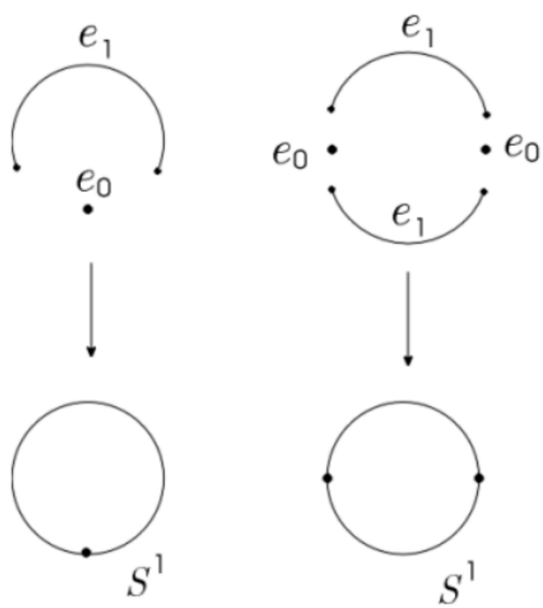


Figure 15.5: Two different cell structures for S^1.

conclude that $P^n(\mathbb{R})$ can be obtained from a point by attaching one cell from each dimension up to n :

$$P^n(\mathbb{R}) = e^0 \cup e^2 \cup \cdots \cup e^n$$

and so $P^n(\mathbb{R})$ is a finite cell complex of dimension n.

Chapter 16

Algebraic Topology

16.1 Axioms for a Homology Theory

Consider the category $\mathcal{TP}$ of all topological pairs (X, A) where X is a topological space, A is a subspace of X and where a morphism $f : (X, A) \to (X', A')$ is given by a map $f : X \to X'$ such that $f(A) \subset A'$. We may consider the category of topological spaces and maps as a subcategory of $\mathcal{TP}$ by identifying $(X, \emptyset)$ with X. We will be interested in functors from some subcategory $\mathcal{NTP}$ to the category $\mathbb{Z}-\mathcal{GAG}$ of $\mathbb{Z}$-graded abelian groups. The subcategory $\mathcal{NTP}$ (tentatively called "nice topological pairs") will vary depending of the situation but one example for which things work out nicely is the category of finite cell complex pairs. Let $\sum A_k$ and $\sum B_k$ be graded abelian groups. A morphism of $\mathbb{Z}$-graded abelian groups is a sequence $\{h_k\}$ of group homomorphisms $h_k : A_k \to B_k$. Such a morphism may also be thought of as combined to give a degree preserving map on the graded group; $h : \sum A_k \to \sum B_k$.

A homology theory H with coefficient group G is a covariant functor h_G from a category of nice topological pairs $\mathcal{NTP}$ to the category $\mathbb{Z}-\mathcal{GAG}$ of $\mathbb{Z}$-graded abelian groups:

$$h_G : \begin{cases} (X, A) \mapsto H(X, A, G) = \sum_{p \in \mathbb{Z}} H_p(X, A, G) \\ \qquad\qquad f \mapsto f_* \end{cases}$$

and which satisfies the following axioms (where we write $H_p(X, \emptyset) = H_p(X)$ etc.):

1. $H_p(X, A) = 0$ for $p < 0$.

2. (**Dimension axiom**) $H_p(pt) = 0$ for all $p \geq 1$ and $H_0(pt) = G$.

3. If $f : (X, A) \to (X', A')$ is homotopic to $g : (X, A) \to (X', A')$ then $f_* = g_*$

4. (**Boundary map axiom**) To each pair (X, A) and each $p \in \mathbb{Z}$ there is a boundary homomorphism $\partial_p : H_p(X, A; G) \to H_{p-1}(A; G)$ such that for all maps $f : (X, A) \to (X', A')$ the following diagram commutes:

$$\begin{array}{ccc} H_p(X, A; G) & \xrightarrow{f_*} & H_p(X', A'; G) \\ \partial_p \downarrow & & \partial_p \downarrow \\ H_{p-1}(A; G) & \xrightarrow[(f|A)_*]{} & H_{p-1}(A'; G) \end{array}$$

5. (**Excision axiom**) For each inclusion $\iota : (B, B \cap A) \to (A \cup B, A)$ the induced map $\iota_* : H(B, B \cap A; G) \to H(A \cup B, A; G)$ is an isomorphism.

6. For each pair (X, A) and inclusions $i : A \hookrightarrow X$ and $j : (X, \emptyset) \hookrightarrow (X, A)$ there is a long exact sequence

$$\begin{array}{ccccc} \cdots \to H_{p+1}(A) \xrightarrow{i_*} H_{p+1}(X) & \xrightarrow{j_*} & H_{p+1}(X, A) & & \\ & & \partial_{p+1} \swarrow & & \\ & & H_{p+1}(A) \xrightarrow{i_*} & H_{p+1}(X) \xrightarrow{j_*} \cdots & \end{array}$$

where we have suppressed the reference to G for brevity.

16.2 Simplicial Homology

Simplicial homology is a perhaps the easiest to understand in principle.

And we have

16.3 Singular Homology

The most often studied homology theory these days is singular homology.

16.4 Cellular Homology

16.5 Universal Coefficient theorem

16.6 Axioms for a Cohomology Theory

16.7 De Rham Cohomology

16.8 Topology of Vector Bundles

In this section we study vector bundles with finite rank. Thus, the typical fiber may be taken to be $\mathbb{R}^n$ (or $\mathbb{C}^n$ for a complex vector bundle) for some positive integer n. We would also like to study vectors bundles over spaces that are not

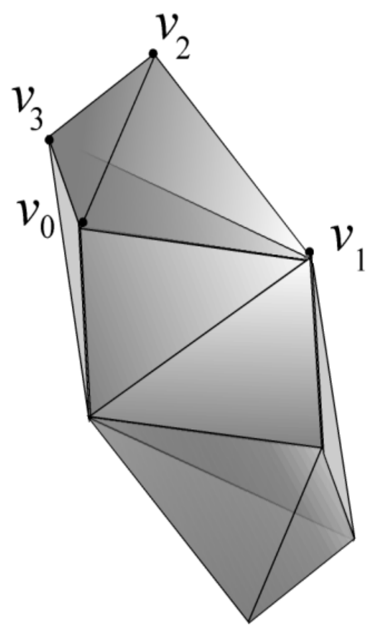


Figure 16.1: Simplicial complex

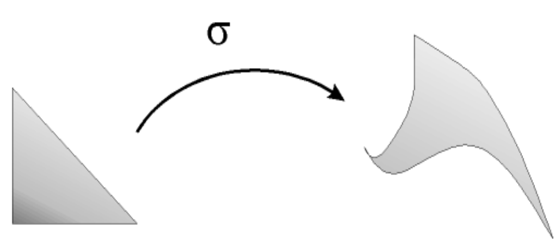


Figure 16.2: Singular 2-simplex

necessarily differentiable manifolds; although this will be our main interest. All the spaces in this section will be assumed to be paracompact Hausdorff spaces. We shall refer to continuous maps simply as maps. In many cases the theorems will makes sense in the differentiable category and in this case one reads map as "smooth map".

Recall that a (rank n) real vector bundle is a triple (π_E, E, M) where E and M are paracompact spaces and $\pi_E : E \to M$ is a surjective map such that there is a cover of M by open sets U_α together with corresponding trivializing maps (VB-charts) $\phi_\alpha : \pi_E^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^n$ of the form $\phi_\alpha = (\pi_E, \Phi_\alpha)$. Here $\Phi_\alpha : \pi_E^{-1}(U_\alpha) \to \mathbb{R}^n$ has the property that $\Phi_\alpha|_{E_x} : E_x \to \mathbb{R}^n$ is a diffeomorphism for each fiber $E_x := \pi_E^{-1}(x)$. Furthermore, in order that we may consistently transfer the linear structure of $\mathbb{R}^n$ over to E_x we must require that when $U_\alpha \cap U_\beta \neq \emptyset$ and $x \in U_\alpha \cap U_\beta$ then function

$$\Phi_{\beta\alpha;x} = \Phi_\beta|_{E_x} \circ \Phi_\alpha|_{E_x}^{-1} : \mathbb{R}^n \to \mathbb{R}^n$$

is a linear isomorphism. Thus the fibers are vectors spaces isomorphic to $\mathbb{R}^n$. For each nonempty overlap $U_\alpha \cap U_\beta$ we have a map $U_\alpha \cap U_\beta \to GL(n)$

$$x \mapsto \Phi_{\beta\alpha;x}.$$

We have already seen several examples of vector bundles but let us add one more to the list:

Example 16.1 *The normal bundle to $S^n \subset \mathbb{R}^{n+1}$ is the subset $N(S^n)$ of $S^n \times \mathbb{R}^{n+1}$ given by*

$$N(S^n) := \{(x, v) : x \cdot v = 0\}.$$

The bundle projection $\pi_{N(S^n)}$ is given by $(x, v) \mapsto x$. We may define bundle charts by taking opens sets $U_\alpha \subset S^n$ which cover S^n and then since any $(x, v) \in \pi_{N(S^n)}^{-1}(U_\alpha)$ is of the form (x, tx) for some $t \in \mathbb{R}$ we may define

$$\phi_\alpha : (x, v) = (x, tx) \mapsto (x, t).$$

Now there is a very important point to be made from the last example. Namely, it seems we could have taken *any* cover $\{U_\alpha\}$ with which to build the VB-charts. But can we just take the cover consisting of the single open set $U_1 := S^n$ and thus get a VB-chart $N(S^n) \to S^n \times \mathbb{R}$? The answer is that in this case we can. This is because $N(S^n)$ is itself a trivial bundle; that is, it is isomorphic to the product bundle $S^n \times \mathbb{R}$. This is not the case for vector bundles in general. In particular, we will later be able to show that the tangent bundle of an even dimensional sphere is always nontrivial. Of course, we have already seen that the Möbius line bundle is nontrivial.

16.9 de Rham Cohomology

In this section we assume that all manifolds are finite dimensional, Hausdorff, second-countable and C^∞. We will define the de Rham cohomology of a smooth manifold which will, of course, be a topological invariant. However, the definition involves the calculus of differential forms and hence uses the differentiable structure of the manifold.

Definition 16.1 *A differential form $\alpha \in \Omega^k(M)$ is called **closed** if $d\alpha = 0$ and is called **exact** if there exists a form β such that $\alpha = d\beta$.*

It is east to check that a linear combination of closed forms is closed and that every exact form is closed. Thus if $Z^k(M)$ denotes the set of all closed forms and $B^k(M)$ the set of all exact forms then

$$B^k(M) = \mathrm{img}(d : \Omega^{k-1}(M) \to \Omega^k(M)),$$
$$Z^k(M) = \ker(d : \Omega^k(M) \to \Omega^{k+1}(M))$$

and are real vector spaces and $B^k(M) \subset Z^k(M)$. Like all vector spaces, these spaces are, a fortiori, also abelian groups.

Definition 16.2 *The quotient (vector) space $H^k(M) := Z^k(M)/B^k(M)$ is called the k-th de Rham cohomology group of M.*

We will start be computing two simple cases. First, let $M = \{p\}$. That is, M consists of a single point and is hence a 0-dimensional manifold. In this case,

$$\Omega^k(\{p\}) = Z^k(\{p\}) = \begin{cases} \mathbb{R} & \text{for } k = 0 \\ \\ 0 & \text{for } k > 0 \end{cases}$$

Furthermore, $B^k(M) = 0$ and so

$$H^k(\{p\}) = \begin{cases} \mathbb{R} & \text{for } k = 0 \\ \\ 0 & \text{for } k > 0 \end{cases}.$$

Next we consider the case $M = \mathbb{R}$. Here, $Z^0(\mathbb{R})$ is clearly just the constant functions and so is (isomorphic to) $\mathbb{R}$. On the other hand, $B^0(\mathbb{R}) = 0$ and so

$$H^0(\mathbb{R}) = \mathbb{R}.$$

Now since $d : \Omega^1(\mathbb{R}) \to \Omega^2(\mathbb{R}) = 0$ we see that $Z^1(\mathbb{R}) = \Omega^1(\mathbb{R})$. If $g(x)dx \in \Omega^1(\mathbb{R})$ then letting

$$f(x) := \int_0^x g(x)dx$$

we get $df = g(x)dx$. Thus, every $\Omega^1(\mathbb{R})$ is exact; $B^1(\mathbb{R}) = \Omega^1(\mathbb{R})$. We are led to

$$H^1(\mathbb{R}) = 0.$$

From this modest beginning we will be able to compute the de Rham cohomology for a large class of manifolds. Our first goal is to compute $H^k(\mathbb{R})$ for all k. In order to accomplish this we will need a good bit of preparation. The methods are largely algebraic and so will need to introduce a small portion of "homological algebra".

Definition 16.3 *Let R be a commutative ring. A **differential** $R-$**complex** is a direct sum of modules $C = \bigoplus_{k\in\mathbb{Z}} C^k$ together with a linear map $d : C \to C$ such that $d \circ d = 0$ and such that $d(C^k) \subset C^{k+1}$. Thus we have a sequence of linear maps*

$$\cdots C^{k-1} \xrightarrow{d} C^k \xrightarrow{d} C^{k+1}$$

where we have denoted the restrictions $d|\, C^k$ all simply by the single letter d.

Let $A = \bigoplus_{k\in\mathbb{Z}} A^k$ and $B = \bigoplus_{k\in\mathbb{Z}} B^k$ be differential complexes. A map $f : A \to B$ is called a chain map if f is a (degree 0) graded map such that $d \circ f = f \circ g$. In other words, if we let $f|\, A^k := f_k$ then we require that $f_k(A^k) \subset B^k$ and that the following diagram commutes for all k:

$$\begin{array}{ccccccccc}
\xrightarrow{d} & A^{k-1} & \xrightarrow{d} & A^k & \xrightarrow{d} & A^{k+1} & \xrightarrow{d} & \\
 & f_{k-1}\downarrow & & f_k\downarrow & & f_{k+1}\downarrow & & . \\
\xrightarrow{d} & B^{k-1} & \xrightarrow{d} & B^k & \xrightarrow{d} & B^{k+1} & \xrightarrow{d} &
\end{array}$$

Notice that if $f : A \to B$ is a chain map then $\ker(f)$ and $\mathrm{img}(f)$ are complexes with $\ker(f) = \bigoplus_{k\in\mathbb{Z}} \ker(f_k)$ and $\mathrm{img}(f) = \bigoplus_{k\in\mathbb{Z}} \mathrm{img}(f_k)$. Thus the notion of **exact sequence of chain maps** may be defined in the obvious way.

Definition 16.4 *The k-th cohomology of the complex $C = \bigoplus_{k\in\mathbb{Z}} C^k$ is*

$$H^k(C) := \frac{\ker(\,d|\, C^k)}{\mathrm{img}(\,d|\, C^{k-1})}$$

The elements of $\ker(\,d|\, C^k)$ (also denoted $Z^k(C)$) are called cocycles while the elements of $\mathrm{img}(\,d|\, C^{k-1})$ (also denoted $B^k(C)$) are called coboundaries.

We already have an example since by letting $\Omega^k(M) := 0$ for $k < 0$ we have a differential complex $d : \Omega(M) \to \Omega(M)$ where d is the exterior derivative. In this case, $H^k(\Omega(M)) = H^k(M)$ by definition. The reader may have noticed that $\Omega(M)$ is a A differential $R-$complex as well as a differential $C^\infty(M)-$complex. In fact, $\Omega(M)$ is a algebra under the exterior product (recall that $\wedge : \Omega^k(M) \times \Omega^l(M) \to \Omega^{l+k}(M)$). This algebra structure actually remains active at the level of cohomology: If $\alpha \in Z^k(M)$ and $\beta \in Z^l(M)$ then for any $\alpha', \beta' \in \Omega^{k-1}(M)$ and any $\beta' \in \Omega^{l-1}(M)$ we have

$$\begin{aligned}
(\alpha + d\alpha') \wedge \beta &= \alpha \wedge \beta + d\alpha' \wedge \beta \\
&= \alpha \wedge \beta + d(\alpha' \wedge \beta) - (-1)^{k-1}\alpha' \wedge d\beta \\
&= \alpha \wedge \beta + d(\alpha' \wedge \beta)
\end{aligned}$$

and similarly $\alpha \wedge (\beta + d\beta') = \alpha \wedge \beta + d(\alpha \wedge \beta')$. Thus we may define a product $H^k(M) \times H^l(M) \to H^{k+l}(M)$ by $[\alpha] \wedge [\beta] := [\alpha \wedge \beta]$.

If $f : A \to B$ is a chain map then it is easy to see that there is a natural (degree 0) graded map $f^* : H \to H$ defined by

$$f^*([x]) := [f(x)] \text{ for } x \in C^k.$$

Definition 16.5 *An exact sequence of chain maps of the form*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is called a ***short exact sequence****.*

Associated to every short exact sequence of chain maps there is a long exact sequence of cohomology groups:

$$\begin{array}{ccccccc}
 & & & & & & \delta \nwarrow \\
 & & H^k(A) & \xrightarrow{f^*} & H^k(B) & \xrightarrow{g^*} & H^k(C) \\
 & & & \delta \nwarrow & & & \\
H^k(A) & \xrightarrow{f^*} & H^k(B) & \xrightarrow{g^*} & H^k(C) & & \\
\nwarrow \delta & & & & & &
\end{array}$$

The maps f^* and g^* are the maps induced by f and g where the "connector map" $\delta : H^k(C) \to H^k(A)$ is defined as follows: Referring to the diagram below, let $c \in Z^k(C) \subset C^k$ so that $dc = 0$.

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A^{k+1} & \xrightarrow{f} & B^{k+1} & \xrightarrow{g} & C^{k+1} & \longrightarrow & 0 \\
 & & d\uparrow & & d\uparrow & & d\uparrow & & \\
0 & \longrightarrow & A^k & \xrightarrow{f} & B^k & \xrightarrow{g} & C^k & \longrightarrow & 0
\end{array}$$

By the surjectivity of g there is an $b \in B^k$ with $g(b) = c$. Also, since $g(db) = d(g(b)) = dc = 0$, it must be that $db = f(a)$ for some $a \in A^{k+1}$. The scheme of the process is

$$c \dashrightarrow b \dashrightarrow a.$$

Certainly $f(da) = d(f(a)) = ddb = 0$ and so since f is 1-1 we must have $da = 0$ which means that $a \in Z^{k+1}(C)$. We would like to define $\delta([c])$ to be $[a]$ but we must show that this is well defined. Suppose that we repeat this process starting with $c' = c + dc_{k-1}$ for some $c_{k-1} \in C^{k-1}$. In our first step we find $b' \in B^k$ with $g(b') = c'$ and then a' with $f(a') = db'$. We wish to show that $[a] = [a']$. We have $g(b - b') = c - c = 0$ and so there is an $a_k \in A^k$ with $f(a_k) = b - b'$. By commutativity we have

$$\begin{aligned} f(d(a_k)) &= d(f(a_k)) = d(b - b') \\ &= db - db' = f(a) - f(a') = f(a - a') \end{aligned}$$

and then since f is 1-1 we have $d(a_k) = a - a'$ which means that $[a] = [a']$. We leave it to the reader to check (if there is any doubt) that δ so defined is linear.

We now return to the de Rham cohomology. If $f : M \to N$ is a C^∞ map then we have $f^* : \Omega(N) \to \Omega(M)$. Since pull back commutes with exterior differentiation and preserves the degree of differential forms, f^* is a chain map.

Thus we have the induced map on the cohomology which we will also denote by f^* :

$$f^* : H^*(M) \to H^*(M)$$
$$f^* : [\alpha] \mapsto [f^*\alpha]$$

where we have used $H^*(M)$ to denote the direct sum $\bigoplus_i H^i(M)$. Notice that $f \mapsto f^*$ together with $M \mapsto H^*(M)$ is a contravariant functor since if $f : M \to N$ and $g : N \to P$ then

$$(g \circ f)^* = f^* \circ g^*.$$

In particular if $\iota_U : U \to M$ is inclusion of an open set U then $\iota_U^*\alpha$ is the same as restriction of the form α to U. If $[\alpha] \in H^*(M)$ then $f^*([\alpha]) \in H^*(U)$;

$$f^* : H^*(M) \to H^*(U).$$

16.10 The Meyer Vietoris Sequence

Suppose that $M = U_0 \cup U_1$ for open sets U. Let $U_0 \sqcup U_1$ denote the disjoint union of U and V. We then have inclusions $\iota_1 : U_1 \to M$ and $\iota_2 : U_2 \to M$ as well as the inclusions

$$\partial_0 : U_0 \cap U_1 \to U_1 \hookrightarrow U_0 \sqcup U_1$$

and

$$\partial_1 : U_0 \cap U_1 \to U_0 \hookrightarrow U_0 \sqcup U_1$$

which we indicate (following [Bott and Tu]) by writing

$$M \underset{\iota_1}{\overset{\iota_0}{\leftleftarrows}} \quad U_0 \sqcup U_1 \underset{\partial_1}{\overset{\partial_0}{\leftleftarrows}} U_0 \cap U_1 \quad .$$

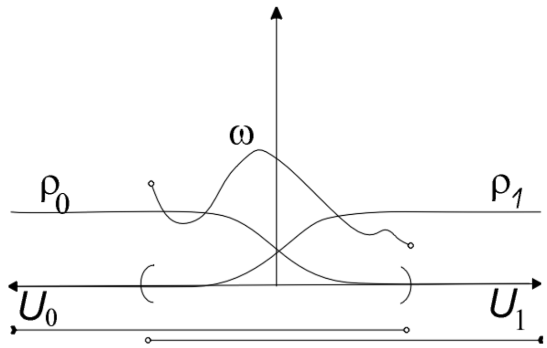
This gives rise to an exact sequence

$$0 \to \Omega(M) \xrightarrow{\iota^*} \Omega(U_0) \oplus \Omega(U_1) \xrightarrow{\partial^*} \Omega(U_0 \cap U_1) \to 0$$

where $\iota(\omega) := (\iota_0^*\omega, \iota_1^*\omega)$ and $\partial^*(\alpha, \beta) := (\partial_0^*(\beta) - \partial_1^*(\alpha))$. Notice that $\iota_0^*\omega \in \Omega(U_0)$ while $\iota_1^*\omega \in \Omega(U_1)$. Also, $\partial_0^*(\beta) = \beta|_{U_0 \cap U_1}$ and $\partial_1^*(\alpha) = \alpha|_{U_0 \cap U_1}$ and live in $\Omega(U_0 \cap U_1)$.

Let us show that this sequence is exact. First if $\iota(\omega) := (\iota_1^*\omega, \iota_0^*\omega) = (0, 0)$ then $\omega|_{U_0} = \omega|_{U_1} = 0$ and so $\omega = 0$ on $M = U_0 \cup U_1$ thus ι^* is 1-1 and exactness at $\Omega(M)$ is demonstrated.

Next, if $\eta \in \Omega(U_0 \cap U_1)$ then we take a smooth partition of unity $\{\rho_0, \rho_1\}$ subordinate to the cover $\{U_0, U_1\}$ and then let $\omega := (-(\rho_1\eta)^{U_0}, (\rho_0\eta)^{U_1})$ where we have extended $\rho_1|_{U_0 \cap U_1} \eta$ by zero to a smooth function $(\rho_1\eta)^{U_0}$ on U_0 and



$\rho_0|_{U_0\cap U_1}\,\eta$ to a function $(\rho_0\eta)^{U_1}$ on U_1 (think about this). Now we have

$$\begin{aligned}
&\partial^*(-(\rho_1\eta)^{U_0},(\rho_0\eta)^{U_1})\\
&=((\rho_0\eta)^{U_1}|_{U_0\cap U_1}+(\rho_1\eta)^{U_0}|_{U_0\cap U_1})\\
&=\rho_0\eta|_{U_0\cap U_1}+\rho_1\eta|_{U_0\cap U_1}\\
&=(\rho_0+\rho_1)\eta=\eta.
\end{aligned}$$

Perhaps the notation is too pedantic. If we let the restrictions and extensions by zero take care of themselves, so to speak, then the idea is expressed by saying that ∂^* maps $(-\rho_1\eta,\rho_0\eta)\in\Omega(U_0)\oplus\Omega(U_1)$ to $\rho_0\eta-(-\rho_1\eta)=\eta\in\Omega(U_0\cap U_1)$. Thus we see that ∂^* is surjective.

It is easy to see that $\partial^*\circ\iota^*=0$ so that $\mathrm{img}(\partial^*)\subset\ker(\iota^*)$. Finally, let $(\alpha,\beta)\in\Omega(U_0)\oplus\Omega(U_1)$ and suppose that $\partial^*(\alpha,\beta)=(0,0)$. This translates to $\alpha|_{U_0\cap U_1}=\beta|_{U_0\cap U_1}$ which means that there is a form $\omega\in\Omega(U_0\cup U_1)=\Omega(M)$ such that ω coincides with α on U_0 and with β on U_0. Thus

$$\begin{aligned}
\iota^*\omega&=(\iota_0^*\omega,\iota_1^*\omega)\\
&=(\alpha,\beta)
\end{aligned}$$

so that $\ker(\iota^*)\subset\mathrm{img}(\partial^*)$ which together with the reverse inclusion gives $\mathrm{img}(\partial^*)=\ker(\iota^*)$.

16.11 Sheaf Cohomology

Under Construction

16.12 Characteristic Classes

Chapter 17

Lie Groups and Lie Algebras

17.1 Lie Algebras

Let $\mathbb{F}$ denote on of the fields $\mathbb{R}$ or $\mathbb{C}$. In definition 7.8 we defined a real Lie algebra $\mathfrak{g}$ as a real algebra with a skew symmetric (bilinear) product (the Lie bracket), usually denoted with a bracket $v, w \mapsto [v, w]$, such that the Jacobi identity holds

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \text{ for all } x, y, z \in \mathfrak{g}. \qquad \text{(Jacobi Identity)}$$

We also have the notion of a complex Lie algebra define analogously.

Remark 17.1 *We will assume that all the Lie algebras we study are finite dimensional unless otherwise indicated.*

Let V be a finite dimensional vector space and recall that $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ is the set of all $\mathbb{F}-$linear maps $\mathrm{V} \to \mathrm{V}$. The space $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ is also denoted $\mathrm{Hom}_{\mathbb{F}}(\mathrm{V}, \mathrm{V})$ or $L_{\mathbb{F}}(\mathrm{V}, \mathrm{V})$ although in the context of Lie algebras we take $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ as the preferred notation. We give $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ its natural Lie algebra structure where the bracket is just the commutator bracket

$$[A, B] := A \circ B - B \circ A.$$

If the field involved is either irrelevant or known from context we will just write $\mathfrak{gl}(\mathrm{V})$. Also, we often identify $\mathfrak{gl}(\mathbb{F}^n)$ with the matrix Lie algebra $\mathbb{M}_{nxn}(\mathbb{F})$ with the bracket $AB - BA$.

For a Lie algebra $\mathfrak{g}$ we can associate to every basis $v_1, ..., v_n$ for $\mathfrak{g}$ the so called structure constants c_{ij}^k defined by

$$[v_i, v_j] = \sum_k c_{ij}^k v_k.$$

It then follows from the skew symmetry of the Lie bracket and the Jacobi identity it follow that the structure constants satisfy

$$\begin{array}{ll} \text{i)} & c_{ij}^k = -c_{ji}^k \\ \text{ii)} & \sum_k c_{rs}^k c_{kt}^i + c_{st}^k c_{kr}^i + c_{tr}^k c_{ks}^i = 0 \end{array} \tag{17.1}$$

Given a real Lie algebra $\mathfrak{g}$ we can extend to a complex Lie algebra $\mathfrak{g}_{\mathbb{C}}$ by defining as $\mathfrak{g}_{\mathbb{C}}$ the complexification $\mathfrak{g}_{\mathbb{C}} := \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$ and then extending the bracket by requiring

$$[v \otimes 1, w \otimes 1] = [v, w] \otimes 1.$$

Then $\mathfrak{g}$ can be identified with its image under the embedding map $v \mapsto v \otimes 1$. In practice one often omits the symbol $\otimes$ and with the aforementioned identification of $\mathfrak{g}$ as a subspace of $\mathfrak{g}_{\mathbb{C}}$ the complexification just amounts to allowing complex coefficients.

Notation 17.1 *Given two subsets S_1 and S_2 of a Lie algebra $\mathfrak{g}$ we let $[\ S_1, S_2]$ denote the linear span of the set defined by $\{[x, y] : x \in S_1$ and $y \in S_2\}$. Also, let $S_1 + S_2$ denote the vector space of all $x + y : x \in S_1$ and $y \in S_2$.*

It is easy to verify that the following relations hold:

1. $[S_1 + S_2, S] \subset [S_1, S] + [S_2, S]$
2. $[S_1, S_2] = [S_2, S_1]$
3. $[S, [S_1, S_2]] \subset [[S, S_1], S_2] + [S_1, [S, S_2]]$

where S_1, S_2 and S are subsets of a Lie algebra $\mathfrak{g}$.

Definition 17.1 *A vector subspace $\mathfrak{a} \subset \mathfrak{g}$ is called a* ***subalgebra*** *if $[\mathfrak{a}, \mathfrak{a}] \subset \mathfrak{g}$ and an* ***ideal*** *if $[\mathfrak{g}, \mathfrak{a}] \subset \mathfrak{g}$.*

If $\mathfrak{a}$ is a subalgebra of $\mathfrak{g}$ and $v_1, ..., v_k, v_{k+1}, ..., v_n$ is a basis for $\mathfrak{g}$ such that $v_1, ..., v_k$ is a basis for $\mathfrak{a}$ then with respect to this basis the structure constants are such that

$$c_{ij}^s = 0 \text{ for } i, j \le k \text{ and } s > k.$$

If $\mathfrak{a}$ is also an ideal then we must have

$$c_{ij}^s = 0 \text{ for } i \le k \text{ and } s > k \text{ with any } j.$$

Remark 17.2 *Notice that in general c_{ij}^s may be viewed as the components of a an element (a tensor) of $T_{1,1}^1(\mathfrak{g})$.*

Example 17.1 *Let $\mathfrak{su}(2)$ denote the set of all traceless and Hermitian 2×2 complex matrices. This is a Lie algebra under the commutator bracket ($AB - BA$). A commonly used basis for $\mathfrak{su}(2)$ is e_1, e_2, e_3 where*

$$e_1 = \frac{1}{2}\begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & \end{bmatrix}, \quad e_2 = \frac{1}{2}\begin{bmatrix} 0 & -1 \\ 1 & \end{bmatrix}, \quad e_2 = \frac{1}{2}\begin{bmatrix} -\mathrm{i} & 0 \\ 0 & \mathrm{i} \end{bmatrix}.$$

The commutation relations satisfied by these matrices are

$$[e_i, e_j] = \epsilon_{ijk} e_k$$

where ϵ_{ijk} *is the totally antisymmetric symbol. Thus in this case the structure constants are* $c_{ij}^k = \epsilon_{ijk}$. *In physics it is common to use the Pauli matrices defined by* $\sigma_i := 2\mathrm{i}e_i$ *in terms of which the commutation relations become* $[\sigma_i, \sigma_j] = 2\mathrm{i}\epsilon_{ijk} e_k$.

Example 17.2 *The Weyl basis for* $\mathfrak{gl}(n, \mathbb{R})$ *is given by the* n^2 *matrices* e_{sr} *defined by*

$$(e_{rs})_{ij} := \delta_{ri}\delta_{sj}.$$

Notice that we are now in a situation where "double indices" will be convenient. For instance, the commutation relations read

$$[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta_{il} e_{kj}$$

while the structure constants are

$$c_{sm,kr}^{ij} = \delta_s^i \delta_{mk} \delta_r^j - \delta_k^i \delta_{rs} \delta_m^j.$$

17.2 Classical complex Lie algebras

If $\mathfrak{g}$ is a real Lie algebra we have seen that the complexification $\mathfrak{g}_{\mathbb{C}}$ is naturally a complex Lie algebra. Is convenient to omit the tensor symbol and use the following convention: Every element of $\mathfrak{g}_{\mathbb{C}}$ may be written at $v + \mathrm{i}w$ for $v, w \in \mathfrak{g}$ and then

$$\begin{aligned}&[v_1 + \mathrm{i}w_1, v_2 + \mathrm{i}w_2] \\ &= [v_1, v_2] - [w_1, w_2] + \mathrm{i}([v_1, w_2] + [w_1, v_2]).\end{aligned}$$

We shall now define a series of complex Lie algebras sometimes denoted by A_n, B_n, C_n and D_n for every integer $n > 0$. First of all, notice that the complexification $\mathfrak{gl}(n, \mathbb{R})_{\mathbb{C}}$ of $\mathfrak{gl}(n, \mathbb{R})$ is really just $\mathfrak{gl}(n, \mathbb{C})$; the set of complex $n \times n$ matrices under the commutator bracket.

The algebra A_n The set of all traceless $n \times n$ matrices is denoted A_{n-1} and also by $\mathfrak{sl}(n, \mathbb{C})$.

We call the readers attention to the follow general fact: If $b(.,.)$ is a bilinear form on a complex vector space V then the set of all $A \in \mathfrak{gl}(n, \mathbb{C})$ such that $b(Az, w) + b(z, Aw) = 0$ is a Lie subalgebra of $\mathfrak{gl}(n, \mathbb{C})$. This follows from

$$\begin{aligned}b([A, B]z, w) &= b(ABz, w) - b(BAz, w) \\ &= -b(Bz, Aw) + b(Az, Bw) \\ &= b(z, BAw) - b(z, ABw) \\ &= b(z, [A, B]w).\end{aligned}$$

The algebras B_n and D_n Let $m = 2n + 1$ and let $b(.,.)$ be a nondegenerate symmetric bilinear form on an m dimensional complex vector space V. Without loss we may assume $V = \mathbb{C}^m$ and we may take $b(z, w) = \sum_{i=1}^{m} z_i w_i$. We define B_n to be $\mathfrak{o}(m, \mathbb{C})$ where

$$\mathfrak{o}(m, \mathbb{C}) := \{A \in \mathfrak{gl}(m, \mathbb{C}) : b(Az, w) + b(z, Aw) = 0\}.$$

Similarly, for $m = 2n$ we define D_n to be $\mathfrak{o}(m, \mathbb{C})$.

The algebra C_n The algebra associated to a skew-symmetric nondegenerate bilinear form which we may take to be $b(z, w) = \sum_{i=1}^{n} z_i w_{n+i} - \sum_{i=1}^{n} z_{n+i} w_i$ on $\mathbb{C}^{2n}$ we have the symplectic algebra

$$C_n = \mathfrak{sp}(n, \mathbb{C}) := \{A \in \mathfrak{gl}(m, \mathbb{C}) : b(Az, w) + b(z, Aw) = 0\}.$$

17.2.1 Basic Definitions

Definition 17.2 *Given two Lie algebras over a field $\mathbb{F}$, say $\mathfrak{a}, [,]_{\mathfrak{a}}$ and $\mathfrak{b}, [,]_{\mathfrak{b}}$, an $\mathbb{F}$-linear map σ is called a **Lie algebra homomorphism** iff*

$$\sigma([v, w]_{\mathfrak{a}}) = [\sigma v, \sigma w]_{\mathfrak{b}}$$

*for all $v, w \in \mathfrak{a}$. A Lie algebra isomorphism is defined in the obvious way. A Lie algebra **isomorphism** $\mathfrak{g} \to \mathfrak{g}$ is called an **automorphism** of $\mathfrak{g}$.*

It is not hard to show that the set of all automorphisms of $\mathfrak{g}$, denoted $\mathrm{Aut}(\mathfrak{g})$, forms a Lie group (actually a Lie subgroup of $\mathfrak{gl}(\mathfrak{g})$).

The expected theorems hold for homomorphisms; the image $\mathrm{img}(\sigma) := \sigma(\mathfrak{a})$ of a homomorphism $\sigma : \mathfrak{a} \to \mathfrak{b}$ is a subalgebra of $\mathfrak{b}$ and the kernel $\ker(\sigma)$ is an ideal of $\mathfrak{a}$.

Definition 17.3 *Let $\mathfrak{h}$ be an ideal in $\mathfrak{g}$. On the quotient vector space $\mathfrak{g}/\mathfrak{h}$ with quotient map π we can define a Lie bracket in the following way: For $\bar{v}, \bar{w} \in \mathfrak{g}/\mathfrak{h}$ choose $v, w \in \mathfrak{g}$ with $\pi(v) = \bar{v}$ and $\pi(w) = \bar{w}$ we define*

$$[\bar{v}, \bar{w}] := \pi([v, w]).$$

*We call $\mathfrak{g}/\mathfrak{h}$ with this bracket a **quotient Lie algebra**.*

Exercise 17.1 *Show that the bracket defined in the last definition is well defined.*

Given two linear subspaces $\mathfrak{a}$ and $\mathfrak{b}$ of $\mathfrak{g}$ the (not necessarily direct) sum $\mathfrak{a} + \mathfrak{b}$ is just the space of all elements in $\mathfrak{g}$ of the form $a + b$ where $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$. It is not hard to see that if $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in $\mathfrak{g}$ then so is $\mathfrak{a} + \mathfrak{b}$.

Exercise 17.2 *Show that for $\mathfrak{a}$ and $\mathfrak{b}$ ideals in $\mathfrak{g}$ we have a natural isomorphism $(\mathfrak{a} + \mathfrak{b})/\mathfrak{b} \cong \mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$.*

Definition 17.4 *If $\mathfrak{g}$ is a Lie algebra, $\mathfrak{b}$ and $\mathfrak{c}$ subsets of $\mathfrak{g}$ then the* ***centralizer of*** $\mathfrak{b}$ ***in*** $\mathfrak{c}$ *is* $\{v \in \mathfrak{c} : [v, \mathfrak{b}] = 0\}$.

Definition 17.5 *If $\mathfrak{a}$ is a (Lie) subalgebra of $\mathfrak{g}$ then the* ***normalizer*** *of $\mathfrak{a}$ in $\mathfrak{g}$ is*

$$\mathfrak{n}(\mathfrak{a}) := \{v \in \mathfrak{g} : [v, \mathfrak{a}] \subset \mathfrak{a}\}.$$

One can check that $\mathfrak{n}(\mathfrak{a})$ is an ideal in $\mathfrak{g}$.

There is also a Lie algebra product. Namely, if $\mathfrak{a}$ and $\mathfrak{b}$ are Lie algebras, then we can define a Lie bracket on $\mathfrak{a} \times \mathfrak{b}$ by

$$[(a_1, a_2), (b_1, b_2)] := ([a_1, b_1], [a_2, b_2]).$$

With this bracket $\mathfrak{a} \times \mathfrak{b}$ is a Lie algebra called the Lie algebra **product** of $\mathfrak{a}$ and $\mathfrak{b}$. The subspaces $\mathfrak{a} \times \{0\}$ and $\{0\} \times \mathfrak{b}$ are ideals in $\mathfrak{a} \times \mathfrak{b}$ which are clearly isomorphic to $\mathfrak{a}$ and $\mathfrak{b}$ respectively.

Definition 17.6 *The* ***center*** $\mathfrak{z}(\mathfrak{g})$ *of a Lie algebra $\mathfrak{g}$ is the subspace* $\mathfrak{z}(\mathfrak{g}) := \{x \in \mathfrak{g} : [x, y] = 0$ *for all* $y \in \mathfrak{g}\}$.

17.3 The Adjoint Representation

The center $\mathfrak{z}(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$ is the largest ideal $\mathfrak{z}$ such that $[\mathfrak{g}, \mathfrak{z}] \subset \mathfrak{z}$. It is easy to see that $\mathfrak{z}(\mathfrak{g})$ is the kernel of the map $v \to \mathrm{ad}(v)$ where $\mathrm{ad}(v) \in \mathfrak{gl}(\mathfrak{g})$ is given by $\mathrm{ad}(v)(x) := [v, x]$.

Definition 17.7 *A derivation of a Lie algebra $\mathfrak{g}$ is a linear map $D : \mathfrak{g} \to \mathfrak{g}$ such that*

$$D[v, w] = [Dv, w] + [v, Dw]$$

for all $v, w \in \mathfrak{g}$.

For each $v \in \mathfrak{g}$ the map $\mathrm{ad}(v) : \mathfrak{g} \to \mathfrak{g}$ is actually a derivation of the Lie algebra $\mathfrak{g}$. Indeed, this is exactly the content of the Jacobi identity. Furthermore, it is not hard to check that the space of all derivations of a Lie algebra $\mathfrak{g}$ is a subalgebra of $\mathfrak{gl}(\mathfrak{g})$. In fact, if D_1 and D_2 are derivations of $\mathfrak{g}$ then so is the commutator $D_1 \circ D_2 - D_2 \circ D_1$. We denote this subalgebra of derivations by $Der(\mathfrak{g})$.

Definition 17.8 *A* ***Lie algebra representation*** *ρ of $\mathfrak{g}$ on a vector space* V *is a Lie algebra homomorphism* $\rho : \mathfrak{g} \to \mathfrak{gl}(\mathrm{V})$.

One can construct Lie algebra representations in various way from given representations. For example, if $\rho_i : \mathfrak{g} \to \mathfrak{gl}(\mathrm{V}_i)$ $(i = 1, .., k)$ are Lie algebra representations then $\oplus \rho_i : \mathfrak{g} \to \mathfrak{gl}(\oplus_i \mathrm{V}_i)$ defined by

$$(\oplus_i \rho_i)(x)(v_1 \oplus \cdots \oplus v_n) = \rho_1(x) v_1 \oplus \cdots \oplus \rho_1(x) v_n \tag{17.2}$$

for $x \in \mathfrak{g}$ is a Lie algebra representation called the **direct sum representation** of the ρ_i. Also, if one defines

$$\begin{aligned}(\otimes_i \rho_i)(x)(\otimes_i v_i) &:= \rho_1(x)v_1 \otimes v_2 \otimes \cdots \otimes v_k \\ &+ v_1 \otimes \rho_2(x)v_2 \otimes \cdots \otimes v_k + \cdots + v_1 \otimes v_2 \otimes \cdots \otimes \rho_k(x)v_k\end{aligned}$$

(and extend linearly) then $\otimes_i \rho_i$ is a representation $\otimes_i \rho_i : \mathfrak{g} \to \mathfrak{gl}(\otimes_i \mathrm{V}_i)$ is Lie algebra representation called a **tensor product representation**.

Lemma 17.1 $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ *is a Lie algebra representation on* $\mathfrak{g}$. *The image of* ad *is contained in the Lie algebra* $\mathrm{Der}(\mathfrak{g})$ *of all derivation of the Lie algebra* $\mathfrak{g}$.

Proof. This follows from the Jacobi identity (as indicated above) and from the definition of ad. ■

Corollary 17.1 $\mathfrak{z}(\mathfrak{g})$ *is an ideal in* $\mathfrak{g}$.

The image $\mathrm{ad}(\mathfrak{g})$ of ad in $\mathrm{Der}(\mathfrak{g})$ is called the adjoint algebra.

Definition 17.9 *The Killing form for a Lie algebra* $\mathfrak{g}$ *is the bilinear form given by*

$$K(X, Y) = Tr(\mathrm{ad}(X) \circ \mathrm{ad}(Y))$$

Lemma 17.2 *For any Lie algebra automorphism* $\rho : \mathfrak{g} \to \mathfrak{g}$ *and any* $X \in \mathfrak{g}$ *we have* $\mathrm{ad}(\rho X) = \rho \mathrm{ad} X \rho^{-1}$

Proof. $\mathrm{ad}(\rho X)(Y) = [\rho X, Y] = [\rho X, \rho\rho^{-1}Y] = \rho[X, \rho^{-1}Y] = \rho \circ \mathrm{ad}X \circ \rho^{-1}(Y)$. ■

Proposition 17.1 *The Killing forms satisfies identities:*
1) $K([X, Y], Z) = K([Z, X], Y)$ *for all* $X, Y, Z \in \mathfrak{g}$
2) $K(\rho X, \rho Y) = K(X, Y)$ *for any Lie algebra automorphism* $\rho : \mathfrak{g} \to \mathfrak{g}$ *and any* $X, Y \in \mathfrak{g}$.

Proof. For (1) we calculate

$$\begin{aligned} K([X, Y], Z) &= Tr(\mathrm{ad}([X, Y]) \circ \mathrm{ad}(Z)) \\ &= Tr([\mathrm{ad}X, \mathrm{ad}Y] \circ \mathrm{ad}(Z)) \\ &= Tr(\mathrm{ad}X \circ \mathrm{ad}Y \circ \mathrm{ad}Z - \mathrm{ad}Y \circ \mathrm{ad}X \circ \mathrm{ad}Z) \\ &= Tr(\mathrm{ad}Z \circ \mathrm{ad}X \circ \mathrm{ad}Y - \mathrm{ad}X \circ \mathrm{ad}Z \circ \mathrm{ad}Y) \\ &= Tr([\mathrm{ad}Z, \mathrm{ad}X] \circ \mathrm{ad}Y) \\ &= Tr(\mathrm{ad}[Z, X] \circ \mathrm{ad}Y) = K([Z, X], Y) \end{aligned}$$

where we have used that $Tr(ABC)$ is invariant under cyclic permutations of A, B, C.

For (2) just observe that

$$\begin{aligned}
K(\rho X, \rho Y) &= Tr(\mathrm{ad}(\rho X) \circ \mathrm{ad}(\rho Y)) \\
&= Tr(\rho \mathrm{ad}(X)\rho^{-1}\rho \mathrm{ad}(Y)\rho^{-1}) \quad \text{(lemma 17.2)} \\
&= Tr(\rho \mathrm{ad}(X) \circ \mathrm{ad}(Y)\rho^{-1}) \\
&= Tr(\mathrm{ad}(X) \circ \mathrm{ad}(Y)) = K(X, Y).
\end{aligned}$$

■

Now clearly, $K(X, Y)$ is symmetric in X, Y and so there must be a basis $\{X_i\}_{1 \le i \le n}$ of $\mathfrak{g}$ for which the matrix (k_{ij}) given by

$$k_{ij} := K(X_i, X_j)$$

is diagonal.

Lemma 17.3 *If $\mathfrak{a}$ is an ideal in $\mathfrak{g}$ then the Killing form of $\mathfrak{a}$ is just the Killing form of $\mathfrak{g}$ restricted to $\mathfrak{a} \times \mathfrak{a}$.*

Proof. Let $\{X_i\}_{1 \le i \le n}$ be a basis of $\mathfrak{g}$ such that $\{X_i\}_{1 \le i \le r}$ is a basis for $\mathfrak{a}$. Now since $[\mathfrak{a}, \mathfrak{g}] \subset \mathfrak{a}$, the structure constants c^i_{jk} with respect to this basis must have the property that $c^k_{ij} = 0$ for $i \le r < k$ and all j. Thus for $1 \le i, j \le r$ we have

$$\begin{aligned}
K_{\mathfrak{a}}(X_i, X_j) &= Tr(\mathrm{ad}(X_i)\mathrm{ad}(X_j)) \\
&= \sum_{i,j=1}^{r} c^i_{ik} c^k_{ji} = \sum_{i,j=1}^{r} c^i_{ik} c^k_{ji} \\
&= K_{\mathfrak{g}}(X_i, X_j).
\end{aligned}$$

■

17.4 The Universal Enveloping Algebra

In a Lie algebra $\mathfrak{g}$ the product $[.,.]$ is usually not associative. On the other hand if $\mathfrak{A}$ is an associative algebra then we can introduce the commutator bracket on $\mathfrak{A}$ by

$$[A, B] := AB - BA$$

which gives $\mathfrak{A}$ the structure of Lie algebra. From the other direction, if we start with a Lie algebra $\mathfrak{g}$ then we can construct an associative algebra called the universal enveloping algebra of $\mathfrak{g}$. This is done, for instance, by first forming the full tensor algebra on $\mathfrak{g}$;

$$T(\mathfrak{g}) = \mathbb{F} \oplus \mathfrak{g} \oplus (\mathfrak{g} \otimes \mathfrak{g}) \oplus \cdots \oplus \mathfrak{g}^{\otimes k} \oplus \cdots$$

and then dividing out by an appropriate ideal:

Definition 17.10 *Associated to every Lie algebra $\mathfrak{g}$ there is an associative algebra $U(\mathfrak{g})$ called the* ***universal enveloping algebra*** *defined by*

$$U(\mathfrak{g}) := T(\mathfrak{g})/J$$

where J is the ideal generated by elements in $T(\mathfrak{g})$ of the form $X \otimes Y - Y \otimes X - [X, Y]$.

There is the natural map of $\mathfrak{g}$ into $U(\mathfrak{g})$ given by the composition $\pi : \mathfrak{g} \hookrightarrow T(\mathfrak{g}) \to T(\mathfrak{g})/J = U(\mathfrak{g})$. For $v \in \mathfrak{g}$, let v^* denote the image of v under this canonical map.

Theorem 17.1 *Let V be a vector space over the field $\mathbb{F}$. For every ρ representation of $\mathfrak{g}$ on V there is a corresponding representation ρ^* of $U(\mathfrak{g})$ on V such that for all $v \in \mathfrak{g}$ we have*

$$\rho(v) = \rho^*(v^*).$$

This correspondence, $\rho \mapsto \rho^*$ is a 1-1 correspondence.

Proof. Given ρ, there is a natural representation $T(\rho)$ on $T(\mathfrak{g})$. The representation $T(\rho)$ vanishes on J since

$$T(\rho)(X \otimes Y - Y \otimes X - [X, Y]) = \rho(X)\rho(Y) - \rho(Y)\rho(X) - \rho([X, Y]) = 0$$

and so $T(\rho)$ descends to a representation ρ^* of $U(\mathfrak{g})$ on $\mathfrak{g}$ satisfying $\rho(v) = \rho^*(v^*)$. Conversely, if σ is a representation of $U(\mathfrak{g})$ on V then we put $\rho(X) = \sigma(X^*)$. The map $\rho(X)$ is linear and a representation since

$$\begin{aligned}\rho([X, Y]) &= \sigma([X, Y]^*) \\ &= \sigma(\pi(X \otimes Y - Y \otimes X)) \\ &= \sigma(X^*Y^* - Y^*X^*) \\ &= \rho(X)\rho(Y) - \rho(Y)\rho(X)\end{aligned}$$

for all $X, Y \in \mathfrak{g}$. ∎

Now let $X_1, X_2, ..., X_n$ be a basis for $\mathfrak{g}$ and then form the monomials $X_{i_1}^* X_{i_2}^* \cdots X_{i_r}^*$ in $U(\mathfrak{g})$. The set of all such monomials for a fixed r, span a subspace of $U(\mathfrak{g})$ which we denote by $U^r(\mathfrak{g})$. Let c_{ik}^j be the structure constants for the basis $X_1, X_2, ..., X_n$. Then under the map π the structure equations become

$$[X_i^*, X_j^*] = \sum_k c_{ij}^k X_k^*.$$

By using this relation we can replace the spanning set $\mathcal{M}_r = \{X_{i_1}^* X_{i_2}^* \cdots X_{i_r}^*\}$ for $U^r(\mathfrak{g})$ by spanning set $\mathcal{M}_{\leq r}$ for $U^r(\mathfrak{g})$ consisting of all monomials of the form $X_{i_1}^* X_{i_2}^* \cdots X_{i_m}^*$ where $i_1 \leq i_2 \leq \cdots \leq i_m$ and $m \leq r$. In fact one can then concatenate these spanning sets $\mathcal{M}_{\leq r}$ and turns out that these combine to form a basis for $U(\mathfrak{g})$. We state the result without proof:

Theorem 17.2 (Birchoff-Poincarè-Witt) *Let* $e_{i_1 \le i_2 \le \cdots \le i_m} = X_{i_1}^* X_{i_2}^* \cdots X_{i_m}^*$ *where* $i_1 \le i_2 \le \cdots \le i_m$*. The set of all such elements* $\{e_{i_1 \le i_2 \le \cdots \le i_m}\}$ *for all* m *is a basis for* $U(\mathfrak{g})$ *and the set* $\{e_{i_1 \le i_2 \le \cdots \le i_m}\}_{m \le r}$ *is a basis for the subspace* $U^r(\mathfrak{g})$*.*

Lie algebras and Lie algebra representations play an important role in physics and mathematics and as we shall see below every Lie group has an associated Lie algebra which, to a surprisingly large extent, determines the structure of the Lie group itself. Let us first explore some of the important abstract properties of Lie algebras. A notion that is useful for constructing Lie algebras with desired properties is that of the **free Lie algebra** $\mathfrak{f}_n$ which is defined to be the free algebra subject only to the relations define the notion of Lie algebra. Every Lie algebra can be realized as a quotient of one of these free Lie algebras.

Definition 17.11 *The* ***descending central series*** $\{\mathfrak{g}_{(k)}\}$ *of a Lie algebra* $\mathfrak{g}$ *is defined inductively by letting* $\mathfrak{g}_{(1)} = \mathfrak{g}$ *and then* $\mathfrak{g}_{(k+1)} = [\mathfrak{g}_{(k)}, \mathfrak{g}]$*.*

From the definition of Lie algebra homomorphism we see that if $\sigma : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism then $\sigma(\mathfrak{g}_{(k)}) \subset \mathfrak{h}_{(k)}$.

Exercise 17.3 (!) *Use the Jacobi identity to prove that for all positive integers* i *and* j*, we have* $[\mathfrak{g}_{(i)}, \mathfrak{g}_{(i)}] \subset \mathfrak{g}_{(i+j)}$*.*

Definition 17.12 *A Lie algebra* $\mathfrak{g}$ *is called* ***k-step nilpotent*** *iff* $\mathfrak{g}_{(k+1)} = 0$ *but* $\mathfrak{g}_{(k)} \neq 0$*.*

The most studied nontrivial examples are the Heisenberg algebras which are 2-step nilpotent. These are defined as follows:

Example 17.3 *The* $2n+1$ *dimensional* ***Heisenberg algebra*** $\mathfrak{h}_n$ *is the Lie algebra (defined up to isomorphism) with a basis* $\{X_1, ..., X_n, Y_1, ..., Y_n, Z\}$ *subject to the relations*

$$[X_j, Y_j] = Z$$

and all other brackets of elements from this basis being zero. A concrete realization of $\mathfrak{h}_n$ *is given as the set of all* $(n+2) \times (n+2)$ *matrices of the form*

$$\begin{bmatrix} 0 & x_1 & ... & x_n & z \\ 0 & 0 & ... & 0 & y_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & y_n \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}$$

where x_i, y_i, z *are all real numbers. The bracket is the commutator bracket as is usually the case for matrices. The basis is realized in the obvious way by putting* *a lone 1 in the various positions corresponding to the potentially nonzero entries. For example,*

$$X_1 = \begin{bmatrix} 0 & 1 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}$$

and

$$Z = \begin{bmatrix} 0 & 0 & ... & 0 & 1 \\ 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}.$$

Example 17.4 *The space of all upper triangular $n \times n$ matrices $\mathfrak{n}_n$ which turns out to be $n-1$ step nilpotent.*

We also have the **free k-step nilpotent Lie algebra** given by the quotient $\mathfrak{f}_{n,k} := \mathfrak{f}_n/(\mathfrak{f}_n)_{(k)}$ where $\mathfrak{f}_n$ is the free Lie algebra mentioned above.

Lemma 17.4 *Every finitely generated k-step nilpotent Lie algebra is isomorphic to a quotient of the free k-step nilpotent Lie algebra.*

Proof. Suppose that $\mathfrak{g}$ is k-step nilpotent and generated by elements $X_1, ..., X_n$. Let $F_1, ..., F_n$ be the generators of $\mathfrak{f}_n$ and define a map $h : \mathfrak{f}_n \to \mathfrak{g}$ by sending F_i to X_i and extending linearly. This map clearly factors through $\mathfrak{f}_{n,k}$ since $h((\mathfrak{f}_n)_k) = 0$. Then we have a homomorphism $(\mathfrak{f}_n)_k \to \mathfrak{g}$ which is clearly onto and so the result follows. ∎

Definition 17.13 *Let $\mathfrak{g}$ be a Lie algebra. We define the commutator series $\{\mathfrak{g}^{(k)}\}$ by letting $\mathfrak{g}^{(1)} = \mathfrak{g}$ and then inductively $\mathfrak{g}^{(k)} = [\mathfrak{g}^{(k-1)}, \mathfrak{g}^{(k-1)}]$. If $\mathfrak{g}^{(k)} = 0$ for some positive integer k, then we call $\mathfrak{g}$ a **solvable** Lie algebra.*

Clearly, the statement $\mathfrak{g}^{(2)} = 0$ is equivalent to the statement that $\mathfrak{g}$ is abelian. Another simple observation is that $\mathfrak{g}^{(k)} \subset \mathfrak{g}_{(k)}$ so that nilpotency implies solvability.

Exercise 17.4 (!) *Every subalgebra and every quotient algebra of a solvable Lie algebra is solvable. In particular, the homomorphic image of a solvable Lie algebra is solvable. Conversely, if $\mathfrak{a}$ is a solvable ideal in $\mathfrak{g}$ and $\mathfrak{g}/\mathfrak{a}$ is solvable, then $\mathfrak{g}$ is solvable. Hint: Use that $(\mathfrak{g}/\mathfrak{a})^{(j)} = \mathfrak{g}^{(j)}/\mathfrak{a}$.*

It follows from this exercise that we have

Corollary 17.2 *Let $h : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra homomorphism. If* $\mathrm{img}(h) := h(\mathfrak{g})$ *and* $\ker(h)$ *are both solvable then $\mathfrak{g}$ is solvable. In particular, if* $\mathrm{img}(\mathrm{ad}) := \mathrm{ad}(\mathfrak{g})$ *is solvable then so is $\mathfrak{g}$.*

Lemma 17.5 *If $\mathfrak{a}$ is a nilpotent ideal in $\mathfrak{g}$ contained in the center $\mathfrak{z}(\mathfrak{g})$ and if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then $\mathfrak{g}$ is nilpotent.*

Proof. First, the reader can verify that $(\mathfrak{g}/\mathfrak{a})_{(j)} = \mathfrak{g}_{(j)}/\mathfrak{a}$. Now if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then $\mathfrak{g}_{(j)}/\mathfrak{a} = \mathfrak{o}$ for some j and so $\mathfrak{g}_{(j)} \subset \mathfrak{a}$ and if this is the case then we have $\mathfrak{g}_{(j+1)} = [\mathfrak{g}, \mathfrak{g}_{(j)}] \subset [\mathfrak{g}, \mathfrak{a}] = 0$. (Here we have $[\mathfrak{g}, \mathfrak{a}] = 0$ since $\mathfrak{a} \subset \mathfrak{z}(\mathfrak{g})$.) Thus $\mathfrak{g}$ is nilpotent. ■

Trivially, the center $\mathfrak{z}(\mathfrak{g})$ of a Lie algebra a solvable ideal.

Corollary 17.3 *Let $h : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra homomorphism. If* $\operatorname{img}(\operatorname{ad}) := \operatorname{ad}(\mathfrak{g})$ *is nilpotent then $\mathfrak{g}$ is nilpotent.*

Proof. Just use the fact that $\ker(\operatorname{ad}) = \mathfrak{z}(\mathfrak{g})$. ■

Theorem 17.3 *The sum of any family of solvable ideals in $\mathfrak{g}$ is a solvable ideal. Furthermore, there is a unique maximal solvable ideal which is the sum of all solvable ideals in $\mathfrak{g}$.*

Sketch of proof. The proof is based on the following idea for ideals $\mathfrak{a}$ and $\mathfrak{b}$ and a maximality argument.. If $\mathfrak{a}$ and $\mathfrak{b}$ are solvable then $\mathfrak{a} \cap \mathfrak{b}$ is an ideal in the solvable $\mathfrak{a}$ and so is solvable. It is easy to see that $\mathfrak{a} + \mathfrak{b}$ is an ideal. We have by exercise 17.2 $(\mathfrak{a} + \mathfrak{b})/\mathfrak{b} \cong \mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$. Since $\mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$ is a homomorphic image of $\mathfrak{a}$ we see that $\mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b}) \cong (\mathfrak{a} + \mathfrak{b})/\mathfrak{b}$ is solvable. Thus by our previous result $\mathfrak{a} + \mathfrak{b}$ is solvable. ■

Definition 17.14 *The maximal solvable ideal in $\mathfrak{g}$ whose existence is guaranteed by the last theorem is called the* ***radical*** *of $\mathfrak{g}$ and is denoted $rad(\mathfrak{g})$*

Definition 17.15 *A Lie algebra $\mathfrak{g}$ is called* ***simple*** *if it contains no ideals other than $\{0\}$ and $\mathfrak{g}$. A Lie algebra $\mathfrak{g}$ is called* ***semisimple*** *if it contains no abelian ideals (other than $\{0\}$).*

Theorem 17.4 (Levi decomposition) *Every Lie algebra is the semi-direct sum of its radical and a semisimple Lie algebra.*

Define semi-direct sum before this.

17.5 The Adjoint Representation of a Lie group

Definition 17.16 *Fix an element $g \in G$. The map $C_g : G \to G$ defined by $C_g(x) = gxg^{-1}$ is called conjugation and the tangent map $T_eC_g : \mathfrak{g} \to \mathfrak{g}$ is denoted Ad_g and called the* ***adjoint map****.*

Proposition 17.2 *$C_g : G \to G$ is a Lie group homomorphism.*

The proof is easy.

Proposition 17.3 *The map $C : g \mapsto C_g$ is a Lie group homomorphism $G \to \operatorname{Aut}(G)$.*

The image of the map C inside $\mathrm{Aut}(G)$ is a Lie subgroup called the group of inner automorphisms and is denoted by $\mathrm{Int}(G)$.

Using 8.3 we get the following

Corollary 17.4 $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ *is Lie algebra homomorphism.*

Proposition 17.4 *The map* $\mathrm{Ad} : g \mapsto \mathrm{Ad}_g$ *is a homomorphism* $G \to \mathrm{GL}(\mathfrak{g})$ *which is called the* ***adjoint representation*** *of* G.

Proof. We have

$$\begin{aligned}\mathrm{Ad}(g_1 g_2) &= T_e C_{g_1 g_2} = T_e(C_{g_1} \circ C_{g_2}) \\ &= T_e C_{g_1} \circ T_e C_{g_2} = \mathrm{Ad}_{g_1} \circ \mathrm{Ad}_{g_2}\end{aligned}$$

which show that Ad is a group homomorphism. The smoothness follows from the following lemma applied to the map $C : (g, x) \mapsto C_g(x)$. ■

Lemma 17.6 *Let* $f : M \times N \to N$ *be a smooth map and define the partial map at* $x \in M$ *by* $f_x(y) = f(x, y)$. *Suppose that for every* $x \in M$ *the point* y_0 *is fixed by* f_x*:*

$$f_x(y_0) = y_0 \text{ for all } x.$$

The the map $A_{y_0} : x \mapsto T_{y_0} f_x$ *is a smooth map from* M *to* $\mathrm{GL}(T_{y_0} N)$.

Proof. It suffices to show that A_{y_0} composed with an arbitrary coordinate function from some atlas of charts on $\mathrm{GL}(T_{y_0} N)$ is smooth. But $\mathrm{GL}(T_{y_0} N)$ has an atlas consisting of a single chart which covers it. Namely, choose a basis $v_1, v_2, ..., v_n$ of $T_{y_0} N$ and let $\upsilon^1, \upsilon^2, ..., \upsilon^n$ the dual basis of $T_{y_0}^* N$, then $\chi_j^i : A \mapsto \upsilon^i(A v_j)$ is a typical coordinate function. Now we compose;

$$\begin{aligned}\chi_j^i \circ A_{y_0}(x) &= \upsilon^i(A_{y_0}(x) v_j) \\ &= \upsilon^i(T_{y_0} f_x \cdot v_j).\end{aligned}$$

Now it is enough to show that $T_{y_0} f_x \cdot v_j$ is smooth in x. But this is just the composition the smooth maps $M \to TM \times TN \cong T(M \times N) \to T(N)$ given by

$$\begin{aligned}x \mapsto ((x, 0), (y_0, v_j)) \mapsto (\partial_1 f)(x, y_0) \cdot 0 + (\partial_2 f)(x, y_0) \cdot v_j \\ = T_{y_0} f_x \cdot v_j.\end{aligned}$$

(The reader might wish to review the discussion leading up to lemma 3.4). ■

Lemma 17.7 *Let* $v \in \mathfrak{g}$. *Then* $L^v(x) = R^{\mathrm{Ad}(x)v}$.

Proof. $L^v(x) = T_e(L_x) \cdot v = T(R_x) T(R_{x^{-1}}) T_e(L_x) \cdot v = T(R_x) T(R_{x^{-1}} \circ L_x) \cdot v = R^{\mathrm{Ad}(x)v}$. ■

We have already defined the group $\mathrm{Aut}(G)$ and the subgroup $\mathrm{Int}(G)$. We have also defined $\mathrm{Aut}(\mathfrak{g})$ which has the subgroup $Int(\mathfrak{g}) := Ad(G)$.

We now go one step further and take the differential of Ad.

Definition 17.17 *For a Lie group G with Lie algebra $\mathfrak{g}$ define the **adjoint representation of** $\mathfrak{g}$, a map* $\mathrm{ad} : \mathfrak{g} \to gl(\mathfrak{g})$ *by*

$$\mathrm{ad} = T_e \,\mathrm{Ad}$$

The following proposition shows that the current definition of ad agrees with that given previously for abstract Lie algebras:

Proposition 17.5 $\mathrm{ad}(v)w = [v, w]$ *for all* $v, w \in \mathfrak{g}$.

Proof. Let $v^1, ..., v^n$ be a basis for $\mathfrak{g}$ so that $\mathrm{Ad}(x)w = \sum a_i(x)v^i$ for some functions a_i. Then we have

$$\begin{aligned} \mathrm{ad}(v)w &= T_e(\mathrm{Ad}()w)v \\ &= d(\sum a_i()v^i)v \\ &= \sum (da_i|_e \, v)v^i \\ &= \sum (L^v a_i)(e)v^i \end{aligned}$$

On the other hand, by lemma 17.7

$$\begin{aligned} L^w(x) = R^{\mathrm{Ad}(x)w} &= R(\sum a_i(x)v^i) \\ &= \sum a_i(x)R^{v^i}(x) \end{aligned}$$

Then we have

$$[L^v, L^w] = [L^v, \sum a_i()R^{v^i}()] = 0 + \sum L^v(a_i)R^{v^i}.$$

Finally, we have

$$\begin{aligned} &[w, v] = [L^w, L^v](e) \\ = \sum L^v(a_i)(e)R^{v^i}(e) &= \sum L^v(a_i)(e)v^i \\ &\mathrm{ad}(v)w. \end{aligned}$$

■

The map $\mathrm{ad} : \mathfrak{g} \to \mathrm{gl}(\mathfrak{g}) = End(T_eG)$ is given as the tangent map at the identity of Ad which is a Lie algebra homomorphism. Thus by 8.3 we have the following

Proposition 17.6 $\mathrm{ad} : \mathfrak{g} \to \mathrm{gl}(\mathfrak{g})$ *is a Lie algebra homomorphism.*

Proof. This follows from our study of abstract Lie algebras and proposition 17.5. ■

Lets look at what this means. Recall that the Lie bracket for $\mathrm{gl}(\mathfrak{g})$ is just $A \circ B - B \circ A$. Thus we have

$$\mathrm{ad}([v, w]) = [\mathrm{ad}(v), \mathrm{ad}(w)] = \mathrm{ad}(v) \circ \mathrm{ad}(w) - \mathrm{ad}(w) \circ \mathrm{ad}(v)$$

which when applied to a third vector z gives

$$[[v,w],z] = [v,[w,z]] - [w,[v,z]]$$

which is just a version of the Jacobi identity. Also notice that using the antisymmetry of the bracket we get

$$[z,[v,w]] = [w,[z,v]] + [v,[z,w]]$$

which is in turn the same as

$$\mathrm{ad}(z)([v,w]) = [\mathrm{ad}(z)v, w] + [v, \mathrm{ad}(z)w]$$

so $\mathrm{ad}(z)$ is a derivation of the Lie algebra $\mathfrak{g}$ as explained before.

Proposition 17.7 *The Lie algebra $Der(\mathfrak{g})$ of all derivation of $\mathfrak{g}$ is the Lie algebra of the group of automorphisms $\mathrm{Aut}(\mathfrak{g})$. The image $\mathrm{ad}(\mathfrak{g}) \subset Der(\mathfrak{g})$ is the Lie algebra of the set of all inner automorphisms $\mathrm{Int}(\mathfrak{g})$.*

$$\begin{array}{ccc} \mathrm{ad}(\mathfrak{g}) & \subset & \mathrm{Der}(\mathfrak{g}) \\ \downarrow & & \downarrow \\ \mathrm{Int}(\mathfrak{g}) & \subset & \mathrm{Aut}(\mathfrak{g}) \end{array}$$

Let $\mu : G \times G \to G$ be the multiplication map. Recall that the tangent space $T_{(g,h)}(G \times G)$ is identified with $T_gG \times T_hG$. Under this identification we have

$$T_{(g,h)}\mu(v,w) = T_hL_gw + T_gR_hv$$

where $v \in T_gG$ and $w \in T_hG$. The following diagrams exhibit the relations:

$$\begin{array}{ccccc} & & G \times G, (g,h) & & \\ & \overset{pr_1}{\swarrow} & \updownarrow & \overset{pr_2}{\searrow} & \\ G, g & \to & G \times G, (g,h) & \leftarrow & G, h \\ & \underset{R_h}{\searrow} & \downarrow \mu & \underset{L_g}{\swarrow} & \\ & & G, gh & & \end{array}$$

The horizontal maps are the insertions $g \mapsto (g,h)$ and $h \mapsto (g,h)$. Applying the tangent functor to the last diagram gives.

$$\begin{array}{ccccc} Tpr_1 & & T_{(g,h)}(G \times G) & & Tpr_2 \\ & \swarrow & \updownarrow & \searrow & \\ T_gG & \to & T_gG \times T_hG & \leftarrow & T_hG \\ & \searrow & \downarrow T\mu & \swarrow & \\ T_gR_h & & T_{gh}G & & T_hL_g \end{array}$$

In lieu of a prove we ask the reader to examine the diagrams and try to construct a proof on that basis.

We have another pair of diagrams to consider. Let $\nu : G \to G$ be the inversion map $\nu : g \mapsto g^{-1}$. We have the following commutative diagrams:

$$
\begin{array}{ccccc}
R_{g^{-1}} & & G, g & & L_{g^{-1}} \\
 & \swarrow & & \searrow & \\
G, e & & \downarrow \nu & & G, e \\
 & \searrow & & \swarrow & \\
L_{g^{-1}} & & G, g^{-1} & & R_{g^{-1}}
\end{array}
$$

Applying the tangent functor we get

$$
\begin{array}{ccccc}
TR_{g^{-1}} & & T_gG & & TL_{g^{-1}} \\
 & \swarrow & & \searrow & \\
T_eG & & \downarrow T\nu & & T_eG \\
 & \searrow & & \swarrow & \\
TL_{g^{-1}} & & T_{g^{-1}}G & & TR_{g^{-1}}
\end{array}
.
$$

The result we wish to express here is that $T_g\nu = TL_{g^{-1}} \circ TR_{g^{-1}} = TR_{g^{-1}} \circ TL_{g^{-1}}$. Again the diagrams more or less give the proof which we leave to the reader.

Chapter 18

Group Actions and Homogenous Spaces

Here we set out our conventions regarding (right and left) group actions and the notion of equivariance. There is plenty of room for confusion just from the issues of right as opposed to left if one doesn't make a few observations and set down the conventions carefully from the start. We will make the usual choices but we will note how these usual choices lead to annoyances like the mismatch of homomorphisms with anti-homomorphisms in proposition 18.1 below.

18.1 Our Choices

1. We have define the Lie derivative by use of the contravariant functor $f \mapsto f^*$ so that $L_X Y = \frac{d}{dt}\big|_0 \left(Fl_t^X\right)^* Y$. Notice that we are implicitly using a right action of $\mathrm{Diff}(M)$ on $\mathfrak{X}(M)$[1]. Namely, $Y \mapsto f^*Y$.

2. We have chosen to make the bracket of vector fields be defined so that $[X,Y] = XY - YX$ rather than by $YX - XY$. This makes it true that $L_X Y = [X,Y]$ so the first choice seems to influence this second choice.

3. We have chosen to define the bracket in a Lie algebra $\mathfrak{g}$ of a Lie group G to be given by using the identifying linear map $\mathfrak{g} = T_e G \to \mathfrak{X}^L(M)$ where $\mathfrak{X}^L(M)$ is **left** invariant vector fields. What if we had used **right** invariant vector fields? Then we would have $[X_e, Y_e]_{new} = [X', Y']_e$ where $X'_g = TR_g \cdot X_e$ is the right invariant vector field:

$$\begin{aligned} R_h^* X'(g) &= TR_h^{-1} X'(gh) = TR_h^{-1} TR_{gh} \cdot X_e \\ &= TR_h^{-1} \circ T(R_h \circ R_g) \cdot X_e = TR_g \cdot X_e \\ &= X'(g) \end{aligned}$$

[1] At least when X is complete since otherwise Fl_t^X is only a diffeomorphism on relatively compact set open sets and even then only for small enough t).

But notice that Now on the other hand, consider the inversion map $\nu : G \to G$. We have $v \circ R_{g^{-1}} = L_g \circ v$ and also $T\nu = -\operatorname{id}$ at T_eG so

$$\begin{aligned}(\nu^* X')(g) &= T\nu \cdot X'(g^{-1}) = T\nu \cdot TR_{g^{-1}} \cdot X_e \\ &= T(L_g \circ v)X_e = TL_g Tv \cdot X_e \\ &= -TL_g X_e = -X(g)\end{aligned}$$

thus $\nu^*[X', Y'] = [\nu^* X', \nu^* Y'] = [-X, -Y] = [X, Y]$. Now at e we have $(\nu^*[X', Y'])(e) = Tv \circ [X', Y'] \circ \nu(e) = -[X', Y']_e$. So we have $[X, Y]_e = -[X', Y']_e$.

So this choice is different by a sign also.

The source of the problem may just be conventions but it is interesting to note that if we consider $\operatorname{Diff}(M)$ as an infinite dimensional Lie group then the vector fields of that manifold would be maps $\overleftrightarrow{X} : \operatorname{Diff}(M) \to \mathfrak{X}(M)$ such $\overleftrightarrow{X}(\phi)$ is a vector field in $\mathfrak{X}(M)$ such that $Fl_0^{\overleftrightarrow{X}(\phi)} = \phi$. In other words, a field for every diffeomorphism, a "field of fields" so to speak. Then in order to get the usual bracket in $\mathfrak{X}(M)$ we would have to use right invariant (fields of) fields (instead of the conventional left invariant choice) and evaluate them at the identity element of $\operatorname{Diff}(M)$ to get something in $T_{\mathrm{id}} \operatorname{Diff}(M) = \mathfrak{X}(M)$. This makes one wonder if right invariant vector fields would have been a better convention to start with. Indeed some authors do make that convention.

18.1.1 Left actions

Definition 18.1 *A left action of a Lie group G on a manifold M is a smooth map $\lambda : G \times M \to M$ such that $\lambda(g_1, \lambda(g_2, m)) = \lambda(g_1 g_2, m))$ for all $g_1, g_2 \in G$. Define the partial map $\lambda_g : M \to M$ by $\lambda_g(m) = \lambda(g, m)$ and then the requirement is that $\widetilde{\lambda} : g \mapsto \lambda_g$ is a group homomorphism $G \to \operatorname{Diff}(M)$. We often write $\lambda(g, m)$ as $g \cdot m$.*

Definition 18.2 *For a left group action as above, we have for every $v \in \mathfrak{g}$ we define a vector field $v^\lambda \in \mathfrak{X}(M)$ defined by*

$$v^\lambda(m) = \left.\frac{d}{dt}\right|_{t=0} \exp(tv) \cdot m$$

which is called the ***fundamental vector field*** *associated with the action λ.*

Notice that $v^\lambda(m) = T\lambda_{(e,m)} \cdot (v, 0)$.

Proposition 18.1 *Given left action $\lambda : G \times M \to M$ of a Lie group G on a manifold M, the map $\widetilde{\lambda} : g \mapsto \lambda_g$ is a group homomorphism $G \to \operatorname{Diff}(M)$ by definition.* ***Despite this, the map $X \mapsto X^\lambda$ is a Lie algebra anti-homomorphism*** *$\mathfrak{g} \to \mathfrak{X}(M)$:*

$$[v, w]^\lambda = -[v^\lambda, w^\lambda]_{\mathfrak{X}(M)}$$

which implies that the bracket for the Lie algebra $\mathfrak{diff}(M)$ *of* $\mathrm{Diff}(M)$ *(as an infinite dimensional Lie group) is in fact* $[X,Y]_{\mathfrak{diff}(M)} := -[X,Y]_{\mathfrak{X}(M)}$.

Proposition 18.2 *If G acts on itself from the* left *by multiplication* $L : G \times G \to G$ *then the fundamental vector fields are the* right *invariant vector fields!*

18.1.2 Right actions

Definition 18.3 *A right action of a Lie group G on a manifold M is a smooth map* $\rho : M \times G \to M$ *such that* $\rho(\rho(m, g_2), g_1) = \rho(m, g_2 g_1))$ *for all* $g_1, g_2 \in G$. *Define the partial map* $\rho^g : M \to M$ *by* $\rho^g(m) = \rho(m, g)$ *and then the requirement is that* $\widetilde{\rho} : g \mapsto \rho^g$ *is a group* ***anti-homomorphism*** $G \to \mathrm{Diff}(M)$. *We often write* $\rho(m, g)$ *as* $m \cdot g$

Definition 18.4 *For a right group action as above, we have for every* $v \in \mathfrak{g}$ *a vector field* $v^\rho \in \mathfrak{X}(M)$ *defined by*

$$v^\rho(m) = \left.\frac{d}{dt}\right|_{t=0} m \cdot \exp(tv)$$

which is called the ***fundamental vector field*** *associated with the right action* ρ.

Notice that $v^\rho(m) = T\rho_{(m,e)} \cdot (0, v)$.

Proposition 18.3 *Given right action* $\rho : M \times G \to M$ *of a Lie group G on a manifold M, the map* $\widetilde{\rho} : g \mapsto \rho^g$ *is a group anti-homomorphism* $G \to \mathrm{Diff}(M)$ *by definition.* ***However, the map*** $X \mapsto X^\lambda$ ***is a true Lie algebra homomorphism*** $\mathfrak{g} \to \mathfrak{X}(M)$*:*

$$[v, w]^\rho = [v^\rho, w^\rho]_{\mathfrak{X}(M)}$$

this disagreement again implies that the Lie algebra $\mathfrak{diff}(M)$ *of* $\mathrm{Diff}(M)$ *(as an infinite dimensional Lie group) is in fact* $\mathfrak{X}(M)$*, but with the bracket* $[X,Y]_{\mathfrak{diff}(M)} := -[X,Y]_{\mathfrak{X}(M)}$.

Proposition 18.4 *If G acts on itself from the* right *by multiplication* $L : G \times G \to G$ *then the fundamental vector fields are the* left *invariant vector fields* $\mathfrak{X}_L(G)$*!*

Proof: Exercise.

18.1.3 Equivariance

Definition 18.5 *Given two left actions* $\lambda_1 : G \times M \to M$ *and* $\lambda_2 : G \times S \to S$ *we say that a map* $f : M \to N$ *is (left)* ***equivariant*** *(with respect to these actions) if*

$$f(g \cdot s) = g \cdot f(s)$$

$$i.e.$$

$$f(\lambda_1(g, s)) = \lambda_2(g, f(s))$$

with a similar definition for right actions.

Notice that if $\lambda : G \times M \to M$ is a left action then we have an **associated right action** $\lambda^{-1} : M \times G \to M$ given by

$$\lambda^{-1}(p, g) = \lambda(g^{-1}, p).$$

Similarly, to a right action $\rho : M \times G \to M$ there is an associated left action

$$\rho^{-1}(g, p) = \rho(p, g^{-1})$$

and then we make the follow conventions concerning equivariance when mixing right with left.

Definition 18.6 *Is is often the case that we have a right action on a manifold P (such as a principle bundle) and a left action on a manifold S. Then equivariance is defined by converting the right action to its associated left action. Thus we have the requirement*

$$f(s \cdot g^{-1}) = g \cdot f(s)$$

or we might do the reverse and define equivariance by

$$f(s \cdot g) = g^{-1} \cdot f(s)$$

18.1.4 The action of $\mathrm{Diff}(M)$ and map-related vector fields.

Given a diffeomorphism $\Phi : M \to N$ define $\Phi_\star : \Gamma(M, TM) \to \Gamma(N, TN)$ by

$$\Phi_* X = T\Phi \circ X \circ \Phi^{-1}$$

and $\Phi^* : \Gamma(M, TN) \to \Gamma(M, TM)$ by

$$\Phi^* X = T\Phi^{-1} \circ X \circ \Phi$$

If $M = N$, this gives a right and left pair of actions of the diffeomorphism group $\mathrm{Diff}(M)$ on the space of vector fields $\mathfrak{X}(M) = \Gamma(M, TM)$.

$$\begin{aligned} \mathrm{Diff}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (\Phi, X) &\mapsto \Phi_* X \end{aligned}$$

and

$$\begin{aligned} \mathfrak{X}(M) \times \mathrm{Diff}(M) &\to \mathfrak{X}(M) \\ (X, \Phi) &\mapsto \Phi^* X \end{aligned}$$

18.1.5 Lie derivative for equivariant bundles.

Definition 18.7 *An equivariant left action for a bundle $E \to M$ is a pair of actions $\gamma^E : G \times E \to E$ and $: G \times M \to M$ such that the diagram below commutes*

$$\begin{array}{cccc} \gamma^E : & G \times E & \to & E \\ & \downarrow & & \downarrow \\ \gamma : & G \times M & \to & M \end{array}$$

In this case we can define an action on the sections $\Gamma(E)$ via

$$\gamma_g^* \mathbf{s} = (\gamma^E)^{-1} \circ \mathbf{s} \circ \gamma_g$$

and then we get a Lie derivative for $\mathbf{X} \in LG$

$$L_{\mathbf{X}}(\mathbf{s}) = \left.\frac{d}{dt}\right|_0 \gamma^*_{\exp t\mathbf{X}} \mathbf{s}$$

18.2 Homogeneous Spaces.

Definition 18.8 *The **orbit** of $x \in M$ under a right action by G is denoted $x \cdot G$ or xG and the set of orbits M/G partition M into equivalence classes. For left actions we write $G\cdot\, x$ and $G \setminus M$ for the orbit space.*

Example 18.1 *If H is a closed subgroup of G then H acts on G from the right by right multiplication.. The space of orbits G/H of this* right *is just the set of **left cosets**.*

Definition 18.9 *A left (resp. right) action is said to be **effective** if $g \cdot p = x$ (resp. $x \cdot g = x$) for every $x \in M$ implies that $g = e$ and is said to be **free** if $g \cdot x = x$ (resp. $x \cdot g = x$) for even one $x \in M$ implies that $g = e$.*

Definition 18.10 *A left (resp. right) action is said to be a **transitive action** if there is only one orbit in the space of orbit which. This single orbit would have to be the M. So in other words, given pair $x, y \in M$, there is a $g \in G$ with $g \cdot x = y$ (resp. $x \cdot g = y$).*

Theorem 18.1 *Let $\lambda : G \times M \to M$ be a left action and fix $x_0 \in M$. Let $H = H_{x_0}$ be the **isotropy subgroup** of x_0 defined by*

$$H = \{g \in G : g \cdot x_0 = x_0\}.$$

Then we have a natural bijection

$$G \cdot x_0 \cong G/H$$

given by $g \cdot x_0 \mapsto gH$. In particular, if the action is transitive then $G/H \cong M$ and x_0 maps to H.

Let us denote the projection onto cosets by π and also write $r^{x_0} : g \longmapsto gx_0$. Then we have the following equivalence of maps

$$\begin{array}{ccc} G & = & G \\ \pi \downarrow & & \downarrow r^{x_0} \\ G/H & \cong & M \end{array}$$

In the transitive action situation from the last theorem we may as well assume that $M = G/H$ and then we have the literal equality $r^{x_0} = \pi$. In this case the left action is just $l_h : gH \mapsto hgH$ or $gx \mapsto hx$. Now H also acts on then we get an action $\tau : H \times G/H \to G/H$ given by $\tau_h : gH \mapsto hgH$ or for getting the coset structure $\tau_h : h \mapsto hx$. Thus τ is just the restriction $l : G \times G/H \to G/H$ to $H \times G/H \subset G \times G/H$ and we call this map the translation map.

$$\begin{array}{ccc} G \times G/H & & \\ \shortparallel & \xrightarrow{l} & G/H = M \\ G \times M & & \\ \cup & & \shortparallel \\ H \times G/H & & \\ \shortparallel & \xrightarrow{\tau} & G/H = M \\ H \times M & & \end{array}.$$

For each $h \in H$ the map $\tau_h : M \to M$ fixes the point x_0 and so the differential $T_{x_0}\tau_h$ maps $T_{x_0}M$ onto itself. Let us abbreviate[2] by writing $d\tau_h := T_{x_0}\tau_h$. So now we have a representation $\rho : H \to Gl(T_{x_0}M)$ given by $h \mapsto d\tau_h$. This representation is called the linear isotropy and the group $\rho(H) \subset Gl(T_{x_0}M)$ is called the linear isotropy subgroup. On the other hand we for each $h \in H$ have the action $C_h : G \to G$ given by $g \longmapsto hgh^{-1}$ which fixes H and whose derivative is $Ad_h : \mathfrak{g} \to \mathfrak{g}$. It is easy to see that this map descends to a map $\widetilde{Ad_h} : \mathfrak{g}/\mathfrak{h} \to \mathfrak{g}/\mathfrak{h}$. We are going to show that there is a natural isomorphism $T_{x_0}M \cong \mathfrak{g}/\mathfrak{h}$ such that for each $h \in H$ the following diagram commutes:

$$\begin{array}{cccc} \widetilde{Ad_h}: & \mathfrak{g}/\mathfrak{h} \to & \mathfrak{g}/\mathfrak{h} \\ & \downarrow & \downarrow \\ d\tau_h: & T_{x_0}M \to & T_{x_0}M \end{array} \tag{18.1}$$

One way to state the meaning of this result is to say that $h \mapsto \widetilde{Ad_h}$ is representation of H on the vector space $\mathfrak{g}/\mathfrak{h}$ which is equivalent to the linear isotropy representation. The isomorphism $T_{x_0}M \cong \mathfrak{g}/\mathfrak{h}$ is given in the following way: Let $\xi \in \mathfrak{g}$ and consider $T_e\pi(\xi) \in T_{x_0}M$. If $\varsigma \in \mathfrak{h}$ then $T_e\pi(\xi+\varsigma) = T_e\pi(\xi)+T_e\pi(\varsigma) = T_e\pi(\xi)$ and so $\xi \mapsto T_e\pi(\xi)$ induces a map on $\mathfrak{g}/\mathfrak{h}$. Now if $T_e\pi(\xi) = 0 \in T_{x_0}M$ then as you are asked to show in exercise 18.1 below $\xi \in \mathfrak{h}$ which in turn means that the induces map $\mathfrak{g}/\mathfrak{h} \to T_{x_0}M$ has a trivial kernel. As usual this implies that the map is in fact an isomorphism since $\dim(\mathfrak{g}/\mathfrak{h}) = \dim(T_{x_0}M)$. Let us now see why the diagram 18.2 commutes. Let us take a the scenic root to the conclusion since it allows us to see the big picture a bit better. First the following diagram clearly commutes:

$$\begin{array}{ccc} \exp t\xi & \xrightarrow{C_h} & h(\exp t\xi)h^{-1} \\ \pi \downarrow & & \pi \downarrow \\ (\exp t\xi)H & \xrightarrow{\tau_h} & h(\exp t\xi)H \end{array}$$

[2]This notation for the tangent map is quite common especially in this type of situation where we have a "canonical space" such as $\mathfrak{g}$ or in this case $T_{x_0}M$.

which under the identification $M = G/H$ is just

$$\begin{array}{ccc} \exp t\xi & \rightarrow & h(\exp t\xi)h^{-1} \\ \pi\downarrow & & \pi\downarrow \\ (\exp t\xi)x_0 & \rightarrow & h(\exp t\xi)x_0 \end{array}.$$

Applying the tangent functor (looking at the differential) we get the commutative diagram

$$\begin{array}{ccc} \xi & \rightarrow & Ad_h\xi \\ \downarrow & & \downarrow \\ T_e\pi(\xi) & \stackrel{d\tau_h}{\rightarrow} & T_e\pi(Ad_h\xi) \end{array}$$

and in turn

$$\begin{array}{ccc} [\xi] & \mapsto & \widetilde{Ad_h}([\xi]) \\ \downarrow & & \downarrow \\ T_e\pi(\xi) & \mapsto & T_e\pi(Ad_h\xi) \end{array}.$$

This latter diagram is in fact the element by element version of 18.2.

Exercise 18.1 *Show that $T_e\pi(\xi) \in T_{x_0}M$ implies that $\xi \in \mathfrak{h}$.*

Chapter 19

Fiber Bundles and Connections

Halmos, Paul R.

Don't just read it; fight it! Ask your own questions, look for your own examples, discover your own proofs. Is the hypothesis necessary? Is the converse true? What happens in the classical special case? What about the degenerate cases? Where does the proof use the hypothesis?

Halmos, Paul R.

I Want to be a Mathematician, Washington: MAA Spectrum, 1985.

19.1 Definitions

A (C^r) **fiber bundle** is a quadruple (π, E, M, F) where $\pi : E \to M$ is a smooth C^r-submersion such that for every $p \in M$ there is an open set U containing p with a C^r-isomorphism $\phi = (\pi, \Phi) : \pi^{-1}(U) \to U \times F$. Let us denote the **fiber at** p by $E_p = \pi^{-1}(p)$. It follows that for each $p \in U$ the map $\Phi|_{E_p} : E_p \to F$ is a C^r-diffeomorphism Given two such trivializations $(\pi, \Phi_\alpha) : \pi^{-1}(U_\alpha) \to U_\alpha \times F$ and $(\pi, \Phi_\beta) : \pi^{-1}(U_\beta) \to U_\beta \times F$ then there is a family of diffeomorphisms $\Phi_{\alpha\beta}|_p : E_p \to E_p$ where $p \in U_\alpha \cap U_\beta$ and so for each α, β we have a map $U_\alpha \cap U_\beta \to \mathrm{Diff}(F)$ defined by $p \mapsto \Phi_{\alpha\beta}(p) = \Phi_{\alpha\beta}|_p$. These are called **transition maps** or **transition functions**.

Remark 19.1 *Recall that a group action $\rho : G \times F \to F$ is equivalently thought of as a representation $\overline{\rho} : G \to \mathrm{Diff}(F)$ given by $\overline{\rho}(g)(f) = \rho(g, f)$. We will forgo the separate notation $\overline{\rho}$ and simple write ρ for the action and the corresponding representation.*

Returning to our discussion of fiber bundles, suppose that there is a Lie group action $\rho : G \times F \to F$ such that for each α, β we have

$$\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p), f)$$

for some smooth map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ then we have presented (π, E, M, F) as G **bundle under the representation** ρ. We also say that the transition functions **live in** G (via ρ). In most but not all cases the representation ρ will be faithful, i.e. the action will be **effective** and so G can be considered as a subgroup of $\mathrm{Diff}(F)$. In this case we say simply that (π, E, M, F) is a G **bundle** and that the transition functions **live in** G. It is common to call G the **structure group** but since the action in question may not be effective we should really refer to the **structure group representation (or action)** ρ.

A fiber bundle is determined if we are given an open cover $\mathcal{U} = \{U_\alpha\}$ and maps $\Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{diff}(F)$ such that for all α, β, γ

$$\begin{aligned} \Phi_{\alpha\alpha}(p) &= \mathrm{id}\,\text{for } p \in U_\alpha \\ \Phi_{\alpha\beta}(p) &= \Phi^{-1}_{\beta\alpha}(p) \text{ for } p \in U_\alpha \cap U_\beta \\ \Phi_{\alpha\beta}(p) \circ \Phi_{\beta\gamma}(p) \circ \Phi_{\gamma\alpha}(p) &= \text{ id for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned}$$

If we want a G bundle under a representation ρ then we further require that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p))(f)$ as above and that the maps $g_{\alpha\beta}$ themselves satisfy the cocycle condition:

$$\begin{aligned} g_{\alpha\alpha}(p) &= \text{id for } p \in U_\alpha \\ g_{\alpha\beta}(p) &= g^{-1}_{\beta\alpha}(p) \text{ for } p \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(p) \circ g_{\beta\gamma}(p) \circ g_{\gamma\alpha}(p) &= \text{id for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned} \tag{19.1}$$

We shall also call the maps $g_{\alpha\beta}$ **transition functions** or **transition maps**. Notice that if ρ is effective the last condition will be automatic. The family $\{U_\alpha\}$ together with the maps $\Phi_{\alpha\beta}$ form a **cocycle** and we can construct a bundle by taking the disjoint union $\bigsqcup(U_\alpha \times F) = \bigcup U_\alpha \times F \times \{\alpha\}$ and then taking the equivalence classes under the relation $(p, f, \beta) \backsim (p, \Phi_{\alpha\beta}(p)(f), \alpha)$ so that

$$E = \left(\bigcup U_\alpha \times F \times \{\alpha\}\right) / \backsim$$

and $\pi([p, f, \beta]) = p$.

Let $H \subset G$ be a closed subgroup. Suppose that we can, by throwing out some of the elements of $\{ U_\alpha, \Phi_\alpha\}$ arrange that all of the transition functions live in H. That is, suppose we have that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p), f)$ where $g_{\alpha\beta} : U_\alpha \cap U_\beta \to H$. Then we have a **reduction the structure group (**or **reduction of the structure representation** in case the action needs to be specified).

Next, suppose that we have an surjective Lie group homomorphism $h : \overline{G} \to G$ **.** We then have the lifted representation $\overline{\rho} : \overline{G} \times F \to F$ given by $\overline{\rho}(\overline{g}, f) = \rho(h(\overline{g}), f)$. Under suitable topological conditions we may be able to lift the maps $g_{\alpha\beta}$ to maps $\overline{g}_{\alpha\beta} : U_\alpha \cap U_\beta \to \overline{G}$ and by choosing a subfamily we can even arrange that the $\overline{g}_{\alpha\beta}$ satisfy the cocycle condition. Note that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}, f) = \rho(h(\overline{g}_{\alpha\beta}), f) = \overline{\rho}(\overline{g}_{\alpha\beta}(p), f)$. In this case we say that we have **lifted** the structure representation to $\overline{\rho}$.

Example 19.1 *The simplest class of examples of fiber bundles over a manifold M are the product bundles. These are just Cartesian products $M \times F$ together with the projection map $pr_1 : M \times F \to M$. Here, the structure group can be reduced to the trivial group $\{e\}$ acting as the identity map on F. On the other hand, this bundle can also be prolonged to any Lie group acting on F.*

Example 19.2 *A covering manifold $\pi : \widetilde{M} \to M$ is a G-bundle where G is the permutation group of the fiber which is a discrete (0-dimensional) Lie group.*

Example 19.3 (The Hopf Bundle) *Identify S^1 as the group of complex numbers of unit modulus. Also, we consider the sphere S^3 as it sits in $\mathbb{C}^2$:*

$$S^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}.$$

The group S^1 acts on S^2 by $u \cdot (z_1, z_2) = (uz_1, uz_2)$. Next we get S^2 into the act. We want to realize S^2 as the sphere of radius 1/2 in $\mathbb{R}^3$ and having two coordinate maps coming from stereographic projection from the north and south poles onto copies of $\mathbb{C}$ embedded as planes tangent to the sphere at the two poles. The chart transitions then have the form $w = 1/z$. Thus we may view S^2 as two copies of $\mathbb{C}$, say the z plane $\mathbb{C}_1$ and the w plane $\mathbb{C}_2$ glued together under the identification $\phi : z \mapsto 1/z = w$

$$S^2 = \mathbb{C}_1 \cup_\phi \mathbb{C}_2.$$

With this in mind define a map $\pi : S^3 \subset \mathbb{C}^2 \to S^2$ by

$$\pi(z_1, z_2) = \begin{cases} z_2/z_1 \in \mathbb{C}_2 \subset \mathbb{C}_1 \cup_\phi \mathbb{C}_2 & \textit{if} \quad z_1 \neq 0 \\ z_1/z_2 \in \mathbb{C}_1 \subset \mathbb{C}_1 \cup_\phi \mathbb{C}_2 & \textit{if} \quad z_2 \neq 0 \end{cases}.$$

Note that this gives a well defined map onto S^2.

Claim 19.1 *$u \cdot (z_1, z_2) = u \cdot (w_1, w_2)$ iff $\pi(z_1, z_2) = \pi(w_1, w_2)$.*

Proof. If $u \cdot (z_1, z_2) = u \cdot (w_1, w_2)$ and $z_1 \neq 0$ then $w_1 \neq 0$ and $\pi(w_1, w_2) = w_2/w_1 = uw_2/uw_1 = \pi(w_1, w_2) = \pi(z_1, z_2) = uz_2/uz_1 = z_2/z_1 = \pi(z_1, z_2)$. A similar calculation show applies when $z_2 \neq 0$. On the other hand, if $\pi(w_1, w_2) = \pi(z_1, z_2)$ then by a similar chain of equalities we also easily get that $u \cdot (w_1, w_2) = ... = \pi(w_1, w_2) = \pi(z_1, z_2) = ... = u \cdot (z_1, z_2)$. ∎

Using these facts we see that there is a fiber bundle atlas on $\pi_{Hopf} = \pi : S^3 \to S^2$ given by the following trivializations:

$$\begin{aligned} &\varphi_1 : \pi^{-1}(C_1) \to C_1 \times S^1 \\ &\varphi_1 : (z_1, z_2) = (z_2/z_1, z_1/|z_1|) \end{aligned}$$

and

$$\begin{aligned} &\varphi_2 : \pi^{-1}(C_2) \to C_2 \times S^1 \\ &\varphi_2 : (z_1, z_2) = (z_1/z_2, z_2/|z_2|). \end{aligned}$$

The transition map is
$$(z,u) \mapsto (1/z, \frac{z}{|z|}u)$$
which is of the correct form since $u \mapsto \frac{z}{|z|} \cdot u$ is a circle action. Thus the Hopf bundle is an S^1-bundle with typical fiber S^1 itself. It can be shown that the inverse image of a circle on S^2 by the Hopf projection π_{Hopf} is a torus. Since the sphere S^2 is foliated by circles degenerating at the poles we have a foliation of S^3-{two circles} by tori degenerating to circles at the fiber over the two poles. Since $S^3\backslash\{\text{pole}\}$ is diffeomorphic to $\mathbb{R}^3$ we expect to be able to get a picture of this foliation by tori. In fact, the following picture depicts this foliation.

19.2 Principal and Associated Bundles

An important case for a bundle with structure group G is where the typical fiber is the group itself. In fact we may obtain such a bundle by taking the transition functions $g_{\alpha\beta}$ from any effective G bundle $E \to M$ or just any smooth maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ that form a cocycle with respect to some cover $\{U_\alpha\}$ of M. We let G act on itself by left multiplication and then use the bundle construction method above. Thus if $\{U_\alpha\}$ is the cover of M corresponding to the cocycle $\{g_{\alpha\beta}\}_{\alpha,\beta}$ then we let
$$P = \left(\bigcup U_\alpha \times G \times \{\alpha\}\right) / \backsim$$
where $(p, g, \alpha) \backsim (p, g_{\alpha\beta}(p)g, \beta)$ gives the equivalence relation. In this way we construct what is called the a principal bundle. Notice that for $g \in G$ we have $(p, g_1, \beta) \backsim (p, g_2, \alpha)$ if and only if $(p, g_1g, \beta) \backsim (p, g_2g, \alpha)$ and so there is a well defined right action on any bundle principal bundle. On the other hand there is a more direct way chart free way to define the notion of principal bundle. The advantage of defining a principal bundle without explicit reference to transitions functions is that we may then use the principal bundle to give another definition of a G-bundle that doesn't appeal directly to the notion of transition functions. We will see that every G bundle is given by a choice of a principal G-bundle and an action of G on some manifold F (the typical fiber).

First we define the **trivial principal G bundle over U** to be the trivial bundle $pr_1 : U \times G \to M$ together with the right G action $(U \times G) \times G$ given by
$$(x, g_1)g := (x, g_1g).$$
An automorphism of the G-space $U \times G$ is a bundle map $\delta : U \times G \to U \times G$ such that $\delta(x, g_1g) = \delta(x, g_1)g$ for all $g_1, g \in G$ and all $x \in U$. Now δ must have the form given by $\delta(x, g) = (x, \Delta(x, g))$ and so
$$\Delta(x, g) = \Delta(x, e)g.$$
If we then let the function $x \mapsto \Delta(x, e)$ be denoted by $g_\delta()$ then we have $\delta(x, g) = (x, g_\delta(x)g)$. Thus we obtain the following

Lemma 19.1 *Every automorphism of a trivial principal G bundle over and open set U has the form $\delta : (x, g) \mapsto (x, g_\delta(x)g)$ for some smooth map $g_\delta : U \to G$.*

Definition 19.1 *A* ***principal*** G***-bundle*** *is a fiber bundle $\pi_P : P \to M$ together with a right G action $P \times G \to P$ which is locally equivalent as a right G space to the trivial principal G bundle over M. This means that for each point $x \in M$ there is an open neighborhood U_x and a trivialization*

$$\begin{array}{ccc} \pi_P^{-1}(U_x) & \overset{\phi}{\to} & U_x \times G \\ & \searrow \quad \swarrow & pr_1 \\ & U_x & \end{array}$$

which is G equivariant. Thus we require that $\phi(pg) = \phi(p)g$. We shall call such a trivialization an ***equivariant trivialization****.*

Note that $\phi(pg) = (\pi_P(pg), \Phi(pg))$ while on the other hand $\phi(p)g = (\pi_P(pg), \Phi(p)g)$ so it is necessary and sufficient that $\Phi(p)g = \Phi(pg)$. Now we want to show that this means that the structure representation of $\pi_P : P \to M$ is left multiplication by elements of G. Let ϕ_1, U_1 and ϕ_2, U_2 be two equivariant trivializations such that $U_1 \cap U_2 \neq \emptyset$. On the overlap we have the diagram

$$\begin{array}{ccccc} U_1 \cap U_2 \times G & \overset{\phi_2}{\leftarrow} & \pi_P^{-1}(U_1 \cap U_2) & \overset{\phi_1}{\to} & U_1 \cap U_2 \times G \\ & \searrow & \downarrow & \swarrow & \\ & & U_1 \cap U_2 & & \end{array}$$

The map $\phi_2 \circ \phi_1^{-1}\big|_{U_1 \cap U_2}$ clearly must an G-bundle automorphism of $U_1 \cap U_2 \times G$ and so by 19.1 must have the form $\phi_2 \circ \phi_1^{-1}\big|_{U_1 \cap U_2}(x, g) = (x, \Phi_{12}(x)g)$. We conclude that a principal G-bundle is a G-bundle with typical fiber G as defined in section 19.1. The maps on overlaps such as Φ_{12} are the transition maps. Notice that $\Phi_{12}(x)$ acts on G by left multiplication and so the structure representation is left multiplication.

Proposition 19.1 *If $\pi_P : P \to M$ is a principal G-bundle then the right action $P \times G \to P$ is free and the action restricted to any fiber is transitive.*

Proof. Suppose $p \in P$ and $pg = p$ for some $g \in G$. Let $\pi_P^{-1}(U_x) \overset{\phi}{\to} U_x \times G$ be an (equivariant) trivialization over U_x where U_x contains $\pi_P(p) = x$. Then we have

$$\begin{aligned} \phi(pg) = \phi(p) \quad &\Rightarrow \\ (x, g_0 g) = (x, g_0) \quad &\Rightarrow \\ g_0 g = g_0 \quad & \end{aligned}$$

and so $g = e$.

Now let $P_x = \pi_P^{-1}(x)$ and let $p_1, p_2 \in P_x$. Again choosing an (equivariant) trivialization over U_x as above we have that $\phi(p_i) = (x, g_i)$ and so letting $g :=$

$g_1^{-1}g_2$ we have $\phi(p_1 g) = (x, g_1 g) = (x, g_2) = \phi(p_2)$ and since ϕ is injective $p_1 g = p_2$. ■

The reader should realize that this result is in some sense "obvious" since the upshot is just that the result is true for the trivial principal bundle and then it follows for the general case since a general principal bundle is locally G-bundle isomorphic to a trivial principal bundle.

Remark 19.2 *Some authors define a principal bundle to be fiber bundle with typical fiber G and with a free right action that is transitive on each fiber.*

Our first and possibly most important example of a principal bundle is the frame bundle of a smooth manifold. The structure group is the general linear group $GL(n, \mathbb{R})$. We define a frame at (or above) a point $x \in M$ to be a basis for the tangent space $T_x M$. Let $F_x(M)$ be the set of all frames at x and define

$$F(M) = \bigcup_{x \in M} F_x(M).$$

Also, let $\pi_{F(M)} : F(M) \to M$ be the natural projection map which takes any frame $\mathbf{f}_x \in F(M)$ with $\mathbf{f}_x \in F_x(M)$ to its base x. We first give $F(M)$ a smooth atlas. Let us adopt the convention that $(\mathbf{f}_x)_i := f_i$ is the i-th vector in the basis $\mathbf{f}_x = (f_1, ..., f_n)$. Let $\{U_\alpha, \psi_\alpha\}_{\alpha \in A}$ be an atlas for M. For each chart $U_\alpha, \psi_\alpha = (x^1, ..., x^1)$ on M shall define a chart $FU_\alpha, F\psi_\alpha$ by letting

$$FU_\alpha := \pi_{F(M)}^{-1}(U_\alpha) = \bigcup_{x \in U_\alpha} F_x(M)$$

and

$$F\psi_\alpha(\mathbf{f}_x) := (f_{ij}) \in \mathbb{R}^{n \times n}$$

where f_{ij} is the n^2 numbers such that $(\mathbf{f}_x)_j = \sum f_j^i \frac{\partial}{\partial x^i}\big|_x$. We leave it to the reader to find the change of coordinate maps and see that they are smooth. The right $GL(n, \mathbb{R})$ action on $F(M)$ is given by matrix multiplication $(\mathbf{f}_x, g) \mapsto \mathbf{f}_x\ g$ where we think of $\mathbf{f}_x$as row of basis vectors.

Chapter 20

Analysis on Manifolds

The best way to escape from a problem is to solve it.
-Alan Saporta

20.1 Basics

20.1.1 Star Operator II

The definitions and basic algebraic results concerning the star operator on a scalar product space globalize to the tangent bundle of a Riemannian manifold in a straight forward way.

Definition 20.1 *Let M, g be a semi-Riemannian manifold. Each tangent space is a scalar product space and so on each tangent space T_pM we have a metric volume element vol_p and then the map $p \mapsto \mathrm{vol}_p$ gives a section of $\bigwedge^n T^*M$ called the metric volume element of M, g. Also on each fiber $\bigwedge T_p^*M$ of $\bigwedge T^*M$ we have a star operator $*_p : \bigwedge^k T_p^*M \to \bigwedge^{n-k} T_p^*M$. These induce a bundle map $* : \bigwedge^k T^*M \to \bigwedge^{n-k} T^*M$ and thus a map on sections (i.e. smooth forms) $* : \Omega^k(M) \to \Omega^{n-k}(M)$.*

Definition 20.2 *The star operator is sometimes referred to as the* ***Hodge star operator****.*

Definition 20.3 *Globalizing the scalar product on the Grassmann algebra we get a* ***scalar product bundle*** *$\Omega(M), \langle .,. \rangle$ where for every $\eta, \omega \in \Omega^k(M)$ we have a smooth function $\langle \eta, \omega \rangle$ defined by*

$$p \mapsto \langle \eta(p), \omega(p) \rangle$$

and thus a $C^\infty(M)$-bilinear map $\langle .,. \rangle : \Omega^k(M) \times \Omega^k(M) \to C^\infty(M)$. Declaring forms of differing degree to be orthogonal as before we extend to a C^∞ bilinear map $\langle .,. \rangle : \Omega(M) \times \Omega(M) \to C^\infty(M)$.

Theorem 20.1 *For any forms $\eta, \omega \in \Omega(M)$ we have* $\langle \eta, \omega \rangle \operatorname{vol} = \eta \wedge *\omega$

Now let M, g be a Riemannian manifold so that $\mathrm{g} = \langle ., . \rangle$ is positive definite. We can then define a Hilbert space of square integrable differential forms:

Definition 20.4 *Let an inner product be defined on $\Omega_c(M)$, the elements of $\Omega(M)$ with compact support, by*

$$(\eta, \omega) := \int_M \eta \wedge *\omega = \int_M \langle \eta, \omega \rangle \operatorname{vol}$$

and let $L^2(\Omega(M))$ denote the L^2 completion of $\Omega_c(M)$ with respect to this inner product.

20.1.2 Divergence, Gradient, Curl

20.2 The Laplace Operator

The exterior derivative operator $d : \Omega^k(M) \to \Omega^{k+1}(M)$ has a formal adjoint $\delta : \Omega^{k+1}(M) \to \Omega^k(M)$ defined by the requirement that for all $\alpha, \beta \in \Omega_c^k(M)$ with compact support we have

$$(d\alpha, \beta) = (\alpha, \delta\beta).$$

On a Riemannian manifold M the Laplacian of a function $f \in C(M)$ is given in coordinates by

$$\Delta f = -\frac{1}{\sqrt{g}} \sum_{j,k} \partial_j (g^{jk} \sqrt{g} \partial_k f)$$

where g^{ij} is the inverse of g_{ij} the metric tensor and g is the determinant of the matrix $G = (g_{ij})$. We can obtain a coordinate free definition as follows. First we recall that the divergence of a vector field $X \in \mathfrak{X}(M)$ is given at $p \in M$ by the trace of the map $\nabla X|_{T_pM}$. Here $\nabla X|_{T_pM}$ is the map

$$v \mapsto \nabla_v X.$$

Thus

$$div(X)(p) := tr(\nabla X|_{T_pM}).$$

Then we have

$$\Delta f := div(grad(f))$$

Eigenvalue problem: For a given compact Riemannian manifold M one is interested in finding all $\lambda \in R$ such that there exists a function $f \neq 0$ in specified subspace $S \subset L^2(M)$ satisfying $\Delta f = \lambda f$ together with certain boundary conditions in case $\partial M \neq 0$.

The reader may be a little annoyed that we have not specified S more clearly. The reason for this is twofold. First, the theory will work even for relatively compact open submanifolds with rather unruly topological boundary and so regularity at the boundary becomes and issue. In general, our choice of S will be influenced by boundary conditions. Second, even though it may appear that S must consist of C^2 functions, we may also seek “weak solutions” by extending Δ in some way. In fact, Δ is essentially self adjoint in the sense that it has a unique extension to a self adjoint unbounded operator in $L^2(M)$ and so eigenvalue problems could be formulated in this functional analytic setting. It turns out that under very general conditions on the form of the boundary conditions, the solutions in this more general setting turn out to be smooth functions. This is the result of the general theory of elliptic regularity.

Definition 20.5 *A boundary operator is a linear map $b : S \to C^0(\partial M)$.*

Using this notion of a boundary operator we can specify boundary conditions as the requirement that the solutions lie in the kernel of the boundary map. In fact, the whole eigenvalue problem can be formulated as the search for λ such that the linear map

$$(\triangle - \lambda) \oplus b : S \to L^2(M) \oplus C^0(\partial M)$$

has a nontrivial kernel. If we find such a λ then this kernel is denoted $E_\lambda \subset L^2(M)$ and by definition $\Delta f = \lambda f$ and $bf = 0$ for all $f \in E_\lambda$. Such a function is called an **eigenfunction** corresponding to the eigenvalue λ. We shall see below that in each case of interest (for compact M) the eigenspaces E_λ will be finite dimensional and the eigenvalues form a sequence of nonnegative numbers increasing without bound. The dimension $\dim(E_\lambda)$ is called the **multiplicity** of λ. We shall present the sequence of eigenvalues in two ways:

1. If we write the sequence so as to include repetitions according to multiplicity then the eigenvalues are written as $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \uparrow \infty$. Thus it is possible, for example, that we might have $\lambda_2 = \lambda_3 = \lambda_4$ if $\dim(E_{\lambda_2}) = 3$.

2. If we wish to list the eigenvalues without repetition then we use an overbar:

$$0 \leq \bar{\lambda}_1 < \bar{\lambda}_2 < \ldots \uparrow \infty$$

The sequence of eigenvalues is sometimes called the **spectrum** of M.

To make thing more precise we divide things up into four cases:

The closed eigenvalue problem: In this case M is a compact Riemannian manifold without boundary the specified subspace of $L^2(M)$ can be taken to be $C^2(M)$. The kernel of the map $\Delta - \lambda : C^2(M) \to C^0(M)$ is the λ eigenspace and denoted by E_λ It consists of eigenfunctions for the eigenvalue λ.

The Dirichlet eigenvalue problem: In this case M is a compact Riemannian manifold without nonempty boundary ∂M. Let $\overset{\circ}{M}$ denote the interior of M. The specified subspace of $L^2(M)$ can be taken to be $C^2(\overset{\circ}{M}) \cap C^0(M)$ and the boundary conditions are $f|\partial M \equiv 0$ (**Dirichlet boundary conditions)** so the appropriate boundary operator is the restriction map $b_D : f \longmapsto f|\partial M$.

The solutions are called **Dirichlet eigenfunctions** and the corresponding sequence of numbers λ for which a nontrivial solution exists is called the **Dirichlet spectrum** of M.

The Neumann eigenvalue problem: In this case M is a compact Riemannian manifold without nonempty boundary ∂M but . The specified subspace of $L^2(M)$ can be taken to be $C^2(\overset{\circ}{M}) \cap C^1(M)$. The problem is to find nontrivial solutions of $\Delta f = \lambda f$ with $f \in C^2(\overset{\circ}{M}) \cap C^0(\partial M)$ which satisfy $\nu f|\,\partial M \equiv 0$ (**Neumann boundary conditions)**.Thus the boundary map here is $b_N : C^1(M) \to C^0(\partial M)$ given by $f \mapsto \nu f|\,\partial M$ where ν is a smooth unit normal vector field defined on ∂M and so the νf is the normal derivative of f . The solutions are called **Neumann eigenfunctions** and the corresponding sequence of numbers λ for which a nontrivial solution exists is called the **Neumann spectrum** of M.

Recall that the completion of $C^k(M)$ (for any $k \geq 0$) with respect to the inner product

$$(f, g) = \int_M fg dV$$

is the Hilbert space $L^2(M)$. The Laplace operator has a natural extension to a self adjoint operator on $L^2(M)$ and a careful reformulation of the above eigenvalue problems in this Hilbert space setting together with the theory of elliptic regularity lead to the following

Theorem 20.2 *1) For each of the above eigenvalue problems the set of eigenvalues (the spectrum) is a sequence of nonnegative numbers which increases without bound:* $0 \leq \bar{\lambda}_1 < \bar{\lambda}_2 < \cdots \uparrow \infty$.

2) Each eigenfunction is a C^∞ *function on* $M = \overset{\circ}{M} \cup \partial M$.

3) Each eigenspace $E_{\bar{\lambda}_i}$ *(or* $E^D_{\bar{\lambda}_i}$ *or* $E^N_{\bar{\lambda}_i}$*) is finite dimensional, that is, each eigenvalue has finite multiplicity.*

4) If $\varphi_{\bar{\lambda}_i,1}, ..., \varphi_{\bar{\lambda}_i,m_i}$ *is an orthonormal basis for the eigenspace* $E_{\bar{\lambda}_i}$ *(or* $E^D_{\bar{\lambda}_i}$ *or* $E^N_{\bar{\lambda}_i}$*) then the set* $B = \cup_i\{\varphi_{\bar{\lambda}_i,1}, ..., \varphi_{\bar{\lambda}_i,m_i}\}$ *is a complete orthonormal set for* $L^2(M)$*. In particular, if we write the spectrum with repetitions by multiplicity,* $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \uparrow \infty$*, then we can reindex this set of functions* B *as* $\{\varphi_1, \varphi_2, \varphi_3, ...\}$ *to obtain an ordered orthonormal basis for* $L^2(M)$ *such that* φ_i *is an eigenfunction for the eigenvalue* λ_i.

The above can be given the following physical interpretation. If we think of M as a vibrating homogeneous membrane then the transverse motion of the membrane is described by a function $f : M \times (0, \infty) \to R$ satisfying

$$\Delta f + \frac{\partial^2 f}{\partial t^2} = 0$$

and if $\partial M \neq \emptyset$ then we could require $f|\partial M \times (0, \infty) = 0$ which means that we are holding the boundary fixed. A similar discussion for the Neumann boundary conditions is also possible and in this case the membrane is free at the boundary.

If we look for the solutions of the form $f(x,t) = \phi(x)T(t)$ then we are led to conclude that ϕ must satisfy $\Delta\phi = \lambda\phi$ for some real number λ with $\phi = 0$ on ∂M. This is the Dirichlet eigenvalue problem discussed above.

Theorem 20.3 *For each of the eigenvalue problems defined above*

Now explicit solutions of the above eigenvalue problems are very difficult to obtain except in the simplest of cases. It is interesting therefore, to see if one can tell something about the eigenvalues from the geometry of the manifold. For instance we may be interested in finding upper and/or lower bounds on the eigenvalues of the manifold in terms of various geometric attributes of the manifold. A famous example of this is the Faber–Krahn inequality which states that if Ω is a regular domain in say $\mathbb{R}^n$ and D is a ball or disk of the same volume then

$$\lambda(\Omega) \geq \lambda(D)$$

where $\lambda(\Omega)$ and $\lambda(D)$ are the lowest nonzero Dirichlet eigenvalues of Ω and D respectively. Now it is of interest to ask whether one can obtain geometric information about the manifold given a degree of knowledge about the eigenvalues. There is a 1966 paper by M. Kac entitled "Can One Hear the Shape of a Drum?" which addresses this question. Kac points out that Weyl's asymptotic formula shows that the sequence of eigenvalues does in fact determine the volume of the manifold. Weyl's formula is

$$(\lambda_k)^{n/2} \sim \left(\frac{(2\pi)^n}{\omega_n}\right)\frac{k}{vol(M)} \quad \text{as} \quad k \longrightarrow \infty$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$ and M is the given compact manifold. In particular,

$$\left(\frac{(2\pi)^n}{\omega_n}\right)\lim_{k\to\infty}\frac{k}{(\lambda_k)^{n/2}} = vol(M).$$

So the volume is indeed determined by the spectrum[1].

20.3 Spectral Geometry

20.4 Hodge Theory

20.5 Dirac Operator

It is often convenient to consider the differential operator $D = i\frac{\partial}{\partial x}$ instead of $\frac{\partial}{\partial x}$ even when one is interested mainly in real valued functions. For one thing

[1] Notice however, one may ask still how far out into the spectrum must one "listen" in order to gain an estimate of $vol(M)$ to a given accuracy.

$D^2 = -\frac{\partial^2}{\partial x^2}$ and so D provides a sort of square root of the positive Euclidean Laplacian $\triangle = -\frac{\partial^2}{\partial x^2}$ in dimension 1. Dirac wanted a similar square root for the wave operator $\Box = \partial_0^2 - \sum_{i=1}^3 \partial_i^2$ (the Laplacian in $\mathbb{R}^4$ for the Minkowski inner metric) and found that an operator of the form $D = \partial_0 - \sum_{i=1}^3 \gamma_i \partial_i$ would do the job if it could be arranged that $\gamma_i\gamma_j + \gamma_i\gamma_j = 2\eta_{ij}$ where

$$(\eta_{ij}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

One way to do this is to allow the γ_i to be matrices.

Now lets consider finding a square root for $\triangle = -\sum_{i=1}^n \partial_i^2$. We accomplish this by an $\mathbb{R}$-linear embedding of $\mathbb{R}^n$ into an $N \times N$ real or complex matrix algebra A by using n linearly independent matrices $\{\gamma_i : i = 1, 2, ..., n\}$ (so called "gamma matrices") and mapping

$$(x^1, ..., x^n) \mapsto x^i\gamma_i \text{ (sum)}.$$

and where $\gamma_1, ..., \gamma_n$ are matrices satisfying the basic condition

$$\gamma_i\gamma_j + \gamma_i\gamma_j = -2\delta_{ij}.$$

We will be able to arrange[2] that $\{1, \gamma_1, ..., \gamma_n\}$ generates an algebra of dimension 2^n and which is spanned as vector space by the identity matrix 1 and all products of the form $\gamma_{i_1} \cdots \gamma_{i_k}$ with $i_1 < i_2 < \cdots < i_k$. Thus we aim to identify $\mathbb{R}^n$ with the linear span of these gamma matrices. Now if we can find matrices with the property that $\gamma_i\gamma_j + \gamma_i\gamma_j = -2\delta_{ij}$ then our "Dirac operator" will be

$$D = \sum_{i=1}^n \gamma_i \partial_i$$

which is now acting on N-tuples of smooth functions.

Now the question arises: What are the differential operators $\partial_i = \frac{\partial}{\partial x^i}$ acting on exactly. The answer is that they act on whatever we take the algebra spanned by the gamma matrices to be acting on. In other words we should have some vector space S which is a module over the algebra spanned by the gamma matrices. Then we take as our "fields" smooth maps $f : \mathbb{R}^n \to S$. Of course since the $\gamma_i \in \mathbb{M}_{N\times N}$ we may always take $S = \mathbb{R}^N$ with the usual action of $\mathbb{M}_{N\times N}$ on $\mathbb{R}^N$. The next example shows that there are other possibilities.

[2] It is possible that gamma matrices might span a space of half the dimension we are interested in. This fact has gone unnoticed in some of the literature. The dimension condition is to assure that we get a universal Clifford algebra.

Example 20.1 *Notice that with* $\frac{\partial}{\partial z} := \frac{\partial}{\partial x} - i\frac{\partial}{\partial y}$ *and* $\frac{\partial}{\partial \bar{z}} := \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}$ *we have*

$$\begin{aligned} &\begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix} \begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}} \\ -\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z} & 0 \end{bmatrix} = \begin{bmatrix} \triangle & 0 \\ 0 & \triangle \end{bmatrix} \\ &= \triangle 1 \end{aligned}$$

where $\triangle = -\sum \partial_i^2$. *On the other hand*

$$\begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{\partial}{\partial x} + \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix} \frac{\partial}{\partial y}.$$

From this we can see that appropriate gamma matrices for this case are $\gamma_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ *and* $\gamma_2 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

Now let E^0 *be the span of* $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ *and* $\gamma_2\gamma_1 = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}$. *Let* E^1 *be the span of* γ_1*and* γ_2. *Refer to* E^0 *and* E^1 *the even and odd parts of* $Span\{1, \gamma_1, \gamma_2, \gamma_2\gamma_1\}$. *Then we have that* $D = \begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix}$ *maps* E^0 *to* E^1 *and writing a typical element of* E^0 *as* $f(x, y) = u(x, y) + \gamma_2\gamma_1 v(x, y)$ *is easy to show that* $Df = 0$ *is equivalent to the Cauchy-Riemann equations.*

The reader should keep this last example in mind as this kind of decomposition into even and odd part will be a general phenomenon below.

20.5.1 Clifford Algebras

A Clifford algebra is the type of algebraic object that allows us to find differential operators which square to give Laplace type operators. The matrix approach described above is in fact quite general but there are other approaches which are more abstract and encourage one to think about a Clifford algebra as something that contains the scalar and the vector space we start with. The idea is similar to that of the complex numbers. We seldom think about complex numbers as "pairs of real numbers" while we are calculating unless push comes to shove. After all, there are other good ways to represent complex numbers; as matrices for example. And yet there is one underlying abstract object called the complex numbers which ironically is quite concrete once one get used to using them. Similarly we encourage the reader to learn to think about abstract Clifford algebras in the same way. Just compute!

Clifford algebras are usually introduced in connection with a quadratic form q on some vector space but in fact we are just as interested in the associated symmetric bilinear form and so in this section we will generically use the same symbol for a quadratic form and the bilinear form obtained by polarization and write both $q(v)$ and $q(v, w)$.

Definition 20.6 *Let* V *be an* n *dimensional vector space over a field* $\mathbb{K}$ *with characteristic not equal to* 2*. Suppose that* q *is a quadratic form on V and let* q *be the associated symmetric bilinear form obtained by polarization. A* ***Clifford algebra based on*** V,q *is an algebra with unity* 1 $Cl(\mathrm{V},q,\mathbb{K})$ *containing V (or an isomorphic image of V) such that the following relations hold:*

$$vw + wv = -2q(v,w)1$$

and such that $Cl(\mathrm{V},q,\mathbb{K})$ *is universal in the following sense: Given any linear map* $L : \mathrm{V} \to A$ *into an associative* $\mathbb{K}-$*algebra with unity* **1** *such that*

$$L(v)L(w) + L(w)L(v) = -2q(v,w)\mathbf{1}$$

then there is a unique extension of L *to an algebra homomorphism* $\bar{L} : Cl(\mathrm{V},q,\mathbb{K}) \to A$*.*

If $e_1, ..., e_n$ is an orthonormal basis for V, q then we must have

$$e_i e_j + e_j e_i = 0 \text{ for } i \neq j$$
$$e_i^2 = -q(e_i) = \pm 1 \text{ or } 0$$

A common choice is the case when q is a nondegenerate inner product on a real vector space. In this case we have a particular realization of the Clifford algebra obtained by introducing a new product into the Grassmann vector space $\wedge$V. The said product is the unique linear extension of the following rule for $v \in \wedge^1\mathrm{V}$ and $w \in \wedge^k\mathrm{V}$:

$$v \cdot w := v \wedge w - v^\flat \lrcorner w$$
$$w \cdot v := (-1)^k (v \wedge w + v^\flat \lrcorner w)$$

We will refer to this as a **geometric algebra** on $\wedge$V and this version of the Clifford algebra will be called the **form presentation of** $Cl(\mathrm{V},q)$. Now once we have a definite inner product on V we have an inner product on V^* and $\mathrm{V} \cong \mathrm{V}^*$. The Clifford algebra on V^* is generated by the following more natural looking formulas

$$\alpha \cdot \beta := \alpha \wedge \beta - (\sharp\alpha) \lrcorner \beta$$
$$\beta \cdot \alpha := (-1)^k (\alpha \wedge \beta + (\sharp\alpha) \lrcorner \beta)$$

for $\alpha \in \wedge^1\mathrm{V}$ and $\beta \in \wedge\mathrm{V}$.

Now we have seen that one can turn $\wedge$V (or $\wedge\mathrm{V}^*$) into a Clifford algebra and we have also seen that one can obtain a Clifford algebra whenever appropriate gamma matrices can be found. A slightly more abstract construction is also common: Denote by $I(q)$ the ideal of the full tensor algebra $T(\mathrm{V})$ generated by elements of the form $x \otimes x - q(x) \cdot 1$. The **Clifford algebra** is (up to isomorphism) given by

$$Cl(\mathrm{V},q,\mathbb{K}) = T(\mathrm{V})/I(q).$$

We can use the canonical injection

$$i : \mathrm{V} \longrightarrow C_K$$

to identify V with its image in $Cl(\mathrm{V},q,\mathbb{K}$. (The map turns out that i is 1–1 onto $i(\mathrm{V})$ and we will just accept this without proof.)

Exercise 20.1 *Use the universal property of $Cl(\mathrm{V},q,\mathbb{K})$ to show that it is unique up to isomorphism.*

Remark 20.1 *Because of the form realization of a Clifford algebra we see that $\wedge\mathrm{V}$ is a $Cl(\mathrm{V},q,\mathbb{R})-$module. But even if we just have some abstract $Cl(\mathrm{V},q,\mathbb{R})$ we can use the universal property to extend the action of V on $\wedge\mathrm{V}$ given by*

$$v \mapsto v \cdot w := v \wedge w - v^\flat \lrcorner w$$

to an action of $Cl(\mathrm{V},q,\mathbb{K})$ on $\wedge\mathrm{V}$ thus making $\wedge\mathrm{V}$ a $Cl(\mathrm{V},q,\mathbb{R})-$module.

Definition 20.7 *Let $\mathbb{R}^n_{(r,s)}$ be the real vector space $\mathbb{R}^n$ with the inner product of signature (r,s) given by*

$$\langle x, y\rangle := \sum_{i=1}^{r} x_i y_i - \sum_{i=r+1}^{r+s=n} x_i y_i.$$

The Clifford algebra formed from this inner product space is denoted $Cl_{r,s}$. In the special case of $(p,q) = (n,0)$ we write Cl_n.

Definition 20.8 *Let $\mathbb{C}^n$ be the complex vector space of n-tuples of complex numbers together with the standard symmetric $\mathbb{C}-$bilinear form*

$$b(z, w) := \sum_{i=1}^{n} z_i w_i.$$

The (complex) Clifford algebra obtained is denoted $\mathbb{C}l_n$.

Remark 20.2 *The complex Clifford algebra $\mathbb{C}l_n$ is based on a complex symmetric form and* not *on a Hermitian form.*

Exercise 20.2 *Show that for any nonnegative integers p, q with $p + q = n$ we have $Cl_{p,q} \otimes \mathbb{C} \cong \mathbb{C}l_n$.*

Example 20.2 *The Clifford algebra based on $\mathbb{R}^1$ itself with the relation $x^2 = -1$ is just the complex number system.*

The Clifford algebra construction can be globalized in the obvious way. In particular, we have the option of using the form presentation so that the above formulas $\alpha \cdot \beta := \alpha \wedge \beta - (\sharp\alpha)\lrcorner\beta$ and $\beta \cdot \alpha := (-1)^k(\alpha \wedge \beta + (\sharp\alpha)\lrcorner\beta)$ are interpreted as equations for differential forms $\alpha \in \wedge^1 T^*M$ and $\beta \in \wedge^k T^*M$ on a semi-Riemannian manifold M, g. In any case we have the following

Definition 20.9 *Given a Riemannian manifold* M, g*, the Clifford algebra bundle is* $Cl(T^*M, \mathrm{g}) = Cl(T^*M) := \cup_x Cl(T^*_x M)$.

Since we take each tangent space to be embedded $T^*_x M \subset Cl(T^*_x M)$, the elements θ^i of a local orthonormal frame $\theta^1, ..., \theta^n \in \Omega^1$ are also local sections of $Cl(T^*M, \mathrm{g})$ and satisfy

$$\theta^i\theta^j + \theta^j\theta^i = -\langle\theta^i, \theta^j\rangle = -\varepsilon^i\delta^{ij}$$

Recall that $\varepsilon^1, ..., \varepsilon^n$ is a list of numbers equal to ±1 (or even 0 if we allow degeneracy) and giving the index of the metric $\mathrm{g}(.,.) = \langle ., .\rangle$.

Obviously, we could also work with the bundle $Cl(TM) := \cup_x Cl(T_x M)$ which is naturally isomorphic to $Cl(T^*M)$ in which case we would have

$$e^i e^j + e^j e^i = -\langle e^i, e^j\rangle = -\varepsilon^i\delta^{ij}$$

for orthonormal frames. Of course it shouldn't make any difference to our development since one can just identify TM with T^*M by using the metric. On the other hand, we could define $Cl(T^*M, \mathrm{b})$ even if b is a *degenerate* bilinear tensor and then we recover the Grassmann algebra bundle $\wedge T^*M$ in case $\mathrm{b} \equiv 0$. These comments should make it clear that $Cl(T^*M, \mathrm{g})$ is in general a sort of deformation of the Grassmann algebra bundle.

There are a couple of things to notice about $Cl(T^*M)$ when we realize it as $\wedge T^*M$ with a new product. First of all if $\alpha, \beta \in \wedge T^*M$ and $\langle\alpha, \beta\rangle = 0$ then $\alpha \cdot \beta = \alpha \wedge \beta$ where as if $\langle\alpha, \beta\rangle \neq 0$ then in general $\alpha\beta$ is not a homogeneous element. Second, $Cl(T^*M)$ is locally generated by $\{1\} \cup \{\theta^i\} \cup \{\theta^i\theta^j : i < j\} \cup \cdots \cup \{\theta^1\theta^2 \cdots \theta^n\}$ where $\theta^1, \theta^2, ..., \theta^n$ is a local orthonormal frame. Now we can immediately define our current objects of interest:

Definition 20.10 *A* ***bundle of modules over*** $Cl(T^*M)$ *is a vector bundle* $\Sigma = (E, \pi, M)$ *such that each fiber* E_x *is a module over the algebra* $Cl(T^*_x M)$ *and such that for each* $\theta \in \Gamma(Cl(T^*M))$ *and each* $\sigma \in \Gamma(\Sigma)$ *the map* $x \mapsto \theta(x)\sigma(x)$ *is smooth. Thus we have an induced map on smooth sections:* $\Gamma(Cl(T^*M)) \times \Gamma(\Sigma) \to \Gamma(\Sigma)$.

Proposition 20.1 *The bundle* $Cl(T^*M)$ *is a Clifford module over itself and the Levi Civita connection* ∇ *on* M *induces a connection on* $Cl(T^*M)$ *(this connection is also denoted* ∇*) such that*

$$\nabla(\sigma_1\sigma_2) = (\nabla\sigma_1)\sigma_2 + \sigma_1\nabla\sigma_2$$

for all $\sigma_1, \sigma_2 \in \Gamma(Cl(T^*M))$*. In particular, if* $X, Y \in \mathfrak{X}(M)$ *and* $\sigma \in \Gamma(Cl(T^*M))$ *then*

$$\nabla_X(Y\sigma) = (\nabla_X Y)\sigma + Y\nabla_X\sigma.$$

Proof. Realize $Cl(T^*M)$ as $\wedge T^*M$ with Clifford multiplication and let ∇ be usual induced connection on $\wedge T^*M \subset \otimes T^*M$. We have for an local o.n frame $e_1, ..., e_n$ with dual frame $\theta^1, \theta^2, ..., \theta^n$. Then $\nabla_\xi \theta^i = -\Gamma^i_j(\xi)\theta^j$

$$\begin{aligned}
\nabla_\xi(\theta^i\theta^j) &= \nabla_\xi(\theta^i \wedge \theta^j) \\
&= \nabla_\xi\theta^i \wedge \theta^j + \theta^i \wedge \nabla_\xi\theta^j \\
&= -\Gamma^i_k(\xi)\theta^k \wedge \theta^j - \Gamma^j_k(\xi)\theta^k \wedge \theta^i \\
&= -\Gamma^i_k(\xi)\theta^k\theta^j - \Gamma^j_k(\xi)\theta^k\theta^i = (\nabla_\xi\theta^i)\theta^j + \theta^i\nabla_\xi\theta^j
\end{aligned}$$

The result follows by linearity and a simple induction since a general section σ can be written locally as $\sigma = \sum a_{i_1 i_2 \ldots i_k}\theta^{i_1}\theta^{i_2}\cdots\theta^{i_k}$. ■

Definition 20.11 *Let M, g be a (semi-) Riemannian manifold. A compatible connection for a* ***bundle of modules*** *Σ* ***over*** *$Cl(T^*M)$ is a connection ∇^Σ on Σ such that*

$$\nabla^\Sigma(\sigma \cdot s) = (\nabla\sigma)\cdot s + \sigma \cdot \nabla^\Sigma s$$

for all $s \in \Gamma(\Sigma)$ and all $\sigma \in \Gamma(Cl(T^*M))$

Definition 20.12 *Let $\Sigma = (E, \pi, M)$ be a bundle of modules over $Cl(T^*M)$ with a compatible connection $\nabla = \nabla^\Sigma$. The associated* ***Dirac operator*** *is defined as a differential operator Σ on by*

$$Ds := \sum \theta^i \cdot \nabla^\Sigma_{e_i} s$$

for $s \in \Gamma(\Sigma)$.

Notice that Clifford multiplication of $Cl(T^*M)$ on $\Sigma = (E, \pi, M)$ is a zeroth order operator and so is well defined as a fiberwise operation $Cl(T^*_xM) \times E_x \to E_x$.

There are still a couple of convenient properties that we would like to have. These are captured in the next definition.

Definition 20.13 *Let $\Sigma = (E, \pi, M)$ be a bundle of modules over $Cl(T^*M)$ such that Σ carries a Riemannian metric and compatible connection $\nabla = \nabla^\Sigma$. We call $\Sigma = (E, \pi, M)$ a* ***Dirac bundle*** *if the following equivalent conditions hold:*

1) $\langle es_1, es_2\rangle = \langle s_1, s_2\rangle$ for all $s_1, s_2 \in E_x$ and all $e \in T^*_xM \subset Cl(T^*_xM))$ with $|e| = 1$. In other words, Clifford multiplication by a unit (co)vector is required to be an isometry of the Riemannian metric on each fiber of Σ. Since , $e^2 = -1$ it follows that this is equivalent to requiring .

2) $\langle es_1, s_2\rangle = -\langle s_1, es_2\rangle$ for all $s_1, s_2 \in E_x$ and all $e \in T^*_xM \subset Cl(T^*_xM))$ with $|e| = 1$.

Assume in the sequel that q is nondegenerate. Let denote the subalgebra generated by all elements of the form $x_1 \cdots x_k$ with k even. And similarly, $Cl_1(\mathrm{V},q)$, with k odd. Thus $Cl(\mathrm{V},q)$ has the structure of a Z_2-graded algebra (also called a superalgebra):

$$Cl(\mathrm{V},q) = Cl_0(\mathrm{V},q) \oplus Cl_1(\mathrm{V},q)$$

$$Cl_0(\mathrm{V},q) \cdot Cl_0(\mathrm{V},q) \subset Cl_0(\mathrm{V},q)$$
$$Cl_0(\mathrm{V},q) \cdot Cl_1(\mathrm{V},q) \subset Cl_1(\mathrm{V},q)$$
$$Cl_1(\mathrm{V},q) \cdot Cl_1(\mathrm{V},q) \subset Cl_0(\mathrm{V},q)$$

$Cl_0(\mathrm{V},q)$ and $Cl_1(\mathrm{V},q\ )$ are referred to as the **even** and **odd** part respectively. There exists a fundamental automorphism α of $Cl(\mathrm{V},q)$ such that $\alpha(x) = -x$ for all $x \in$V. Note that $\alpha^2 = id$. It is easy to see that $Cl_0(\mathrm{V},q)$ and $Cl_1(\mathrm{V},q\ )$ are the $+1$ and -1 eigenspaces of $\alpha : Cl(\mathrm{V},q) \to Cl(\mathrm{V},q)$.

20.5.2 The Clifford group and Spinor group

Let G be the group of all invertible elements $s \in C_K$ such that $s\mathrm{V}s^{-1} =$V. This is called the **Clifford group** associated to q. The special Clifford group is $G^+ = G \cap C_0$. Now for every $s \in G$ we have a map $\phi_s : v \longrightarrow svs^{-1}$ for $v \in$V. It can be shown that ϕ is a map from G into $O(q)$, the orthogonal group of q. The kernel is the invertible elements in the center of C_K.

It is a useful and important fact that if $x \in G\cap$V then $q(x) \neq 0$ and $-\phi_x$ is reflection through the hyperplane orthogonal to x. Also, if s is in G^+ then ϕ_s is in $SO(q)$. In fact, $\phi(G^+) = SO(q)$.

Besides the fundamental automorphism α mentioned above, there is also a fundament anti-automorphism or **reversion** $\beta : Cl(\mathrm{V},q) \to Cl(\mathrm{V},q)$ which is determined by the requirement that $\beta(v_1 v_2 \cdots v_k) = v_k v_{k-1} \cdots v_1$ for $v_1, v_2, ..., v_k \in$V$\subset$ $Cl(\mathrm{V},q)$. We can use this anti-automorphism β to put a kind of "norm" on G^+;

$$N; G^+ \longrightarrow \mathbb{K}^*$$

where $\mathbb{K}^*$ is the multiplicative group of nonzero elements of $\mathbb{K}$ and $N(s) = \beta(s)s$. This is a homomorphism and if we "mod out" the kernel of N we get the so called reduced Clifford group G_0^+.

We now specialize to the real case $\mathbb{K} = \mathbb{R}$. The identity component of G_0^+ is called the **spin group** and is denoted by $\mathrm{Spin}(\mathrm{V}, q)$.

20.6 The Structure of Clifford Algebras

Now if $\mathbb{K} = \mathbb{R}$ and

$$q(x) = \sum_{i=1}^{r} (x_i)^2 - \sum_{i=r+1}^{r+s} (x_i)^2$$

we write $C(\mathbb{R}^{r+s}, q, \mathbb{R}) = Cl(r, s)$. Then one can prove the following isomorphisms.

$$Cl(r+1, s+1) \cong Cl(1,1) \otimes C(r,s)$$

$$Cl(s+2, r) \cong Cl(2,0) \otimes Cl(r,s)$$

$$Cl(s, r+2) \cong Cl(0,2) \otimes Cl(r,s)$$

and

$$Cl(p,p) \cong \bigotimes^{p} Cl(1,1)$$

$$Cl(p+k,p) \cong \bigotimes^{p} Cl(1,1) \bigotimes Cl(k,0)$$

$$Cl(k,0) \cong Cl(2,0) \otimes Cl(0,2) \otimes Cl(k-4,0) \qquad k > 4$$

Using the above type of periodicity relations together with

$$Cl(2,0) \cong Cl(1,1) \cong M_2(\mathbb{R})$$

$$Cl(1,0) \cong \mathbb{R} \oplus \mathbb{R}$$

and

$$Cl(0,1) \cong \mathbb{C}$$

we can piece together the structure of $Cl(r,s)$ in terms of familiar matrix algebras. We leave out the resulting table since for one thing we are more interested in the simpler complex case. Also, we will explore a different softer approach below.

The complex case . In the complex case we have a much simpler set of relations;

$$Cl(2r) \cong Cl(r,r) \otimes \mathbb{C} \cong M_{2^r}(\mathbb{C})$$

$$Cl(2r+1) \cong Cl(1) \otimes Cl(2r)$$

$$\cong Cl(2r) \oplus Cl(2r) \cong M_{2^r}(\mathbb{C}) \oplus M_{2^r}(\mathbb{C})$$

These relations remind us that we may use matrices to represent our Clifford algebras. Lets return to this approach and explore a bit.

20.6.1 Gamma Matrices

Definition 20.14 *A set of real or complex matrices $\gamma_1, \gamma_2, ..., \gamma_n$ are called gamma matrices for $Cl(r,s)$ if*

$$\gamma_i \gamma_j + \gamma_j \gamma_i = -2g_{ij}$$

where $(g_{ij}) = diag(1, ..., 1, -1, ...,)$ is the diagonalized matrix of signature (r,s).

Example 20.3 *Recall the Pauli matrices:*

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

It is easy to see that σ_1, σ_2 serve as gamma matrices for $Cl(0,2)$ while $i\sigma_1, i\sigma_2$ serve as gamma matrices for $Cl(2,0)$.

$Cl(2,0)$ is spanned as a vector space of matrices by $\sigma_0, i\sigma_1, i\sigma_2, i\sigma_3$ and is (algebra) isomorphic to the quaternion algebra $\mathbb{H}$ under the identification

$$\begin{aligned} \sigma_0 &\mapsto 1 \\ i\sigma_1 &\mapsto I \\ i\sigma_2 &\mapsto J \\ i\sigma_3 &\mapsto K \end{aligned}$$

Example 20.4 *The matrices $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ are gamma matrices for $Cl(1,3)$.*

20.7 Clifford Algebra Structure and Representation

20.7.1 Bilinear Forms

We will need some basic facts about bilinear forms. We review this here.

(1) Let E be a module over a commutative ring R. Typically E is a vector space over a field $\mathbb{K}$. A bilinear map g :E$\times$E$\longrightarrow$ R is called symmetric if $g(x,y) = g(y,x)$ and antisymmetric if $g(x,y) = -g(y,x)$ for (x,y) $\in$E$\times$E. If R has an automorphism of order two, $a \mapsto \bar{a}$ we say that g is Hermitian if $g(ax,y) = ag(x,y)$ and $g(x,ay) = \bar{a}g(x,y)$ for all $a \in R$ and (x,y) $\in$E$\times$E. If g is any of symmetric,antisymmetric,or Hermitian then the “ left kernel” of g is equal to the “right kernel”. That is

$$\begin{aligned} \ker g &= \{x \in \mathrm{E} : g(x,y) = 0 \quad \forall y \in \mathrm{E}\} \\ &= \{y \in \mathrm{E} : g(x,y) = 0 \quad \forall x \in \mathrm{E}\} \end{aligned}$$

If $\ker g = 0$ we say that g is nondegenerate. In case E is a vector space of finite dimension g is nondegenerate iff $x \mapsto g(x,\cdot)$ $\in$E* is an isomorphism. An orthogonal basis for g is a basis $\{v_i\}$ for E such that $g(v_i, v_i) = 0$ for $i \neq j$.

Definition 20.15 *Let E be a vector space over a three types above. If E=E$\oplus$E$_2$ for subspaces E$_i$ $\subset$E and $g(x_1, x_2) = 0 \qquad \forall x_1$ $\in$E$, x_2$ $\in$E$_2$ then we write*

$$\mathrm{E} = \mathrm{E}_1 \perp \mathrm{E}_2$$

and say that E is the orthogonal direct sum of E$_1$ and E$_2$.

Proposition 20.2 *Suppose E$, g$ is as above with*

$$\mathrm{E} = \mathrm{E}_1 \perp \mathrm{E}_2 \perp \cdots \perp \mathrm{E}_k$$

Then g is non-degenerate iff its restrictions $g|_{\mathrm{E}_i}$ are and

$$ker \mathrm{E} = \mathrm{E}_1^o \perp \mathrm{E}_2^o \perp \cdots \perp \mathrm{E}_k^o$$

Proof. Nearly obvious. ∎

Terminology: If g is one of symmetric, antisymmetric or Hermitian we say that g is geometric.

Proposition 20.3 *Let g be a geometric bilinear form on a vector space E (over $\mathbb{K}$). Suppose g is nondegenerate. Then g is nondegenerate on a subspace F iff E= $F \perp F^\perp$ where*

$$F^\perp = \{x \in \mathrm{E} : g(x,f) = 0 \qquad \forall f \in F\}$$

Definition 20.16 *A map q is called quadratic iff there is a symmetric g such that $q(x) = g(x, x)$. Note that g can be recovered from q:*

$$2g(x, y) = q(x + y) - q(x) - q(y)$$

20.7.2 Hyperbolic Spaces And Witt Decomposition

E, g is a vector space with symmetric form g. If E has dimension 2 we call E a hyperbolic plane. If $\dim\mathrm{E} \geq 2$ and $\mathrm{E}=\mathrm{E}_1 \perp \mathrm{E}_2 \perp \cdots \perp \mathrm{E}_k$ where each E_i is a hyperbolic plane for $g|_{\mathrm{E}_i}$ then we call E a hyperbolic space. For a hyperbolic plane one can easily construct a basis f_1, f_2 such that $g(f_1, f) = g(f_2, f_2) = 0$ and $g(f_1, f_2) = 1$. So that with respect to this basis g is given by the matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This pair $\{f_1, f_2\}$ is called a hyperbolic pair for E, g. Now we return to $\dim\mathrm{E} \geq 2$. Let $\mathrm{rad}F \equiv F^{\perp} \cap F = \ker g|_F$

Lemma 20.1 *There exists a subspace $U \subset$E such that E$=$ radE$\perp$ U and U is nondegenerate.*

Proof. It is not to hard to see that $\mathrm{rad}U = \mathrm{rad}U^{\perp}$. If $\mathrm{rad}U = 0$ then $\mathrm{rad}U^{\perp} = 0$ and visa vera. Now $U + U^{\perp}$ is clearly direct since $0 = \mathrm{rad}U = U \cap U^{\perp}$. Thus E$=$ $U \perp U^{\perp}$. ■

Lemma 20.2 *Let g be nondegenerate and $U \subset$E some subspace. Suppose that $U = \mathrm{rad}U \perp W$ where $\mathrm{rad}W = 0$. Then given a basis $\{u_1, \cdots, u_s\}$ for radU there exists $v_1, \cdots, v_s \in W^{\perp}$ such that each $\{u_i, v_i\}$ is a hyperbolic pair. Let $P_i = \mathrm{span}\{u_i, v_i\}$. Then*

$$\mathrm{E} = W \perp P_1 \perp \cdots \perp P_s.$$

Proof. Let $W_1 = \mathrm{span}\{u_2, u_3, \cdots, u_s\} \oplus W$. Then $W_1 \subsetneq \mathrm{rad}U \oplus W$ so $(\mathrm{rad}U \oplus W)^{\perp} \subsetneq W_1^{\perp}$. Let $w_1 \in W_1^{\perp}$ but assume $w_1 \notin (\mathrm{rad}U \oplus W)^{\perp}$. Then we have $g(u_1, w_1) \neq 0$ so that $P_1 = \mathrm{span}\{u_1, w_1\}$ is a hyperbolic plane. Thus we can find v_1 such that u_1, v_1 is a hyperbolic pair for P_1. We also have

$$U_1 = (u_2, u_3 \cdots u_s) \perp P_1 \perp W$$

so we can proceed inductively since $u_2, U_3, \ldots u_s \in \mathrm{rad}U_1$. ■

Definition 20.17 *A subspace $U \subset$E is called totally isotropic if $g|_U \equiv 0$.*

Proposition 20.4 *(Witt decomposition) Suppose that $U \subset$E is a maximal totally isotropic subspace and $e_1, e_2, \ldots e_r$ a basis for U. Then there exist (null) vectors $f_1, f_2, \ldots, f_r$ such that each $\{e_i, f_i\}$ is a hyperbolic pair and $U' = \mathrm{span}\{f_i\}$ is totally isotropic. Further*

$$\mathrm{E} = U \oplus U' \perp G$$

where $G = (U \oplus U')^{\perp}$.

Proof. Using the proof of the previous theorem we have rad$U = U$ and $W = 0$. The present theorem now follows. ■

Proposition 20.5 *If g is symmetric then $g|_G$ is definite.*

EXAMPLE: Let E, $g = \mathbb{C}^{2k}, g_0$ where

$$g_0(z, w) = \sum_{i=1}^{2k} z_i w_i$$

Let $\{e_1, ..., e_k, e_{k+1}, ..., e_{2k}\}$ be the standard basis of $\mathbb{C}^{2k}$. Define

$$\varepsilon_j = \frac{1}{\sqrt{2}}(e_j + ie_{k+j}) \qquad j = 1, ..., k$$

and

$$\eta_j = \frac{1}{\sqrt{2}}(e_i - ie_{k+j}).$$

Then letting $F = \text{span}\{\varepsilon_i\}$, $F' = \text{span}\{\eta_j\}$ we have $\mathbb{C}^{2k} = F \oplus F'$ and F is a maximally isotropic subspace. Also, each $\{\varepsilon_j, \eta_j\}$ is a hyperbolic pair.

This is the most important example of a neutral space:

Proposition 20.6 *A vector space E with quadratic form is called neutral if the rank, that is, the dimension of a totally isotropic subspace, is $r = dim E/2$. The resulting decomposition $F \oplus F'$ is called a (weak) polarization.*

20.7.3 Witt's Decomposition and Clifford Algebras

Even Dimension Suppose that V, Q is quadratic space over $\mathbb{K}$. Let dimV$= r$ and suppose that V, Q is neutral. Then we have that $C_{\mathbb{K}}$ is isomorphic to End(S) for an r dimensional space S (spinor space). In particular, $C_{\mathbb{K}}$ is a simple algebra.

Proof. Let $F \oplus F'$ be a polarization of V. Here, F and F' are maximal totally isotropic subspaces of V. Now let $\{x_1, ..., x_r, y_1, ..., y_r\}$ be a basis for V such that $\{x_i\}$ is a basis for F and $\{y_i\}$ a basis for F'. Set $f = y_1 y_2 \cdots y_h$. Now let S be the span of elements of the form $x_{i_1} x_{i_2} \cdots x_{i_r} f$ where $1 \leq i_1 < ... < i_h \leq r$. S is an ideal of $C_{\mathbb{K}}$ of dimension 2^r. We define a representation ρ of $C_{\mathbb{K}}$ in S by

$$\rho(u)s = us$$

This can be shown to be irreducible so that we have the desired result. ■

Now since we are interested in spin which sits inside and in fact generates C_0 we need the following

Proposition 20.7 *C_0 is isomorphic to $End(S^+) \times End(S^-)$ where $S^+ = C_0 \cap S$ and $S^- = C_1 \cap S$.*

This follows from the obvious fact that each of $C_0 f$ and $C_1 f$ are invariant under multiplication by C_0.

Now consider a real quadratic space V, Q where Q is positive definite. We have $\mathrm{Spin}(n) \subset Cl^0(0) \subset C_0$ and $\mathrm{Spin}(n)$ generates C_0. Thus the complex spin group representation of is just given by restriction and is semisimple factoring as $S^+ \oplus S^-$.

Odd Dimension In the odd dimensional case we can not expect to find a polarization but this cloud turns out to have a silver lining. Let x_0 be a non-isotropic vector from V and set $\mathrm{V}_1 = (x_0)^\perp$. On V_1 we define a quadratic form Q_1 by

$$Q_1(y) = -Q(x_0)Q(y)$$

for $y \in \mathrm{V}_1$. It can be shown that Q_1 is non-degenerate. Now notice that for $y \in \mathrm{V}_1$ then $x_0 y = -y x_0$ and further

$$(x_0 y)^2 = -x_0^2 y^2 = -Q(x_0)Q(y) = Q_1(y)$$

so that by the universal mapping property the map

$$y \longrightarrow x_0 y$$

can be extended to an algebra morphism h from $\mathrm{Cl}(Q_1,\mathrm{V}_1)$ to $C_{\mathbb{K}}$. Now these two algebras have the same dimension and since C_o is simple it must be an isomorphism. Now if Q has rank r then Q_1,V_1 is neutral and we obtain the following

Theorem 20.4 *If the dimension of V is odd and Q has rank r then C_0 is represented irreducibly in a space S^+ of dimension 2^r. In particular $C_0 \cong End(S^+)$.*

20.7.4 The Chirality operator

Let V be a Euclidean vector space with associated positive definite quadratic form Q. Let $\{e_1, ..., e_n\}$ be an oriented orthonormal. frame for V. We define the Chirality operator τ to be multiplication in the associated (complexified) Clifford algebra by the element

$$\tau = (\sqrt{-1})^{n/2} e_1 \cdots e_n$$

if n is even and by

$$\tau = (\sqrt{-1})^{(n+1)/2} e_1 \cdots e_n$$

if n is odd. Here $\tau \in Cl(n)$ and does not depend on the choice of orthonormal. oriented frame. We also have that $\tau v = -v\tau$ for $v \in$V and $\tau^2 = 1$.

Let us consider the case of n even. Now we have seen that we can write $\mathrm{V} \otimes C = F \oplus \bar{F}$ where F is totally isotropic and of dimension n. In fact we may assume that F has a basis $\{e_{2j-1} - ie_{2j} \ : 1 \leq j \leq n/2\}$, where the e_i come from an oriented orthonormal basis. Lets use this polarization to once again construct the spin representation.

First note that Q (or its associated bilinear form) places F and $\bar{F}$ in duality so that we can identify $\bar{F}$ with the dual space F'. Now set $S = \wedge F$. First we show how V act on S. Given $v \in$V consider $v \in$V$\otimes C$ and decompose $v = w + \bar{w}$ according to our decomposition above. Define $\phi_w s = \sqrt{2} w \wedge s$ and

$$\phi_{\bar{w}} s = -\iota(\bar{w}) s.$$

where ι is interior multiplication. Now extend ϕ linearly to V. Exercise Show that ϕ extends to a representation of $C \otimes Cl(n)$. Show that $S^+ = \wedge^+ F$ is invariant under C_0. It turns out that ϕ_τ is $(-1)^k$ on $\wedge^k F$

20.7.5 Spin Bundles and Spin-c Bundles

20.7.6 Harmonic Spinors

Chapter 21

Complex Manifolds

21.1 Some complex linear algebra

The set of all n-tuples of complex $\mathbb{C}^n$ numbers is a complex vector space and by choice of a basis, every complex vector space of finite dimension (over $\mathbb{C}$) is linearly isomorphic to $\mathbb{C}^n$ for some n. Now multiplication by $i := \sqrt{-1}$ is a complex linear map $\mathbb{C}^n \to \mathbb{C}^n$ and since $\mathbb{C}^n$ is also a real vector space $\mathbb{R}^{2n}$ under the identification

$$(x^1 + iy^1, ..., x^n + iy^n) \rightleftharpoons (x^1, y^1, ..., x^n, y^n)$$

we obtain multiplication by i as a real linear map $J_0 : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by the matrix

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 0 & & & & \\ & & 0 & -1 & & \\ & & 1 & 0 & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix}.$$

Conversely, if V is a real vector space of dimension $2n$ and there is a map $J : \mathrm{V} \to \mathrm{V}$ with $J^2 = -1$ then we can define the structure of a complex vector space on V by defining the scalar multiplication by complex numbers via the formula

$$(x + \mathrm{i}y)v := xv + yJv \text{ for } v \in \mathrm{V}.$$

Denote this complex vector space by V_J. Now if $e_1, ...e_n$ is a basis for V_J (over $\mathbb{C}$) then we claim that $e_1, ..., e_n, Je_1, ..., Je_n$ is a basis for V over $\mathbb{R}$. We only need to show that $e_1, ..., e_n, Je_1, ..., Je_n$ span. For this let $v \in \mathrm{V}$ and then for some complex numbers $c^i = a^i + \mathrm{i}b^j$ we have $\sum c^i e_i = \sum (a^j + \mathrm{i}b^j) e_j = \sum a^j e_j + \sum b^j J e_j$.

Next we consider the complexification of V which is $\mathrm{V}_{\mathbb{C}} := \mathbb{C} \otimes \mathrm{V}$. Now any real basis $\{f_j\}$ of V is also a basis for $\mathrm{V}_{\mathbb{C}}$ iff we identify f_j with $1 \otimes f_j$.

Furthermore, the linear map $J : \mathrm{V} \to \mathrm{V}$ extends to a complex linear map $J : \mathrm{V}_{\mathbb{C}} \to \mathrm{V}_{\mathbb{C}}$ and still satisfies $J^2 = -1$. Thus this extension has eigenvalues i and $-$i. Let $\mathrm{V}^{1,0}$ be the i eigenspace and $\mathrm{V}^{0,1}$ be the $-$i eigenspace. Of course we must have $\mathrm{V}_{\mathbb{C}} = \mathrm{V}^{1,0} \oplus \mathrm{V}^{0,1}$. The reader may check that the set of vectors $\{e_1 - \mathrm{i}Je_1, ..., e_n - \mathrm{i}Je_n\}$ span $\mathrm{V}^{1,0}$ while $\{e_1 + \mathrm{i}Je_1, ..., e_n + \mathrm{i}Je_n\}$ span $\mathrm{V}^{0,1}$. Thus we have a convenient basis for $\mathrm{V}_{\mathbb{C}} = \mathrm{V}^{1,0} \oplus \mathrm{V}^{0,1}$.

Lemma 21.1 *There is a natural complex linear isomorphism* $\mathrm{V}_J \cong \mathrm{V}^{1,0}$ *given by* $e_i \mapsto e_i - \mathrm{i}Je_i$. *Furthermore, the conjugation map on* $\mathrm{V}_{\mathbb{C}}$ *interchanges the spaces* $\mathrm{V}^{1,0}$ *and* $\mathrm{V}^{0,1}$.

Let us apply these considerations to the simple case of the complex plane $\mathbb{C}$. The realification is $\mathbb{R}^2$ and the map J is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

If we identify the tangent space of $\mathbb{R}^{2n}$ at 0 with $\mathbb{R}^{2n}$ itself then $\{\left.\frac{\partial}{\partial x^i}\right|_0, \left.\frac{\partial}{\partial y^i}\right|_0\}_{1\le i\le n}$ is basis for $\mathbb{R}^{2n}$. A | complex basis for $\mathbb{C}^n \cong (\mathbb{R}^{2n}, J_0)$ is, for instance, $\{\left.\frac{\partial}{\partial x^i}\right|_0\}_{1\le i\le n}$. A complex basis for $\mathbb{R}^2_J \cong \mathbb{C}$ is $e_1 = \left.\frac{\partial}{\partial x}\right|_0$ and so $\left.\frac{\partial}{\partial x}\right|_0, J\left.\frac{\partial}{\partial x}\right|_0$ is a basis for $\mathbb{R}^2$. This is clear anyway since $J\left.\frac{\partial}{\partial x}\right|_0 = \left.\frac{\partial}{\partial y}\right|_0$. Now the complexification of $\mathbb{R}^2$ is $\mathbb{R}^2_{\mathbb{C}}$ which has basis consisting of $e_1 - \mathrm{i}Je_1 = \left.\frac{\partial}{\partial x}\right|_0 - \mathrm{i}\left.\frac{\partial}{\partial y}\right|_0$ and $e_1 + \mathrm{i}Je_1 = \left.\frac{\partial}{\partial x}\right|_0 + \mathrm{i}\left.\frac{\partial}{\partial y}\right|_0$. These are usually denoted by $\left.\frac{\partial}{\partial z}\right|_0$ and $\left.\frac{\partial}{\partial \bar{z}}\right|_0$. More generally, we see that if $\mathbb{C}^n$ is reified to $\mathbb{R}^{2n}$ which is then complexified to $\mathbb{R}^{2n}_{\mathbb{C}} := \mathbb{C} \otimes \mathbb{R}^{2n}$ then a basis for $\mathbb{R}^{2n}_{\mathbb{C}}$ is given by

$$\{\left.\frac{\partial}{\partial z^1}\right|_0, .., \left.\frac{\partial}{\partial z^n}\right|_0, \left.\frac{\partial}{\partial \bar{z}^1}\right|_0 ..., \left.\frac{\partial}{\partial \bar{z}^n}\right|_0\}$$

where

$$2\left.\frac{\partial}{\partial z^i}\right|_0 := \left.\frac{\partial}{\partial x^i}\right|_0 - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_0$$

and

$$2\left.\frac{\partial}{\partial \bar{z}^i}\right|_0 := \left.\frac{\partial}{\partial x^i}\right|_0 + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_0.$$

Now if we consider the tangent bundle $U \times \mathbb{R}^{2n}$ of an open set $U \subset \mathbb{R}^{2n}$ then we have the vector fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$. We can complexify the tangent bundle of $U \times \mathbb{R}^{2n}$ to get $U \times \mathbb{R}^{2n}_{\mathbb{C}}$ and then following the ideas above we have that the fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ also span each tangent space $T_pU := \{p\} \times \mathbb{R}^{2n}_{\mathbb{C}}$. On the other hand, so do the fields $\{\frac{\partial}{\partial z^1}, .., \frac{\partial}{\partial z^n}, \frac{\partial}{\partial \bar{z}^1}, .., \frac{\partial}{\partial \bar{z}^n}\}$. Now if $\mathbb{R}^{2n}$ had a complex vector space structure, say $\mathbb{C}^n \cong (\mathbb{R}^{2n}, J_0)$, then J_0 defines a bundle map $J_0 : T_pU \to T_pU$ given by $(p, v) \mapsto (p, J_0 v)$. This can be extended to a complex bundle map $J_0 : TU_{\mathbb{C}} = \mathbb{C} \otimes TU \to TU_{\mathbb{C}} = \mathbb{C} \otimes TU$ and we get a bundle decomposition

$$TU_{\mathbb{C}} = T^{1.0}U \oplus T^{0.1}U$$

where $\frac{\partial}{\partial z^1}, .., \frac{\partial}{\partial z^n}$ spans $T^{1.0}U$ at each point and $\frac{\partial}{\partial \bar{z}^1}, .., \frac{\partial}{\partial \bar{z}^n}$ spans $T^{0.1}U$.

Now the symbols $\frac{\partial}{\partial z^1}$ etc., already have a meaning a differential operators. Let us now show that this view is at least consistent with what we have done above. For a smooth complex valued function $f : U \subset \mathbb{C}^n \to \mathbb{C}$ we have for $p = (z_1, ..., z_n) \in U$

$$\begin{aligned}
\left.\frac{\partial}{\partial z^i}\right|_p f &= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p f - \mathrm{i} \left.\frac{\partial}{\partial y^i}\right|_p f \right) \\
&= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p u - \mathrm{i} \left.\frac{\partial}{\partial y^i}\right|_p u - \left.\frac{\partial}{\partial x^i}\right|_p \mathrm{i}v - \mathrm{i} \left.\frac{\partial}{\partial y^i}\right|_p \mathrm{i}v \right) \\
&\frac{1}{2}\left(\left.\frac{\partial u}{\partial x^i}\right|_p + \left.\frac{\partial v}{\partial y^i}\right|_p \right) + \frac{\mathrm{i}}{2}\left(\left.\frac{\partial u}{\partial y^i}\right|_p - \left.\frac{\partial v}{\partial x^i}\right|_p \right).
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f &= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p f + \mathrm{i} \left.\frac{\partial}{\partial y^i}\right|_p f \right) \\
&= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p u + \mathrm{i} \left.\frac{\partial}{\partial y^i}\right|_p u + \left.\frac{\partial}{\partial x^i}\right|_p \mathrm{i}v + \mathrm{i} \left.\frac{\partial}{\partial y^i}\right|_p \mathrm{i}v \right) \\
&= \frac{1}{2}\left(\left.\frac{\partial u}{\partial x^i}\right|_p - \left.\frac{\partial v}{\partial y^i}\right|_p \right) + \frac{\mathrm{i}}{2}\left(\left.\frac{\partial u}{\partial y^i}\right|_p + \left.\frac{\partial v}{\partial x^i}\right|_p \right).
\end{aligned}$$

Definition 21.1 *A function $f : U \subset \mathbb{C}^n \to \mathbb{C}$ is called **holomorphic** if*

$$\frac{\partial}{\partial \bar{z}^i} f \equiv 0 \qquad \textit{(all i)}$$

*on U. A function f is called **antiholomorphic** if*

$$\frac{\partial}{\partial z^i} f \equiv 0 \qquad \textit{(all i).}$$

Definition 21.2 *A map $f : U \subset \mathbb{C}^n \to \mathbb{C}^m$ given by functions $f_1, ..., f_m$ is called **holomorphic** (resp. **antiholomorphic**) if each component function $f_1, ..., f_m$ is holomorphic (resp. antiholomorphic).*

Now if $f : U \subset \mathbb{C}^n \to \mathbb{C}$ is holomorphic then by definition $\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f \equiv 0$ for all $p \in U$ and so we have the **Cauchy-Riemann equations**

$$\begin{aligned}
\frac{\partial u}{\partial x^i} &= \frac{\partial v}{\partial y^i} \\
\frac{\partial v}{\partial x^i} &= -\frac{\partial u}{\partial y^i}
\end{aligned} \qquad \text{(Cauchy-Riemann)}$$

and from this we see that for holomorphic f

$$\frac{\partial f}{\partial z^i}$$
$$= \frac{\partial u}{\partial x^i} + \mathrm{i}\frac{\partial v}{\partial x^i}$$
$$= \frac{\partial f}{\partial x^i}$$

which means that as derivations on the sheaf $\mathcal{O}$ of locally defined holomorphic functions on $\mathbb{C}^n$, the operators $\frac{\partial}{\partial z^i}$ and $\frac{\partial}{\partial x^i}$ are equal. This corresponds to the complex isomorphism $T^{1.0}U \cong TU, J_0$ which comes from the isomorphism in lemma **??**. In fact, if one looks at a function $f : \mathbb{R}^{2n} \to \mathbb{C}$ as a differentiable map of real manifolds then with J_0 given the isomorphism $\mathbb{R}^{2n} \cong \mathbb{C}^n$, out map f is holomorphic iff

$$Tf \circ J_0 = J_0 \circ Tf$$

or in other words

$$\begin{pmatrix} \frac{\partial u}{\partial x^1} & \frac{\partial u}{\partial y^1} & \\ \frac{\partial v}{\partial x^1} & \frac{\partial v}{\partial y^1} & \\ & & \ddots \end{pmatrix}\begin{pmatrix} 0 & -1 & \\ 1 & 0 & \\ & & \ddots \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x^1} & \frac{\partial u}{\partial y^1} & \\ \frac{\partial v}{\partial x^1} & \frac{\partial v}{\partial y^1} & \\ & & \ddots \end{pmatrix}\begin{pmatrix} 0 & -1 & \\ 1 & 0 & \\ & & \ddots \end{pmatrix}.$$

This last matrix equation is just the Cauchy-Riemann equations again.

21.2 Complex structure

Definition 21.3 *A manifold M is said to be an **almost complex manifold** if there is a smooth bundle map $J : TM \to TM$, called an **almost complex structure** , having the property that $J^2 = -1$.*

Definition 21.4 *A **complex manifold** M is a manifold modelled on $\mathbb{C}^n$ for some n, together with an atlas for M such that the transition functions are all holomorphic maps. The charts from this atlas are called **holomorphic charts**. We also use the phrase "**holomorphic coordinates**".*

Example 21.1 *Let $S^2(1/2) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1/4\}$ be given coordinates $\psi^+ : (x_1, x_2, x_3) \mapsto \frac{1}{1-x_3}(x_1 + \mathrm{i}x_2) \in \mathbb{C}$ on $U^+ := \{(x_1, x_2, x_3) \in S^2 : 1 - x_3 \neq 0\}$ and $\psi^- : (x_1, x_2, x_3) \mapsto \frac{1}{1+x_3}(x_1 + \mathrm{i}x_2) \in \mathbb{C}$ on $U^- := \{(x_1, x_2, x_3) \in S^2 : 1 + x_3 \neq 0\}$. Then z and w are coordinates on $S^2(1/2)$ with transition function $\psi^- \circ \psi^+(z) = 1/z$. Since on $\psi^+U^+ \cap \psi^-U^-$ the map $z \mapsto 1/z$ is a biholomorphism we see that $S^2(1/2)$ can be given the structure of a complex 1-manifold.*

Another way to get the same complex 1-manifold is by taking two copies of the complex pane, say $\mathbb{C}_z$ with coordinate z and $\mathbb{C}_w$ with coordinate z and then identify $\mathbb{C}_z$ with $\mathbb{C}_w - \{0\}$ via the map $w = 1/z$. This complex surface *is of course topologically a sphere and is also the 1 point compactification of the complex plane. As the reader will not doubt already be aware, this complex 1-manifold is called the* ***Riemann sphere****.*

Example 21.2 *Let $P_n(\mathbb{C})$ be the set of all complex lines through the origin in $\mathbb{C}^{n+1}$ which is to say the set of all equivalence classes of nonzero elements of $\mathbb{C}^{n+1}$ under the equivalence relation*

$$(z^1,...,z^{n+1}) \sim \lambda(z^1,...,z^{n+1}) \text{ for } \lambda \in \mathbb{C}$$

For each i with $1 \leq i \leq n+1$ define the set

$$U_i := \{[z^1,...,z^{n+1}] \in P_n(\mathbb{C}) : z^i \neq 0\}$$

and corresponding map $\psi_i : U_i \to \mathbb{C}^n$ by

$$\psi_i([z^1,...,z^{n+1}]) = \frac{1}{z^i}(z^1,...,\widehat{z^i},...,z^{n+1}) \in \mathbb{C}^n.$$

*One can check that these maps provide a holomorphic atlas for $P_n(\mathbb{C})$ which is therefore a complex manifold (**complex projective n-space**).*

Example 21.3 *Let $\mathbb{C}^m_n$ be the space of $m \times n$ complex matrices. This is clearly a complex manifold since we can always "line up" the entries to get a map $\mathbb{C}^m_n \to \mathbb{C}^{mn}$ and so as complex manifolds $\mathbb{C}^m_n \cong \mathbb{C}^{mn}$. A little less trivially we have the complex general linear group $\mathrm{GL}(n,\mathbb{C})$ which is an open subset of $\mathbb{C}^n_n$ and so is an n^2 dimensional complex manifold.*

Example 21.4 (Grassmannian manifold) *To describe this important example we start with the set $(\mathbb{C}^n_k)_*$ of $n \times k$ matrices with rank $k < n$ (maximal rank). The columns of each matrix from $(\mathbb{C}^n_k)_*$ span a k-dimensional subspace of $\mathbb{C}^n$. Define two matrices from $(\mathbb{C}^n_k)_*$ to be equivalent if they span the same k-dimensional subspace. Thus the set $G(k,n)$ of equivalence classes is in one to one correspondence with the set of complex k dimensional subspaces of $\mathbb{C}^n$. Now let U be the set of all $[A] \in G(k,n)$ such that A has its first k rows linearly independent. This property is independent of the representative A of the equivalence class $[A]$ and so U is a well defined set. This last fact is easily proven by a Gaussian reduction argument. Now every element $[A] \in U \subset G(k,n)$ is an equivalence class that has a unique member A_0 of the form*

$$\begin{pmatrix} I_{k\times k} \\ Z \end{pmatrix}.$$

Thus we have a map on U defined by $\Psi : [A] \mapsto Z \in \mathbb{C}^{n-k}_k \cong \mathbb{C}^{k(n-k)}$. We wish to cover $G(k,n)$ with sets U_σ similar to U and defined similar maps. Let $\sigma_{i_1...i_k}$ be the shuffle permutation that puts the k columns indexed by $i_1,...,i_k$ into the positions $1,...,k$ without changing the relative order of the remaining columns. Now consider the set $U_{i_1...i_k}$ of all $[A] \in G(k,n)$ such that any representative

*A has its k rows indexed by $i_1, ..., i_k$ linearly independent. The the permutation induces an obvious 1-1 onto map $\widetilde{\sigma_{i_1 ... i_k}}$ from $U_{i_1 ... i_k}$ onto $U = U_{1...k}$. We now have maps $\Psi_{i_1 ... i_k} : U_{i_1 ... i_k} \to \mathbb{C}_k^{n-k} \cong \mathbb{C}^{k(n-k)}$ given by composition $\Psi_{i_1 ... i_k} := \Psi \circ \widetilde{\sigma_{i_1 ... i_k}}$. These maps form an atlas $\{\Psi_{i_1 ... i_k}, U_{i_1 ... i_k}\}$ for $G(k, n)$ which turns out to be a holomorphic atlas (biholomorphic transition maps) and so gives $G(k, n)$ the structure of a complex manifold called the **Grassmannian manifold** of complex k-planes in $\mathbb{C}^n$.*

Definition 21.5 *A complex 1-manifold (so real dimension is 2) is called a **Riemann surface**.*

If S is a subset of a complex manifold M such that near each $p_0 \in S$ there exists a holomorphic chart $U, \psi = (z^1, ..., z^n)$ such that $0 \in S \cap U$ iff $z^{k+1}(p) = \cdots = z^n(p) = 0$ then the coordinates $z^1, ..., z^k$ restricted to $U \cap S$ is a chart on the set S and the set of all such charts gives S the structure of a complex manifold. In this case we call S a **complex submanifold** of M.

Definition 21.6 *In the same way as we defined differentiability for real manifolds we define the notion of a **holomorphic map** (resp. **antiholomorphic map**) from one complex manifold to another. Note however, that we must use holomorphic charts for the definition.*

The proof of the following lemma is straightforward.

Lemma 21.2 *Let $\psi : U \to \mathbb{C}^n$ be a holomorphic chart with $p \in U$. Then writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i} y^k$ we have that the map $J_p : T_pM \to T_pM$ defined by*

$$J_p \left. \frac{\partial}{\partial x^i} \right|_p = \left. \frac{\partial}{\partial y^i} \right|_p$$
$$J_p \left. \frac{\partial}{\partial y^i} \right|_p = - \left. \frac{\partial}{\partial x^i} \right|_p$$

is well defined independent of the choice of coordinates.

The maps J_p combine to give a bundle map $J : TM \to TM$ and so an almost complex structure on M called the almost complex structure induced by the holomorphic atlas.

Definition 21.7 *An almost complex structure J on M is said to be **integrable** if there it has a holomorphic atlas giving the map J as the induced almost complex structure. That is if there is an family of admissible charts $\psi_\alpha : U_\alpha \to \mathbb{R}^{2n}$ such that after identifying $\mathbb{R}^{2n}$ with $\mathbb{C}^n$ the charts form a holomorphic atlas with J the induced almost complex structure. In this case, we call J a **complex structure**.*

21.3 Complex Tangent Structures

Let $\mathcal{F}_p(\mathbb{C})$ denote the algebra germs of *complex* valued smooth functions at p on a complex n-manifold M thought of as a smooth real $2n-$manifold with real tangent bundle TM. Let $\mathrm{Der}_p(\mathcal{F})$ be the space of derivations this algebra. It is not hard to see that this space is isomorphic to the complexified tangent space $T_pM_{\mathbb{C}} = \mathbb{C}\otimes T_pM$. The (complex) algebra of germs of holomorphic functions at a point p in a complex manifold is denoted $\mathcal{O}_p$ and the set of derivations of this algebra denoted $\mathrm{Der}_p(\mathcal{O})$. We also have the algebra of germs of antiholomorphic functions at p which is $\overline{\mathcal{O}}_p$ and also $\mathrm{Der}_p(\overline{\mathcal{O}})$.

If $\psi : U \to \mathbb{C}^n$ is a holomorphic chart then writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have the differential operators at $p \in U$:

$$\left\{ \left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\}$$

(now transferred to the manifold). To be pedantic about it, we now denote the coordinates on $\mathbb{C}^n$ by $w_i = u_i + \mathrm{i}v_i$ and then

$$\begin{aligned} \left.\frac{\partial}{\partial z^i}\right|_p f &:= \left.\frac{\partial f \circ \psi^{-1}}{\partial w^i}\right|_{\psi(p)} \\ \left.\frac{\partial}{\partial \bar{z}^i}\right|_p f &:= \left.\frac{\partial f \circ \psi^{-1}}{\partial \bar{w}^i}\right|_{\psi(p)} \end{aligned}$$

Thought of derivations these span $\mathrm{Der}_p(\mathcal{F})$ but we have also seen that they span the complexified tangent space at p. In fact, we have the following:

$$\begin{aligned} T_pM_{\mathbb{C}} &= \mathrm{span}_{\mathbb{C}} \left\{ \left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\} = \mathrm{Der}_p(\mathcal{F}) \\ T_pM^{1,0} &= \mathrm{span}_{\mathbb{C}} \left\{ \left.\frac{\partial}{\partial z^i}\right|_p \right\} \\ &= \{v \in \mathrm{Der}_p(\mathcal{F}) : vf = 0 \text{ for all } f \in \overline{\mathcal{O}}_p\} \\ T_pM^{0,1} &= \mathrm{span}_{\mathbb{C}} \left\{ \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\} \\ &= \{v \in \mathrm{Der}_p(\mathcal{F}) : vf = 0 \text{ for all } f \in \mathcal{O}_p\} \end{aligned}$$

and of course

$$T_pM = \mathrm{span}_{\mathbb{R}} \left\{ \left.\frac{\partial}{\partial x^i}\right|_p, \left.\frac{\partial}{\partial y^i}\right|_p \right\}.$$

The reader should go back and check that the above statements are consistent with our definitions as long as we view the $\left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p$ not only as the algebraic objects constructed above but also as derivations. Also, the definitions of

$T_pM^{1,0}$ and $T_pM^{0,1}$ are independent of the holomorphic coordinates since we also have

$$T_pM^{1,0} = \ker\{J_p : T_pM \to T_pM\}$$

21.4 The holomorphic tangent map.

We leave it to the reader to verify that the construction that we have at each tangent space globalize to give natural vector bundles $TM_{\mathbb{C}}, TM^{1,0}$ and $TM^{0,1}$ (all with M as base space).

Let M and N be complex manifolds and let $f : M \to N$ be a smooth map. The tangent map extend to a map of the complexified bundles $Tf : TM_{\mathbb{C}} \to TN_{\mathbb{C}}$. Now $TM_{\mathbb{C}} = TM^{1,0} \oplus TM^{0,1}$ and similarly $TM_{\mathbb{C}} = TN^{1,0} \oplus TN^{0,1}$. If f is holomorphic then $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$. In fact since it is easily verified that $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$ equivalent to the Cauchy-Riemann equations being satisfied by the local representative on F in any holomorphic chart we obtain the following

Proposition 21.1 $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$ *if and only if* f *is a holomorphic map.*

The map given by the restriction $T_pf : T_pM^{1,0} \to T_{fp}N^{1,0}$ is called the **holomorphic tangent** map at p. Of course, these maps concatenate to give a bundle map

21.5 Dual spaces

Let M, J be a complex manifold. The dual of $T_pM_{\mathbb{C}}$ is $T_p^*M_{\mathbb{C}} = \mathbb{C} \otimes T_p^*M$. Now the map J has a dual bundle map $J^* : T^*M_{\mathbb{C}} \to T^*M_{\mathbb{C}}$ which must also satisfy $J^* \circ J^* = -1$ and so we have the at each $p \in M$ the decomposition by eigenspaces

$$T_p^*M_{\mathbb{C}} = T_p^*M^{1,0} \oplus T_p^*M^{0,1}$$

corresponding to the eigenvalues $\pm \mathrm{i}$.

Definition 21.8 *The space* $T_p^*M^{1,0}$ *is called the space of holomorphic co-vectors at* p *while* $T_p^*M^{0,1}$ *is the space of antiholomorphic co-vector at* p.

We now choose a holomorphic chart $\psi : U \to \mathbb{C}^n$ at p. Writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have the 1-forms

$$dz^k = dx^k + \mathrm{i}dy^k$$

and

$$d\bar{z}^k = dx^k - \mathrm{i}dy^k.$$

Equivalently, the pointwise definitions are $dz^k\big|_p = dx^k\big|_p + \mathrm{i}\ dy^k\big|_p$ and $d\bar{z}^k\big|_p = dx^k\big|_p - \mathrm{i}\ dy^k\big|_p$. Notice that we have the expected relations:

$$
\begin{aligned}
dz^k(\frac{\partial}{\partial z^i}) &= (dx^k + \mathrm{i}dy^k)(\frac{1}{2}\frac{\partial}{\partial x^i} - \mathrm{i}\frac{1}{2}\frac{\partial}{\partial y^i}) \\
&= \frac{1}{2}\delta^k_j + \frac{1}{2}\delta^k_j = \delta^k_j \\
dz^k(\frac{\partial}{\partial \bar{z}^i}) &= (dx^k + \mathrm{i}dy^k)(\frac{1}{2}\frac{\partial}{\partial x^i} + \mathrm{i}\frac{1}{2}\frac{\partial}{\partial y^i}) \\
&= 0 \\
&\text{and similarly} \\
d\bar{z}^k(\frac{\partial}{\partial \bar{z}^i}) &= \delta^k_j \text{ and } d\bar{z}^k(\frac{\partial}{\partial z^i}) = \delta^k_j.
\end{aligned}
$$

Let us check the action of J^* on these forms:

$$
\begin{aligned}
J^*(dz^k)(\frac{\partial}{\partial z^i}) &= J^*(dx^k + \mathrm{i}dy^k)(\frac{\partial}{\partial z^i}) \\
&= (dx^k + \mathrm{i}dy^k)(J\frac{\partial}{\partial z^i}) \\
&= \mathrm{i}(dx^k + \mathrm{i}dy^k)\frac{\partial}{\partial z^i} \\
&= \mathrm{i}dz^k(\frac{\partial}{\partial z^i})
\end{aligned}
$$

and

$$
\begin{aligned}
J^*(dz^k)(\frac{\partial}{\partial \bar{z}^i}) &= dz^k(J\frac{\partial}{\partial \bar{z}^i}) \\
&= -\mathrm{i}dz^k(\frac{\partial}{\partial \bar{z}^i}) = 0 = \\
&= \mathrm{i}dz^k(\frac{\partial}{\partial \bar{z}^i}).
\end{aligned}
$$

Thus we conclude that $dz^k\big|_p \in T^*_pM^{1,0}$. A similar calculation shows that $d\bar{z}^k\big|_p \in T^*_pM^{0,1}$ and in fact

$$
\begin{aligned}
T^*_pM^{1,0} &= \operatorname{span}\left\{ dz^k\big|_p : k = 1, ..., n \right\} \\
T^*_pM^{0,1} &= \operatorname{span}\left\{ d\bar{z}^k\big|_p : k = 1, ..., n \right\}
\end{aligned}
$$

and $\{dz^1\big|_p, ..., dz^n|_p, d\bar{z}^1\big|_p, ..., d\bar{z}^n|_p\}$ is a basis for $T^*_pM_{\mathbb{C}}$.

Remark 21.1 *If we don't specify base points then we are talking about fields (over some open set) which form a basis for each fiber separately. These are called frame fields (e.g.* $\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^i}$*) or co-frame fields (e.g.* $dz^k, d\bar{z}^k$*).*

21.6 Examples

21.7 The holomorphic inverse and implicit functions theorems.

Let $(z^1, ..., z^n)$ and $(w^1, ..., w^m)$ be local coordinates on complex manifolds M and N respectively. Consider a smooth map $f : M \to N$. We suppose that $p \in M$ is in the domain of $(z^1, ..., z^n)$ and that $q = f(p)$ is in the domain of the coordinates $(w^1, ..., w^m)$. Writing $z^i = x^i + \mathrm{i}y^i$ and $w^i = u^i + \mathrm{i}v^i$ we have the following Jacobian matrices:

1. In we consider the underlying real structures then we have the Jacobian given in terms of the frame $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ and $\frac{\partial}{\partial u^i}, \frac{\partial}{\partial v^i}$

$$J_p(f) = \begin{bmatrix} \frac{\partial u^1}{\partial x^1}(p) & \frac{\partial u^1}{\partial y^1}(p) & \frac{\partial u^1}{\partial x^2}(p) & \frac{\partial u^1}{\partial y^2}(p) & \cdots \\ \frac{\partial v^1}{\partial x^1}(p) & \frac{\partial v^1}{\partial y^1}(p) & \frac{\partial v^1}{\partial x^2}(p) & \frac{\partial v^1}{\partial y^2}(p) & \cdots \\ \frac{\partial u^2}{\partial x^1}(p) & \frac{\partial u^2}{\partial y^1}(p) & \frac{\partial u^2}{\partial x^2}(p) & \frac{\partial u^2}{\partial y^2}(p) & \cdots \\ \frac{\partial v^2}{\partial x^1}(p) & \frac{\partial v^2}{\partial y^1}(p) & \frac{\partial v^2}{\partial x^2}(p) & \frac{\partial v^2}{\partial y^2}(p) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

2. With respect to the bases $\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^i}$ and $\frac{\partial}{\partial w^i}, \frac{\partial}{\partial \bar{w}^i}$ we have

$$J_{p,\mathbb{C}}(f) = \begin{bmatrix} J_{11} & J_{12} & \cdots \\ J_{12} & J_{22} & \\ \vdots & & \end{bmatrix}$$

where the J_{ij} are blocks of the form

$$\begin{bmatrix} \frac{\partial w^i}{\partial z^j} & \frac{\partial w^i}{\partial \bar{z}^j} \\ \frac{\partial \bar{w}^i}{\partial z^j} & \frac{\partial \bar{w}^i}{\partial \bar{z}^j} \end{bmatrix}.$$

If f is holomorphic then these block reduce to the form

$$\begin{bmatrix} \frac{\partial w^i}{\partial z^j} & 0 \\ 0 & \frac{\partial \bar{w}^i}{\partial \bar{z}^j} \end{bmatrix}.$$

It is convenient to put the frame fields in the order $\frac{\partial}{\partial z^1}, ..., \frac{\partial}{\partial z^n}, \frac{\partial}{\partial \bar{z}^1}, ..., \frac{\partial}{\partial \bar{z}^n}$ and similarly for the $\frac{\partial}{\partial w^i}, \frac{\partial}{\partial \bar{w}^i}$. In this case we have for holomorphic f

$$\mathcal{J}_{p,\mathbb{C}}(f) = \begin{bmatrix} J^{1,0} & 0 \\ 0 & \overline{J^{1,0}} \end{bmatrix}$$

where

$$J^{1,0}(f) = \left[\frac{\partial w^i}{\partial z^j}\right]$$
$$\overline{J^{1,0}}(f) = \left[\frac{\partial \bar{w}^i}{\partial \bar{z}^j}\right].$$

We shall call a basis arising from a holomorphic coordinate system "separated" when arranged this way. Note that $J^{1,0}$ is just the Jacobian of the holomorphic tangent map $T^{1,0}f : T^{1,0}M \to T^{1,0}N$ with respect to this the **holomorphic frame** $\frac{\partial}{\partial z^1}, ..., \frac{\partial}{\partial z^n}$.

We can now formulate the following version of the inverse mapping theorem:

Theorem 21.1 *(1) Let U and V be open set in $\mathbb{C}^n$ and suppose that the map $f : U \to V$ is holomorphic with $J^{1,0}(f)$ nonsingular at $p \in U$. Then there exists an open set $U_0 \subset U$ containing p such that $f|_{U_0} : U_0 \to f(U_0)$ is a 1-1 holomorphic map with holomorphic inverse. That is, $f|_{U_0}$ is **biholomorphic**.*

(2) Similarly, if $f : U \to V$ is holomorphic map between open sets of complex manifolds M and N then if $T_p^{1,0}f : T_p^{1,0}M \to T_{fp}^{1,0}N$ is a linear isomorphism then f is a biholomorphic map when restricted to a possibly smaller open set containing p.

We also have a holomorphic version of the implicit mapping theorem.

Theorem 21.2 *(1) Let $f : U \subset \mathbb{C}^n \to V \subset \mathbb{C}^k$ and let the component functions of f be $f_1, ..., f_k$. If $J_p^{1,0}(f)$ has rank k then there are holomorphic functions $g^1, g^2, ..., g^k$ defined near $0 \in \mathbb{C}^{n-k}$ such that*

$$f(z^1, ..., z^n) = p$$
$$\Leftrightarrow$$
$$z^j = g^j(z^{k+1}, ..., z^n) \text{ for } j = 1, .., k$$

(2) If $f : M \to N$ is a holomorphic map of complex manifolds and if for fixed $q \in N$ we have that each $p \in f^{-1}(p)$ is regular in the sense that $T_p^{1,0}f : T_p^{1,0}M \to T_{fp}^{1,0}N$ is surjective, then $S := f^{-1}(p)$ is a complex submanifold of (complex) dimension $n-k$.

Example 21.5 *The map $\varphi : \mathbb{C}^{n+1} \to \mathbb{C}$ given by $(z^1, ..., z^{n+1}) \mapsto (z^1)^2 + \cdots + (z^{n+1})^2$ has Jacobian at any $(z^1, ..., z^{n+1})$ given by*

$$\left[\begin{array}{cccc} 2z^1 & 2z^2 & \cdots & 2z^{n+1} \end{array}\right]$$

*which has rank 1 as long as $(z^1, ..., z^{n+1}) \neq 0$. Thus $\varphi^{-1}(1)$ is a complex submanifold of $\mathbb{C}^{n+1}$ having (complex) dimension n. **Warning**: This is not the same as the sphere given by $\left|z^1\right|^2 + \cdots + \left|z^{n+1}\right|^2 = 1$ which is a real submanifold of $\mathbb{C}^{n+1} \cong \mathbb{R}^{2n+2}$ of real dimension $2n+1$.*

Chapter 22

Classical Mechanics

Every body continues in its state of rest or uniform motion in a straight line, except insofar as it doesn't.
Arthur Eddington, Sir

22.1 Particle motion and Lagrangian Systems

If we consider a *single* particle of mass m then Newton's law is

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F}(\mathbf{x}(t), t)$$

The force $\mathbf{F}$ is **conservative** if it doesn't depend on time and there is a potential function $V : \mathbb{R}^3 \to \mathbb{R}$ such that $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$. Assume this is the case. Then Newton's law becomes

$$m\frac{d^2}{dt^2}\mathbf{x}(t) + \operatorname{grad} V(\mathbf{x}(t)) = 0.$$

Consider the Affine space $C(I, \mathbf{x}_1, \mathbf{x}_2)$ of all C^2 paths from $\mathbf{x}_1$ to $\mathbf{x}_2$ in $\mathbb{R}^3$ defined on the interval $I = [t_1, t_2]$. This is an Affine space modelled on the Banach space $C_0^r(I)$ of all C^r functions $\varepsilon : I \to \mathbb{R}^3$ with $\varepsilon(t_1) = \varepsilon(t_1) = 0$ and with the norm

$$\|\varepsilon\| = \sup_{t\in I}\{|\varepsilon(t)| + |\varepsilon'(t)| + |\varepsilon''(t)|\}.$$

If we define the fixed affine linear path $\mathbf{a} : I \to \mathbb{R}^3$ by $\mathbf{a}(t) = \mathbf{x}_1 + \frac{t-t_1}{t_2-t_1}(\mathbf{x}_2 - \mathbf{x}_1)$ then all we have a coordinatization of $C(I, \mathbf{x}_1, \mathbf{x}_2)$ by $C_0^r(I)$ given by the single chart $\psi : \mathbf{c} \mapsto \mathbf{c} - \mathbf{a} \in C_0^2(I)$. Then the tangent space to $C(I, \mathbf{x}_1, \mathbf{x}_2)$ at a fixed path c_0 is just $C_0^2(I)$. Now we have the function S defined on $C(I, \mathbf{x}_1, \mathbf{x}_2)$ by

$$S(\mathbf{c}) = \int_{t_1}^{t_2}\left(\frac{1}{2}m\left\|\mathbf{c}'(t)\right\|^2 - V(\mathbf{c}(t))\right)dt$$

The variation of S is just the 1-form $\delta S : C_0^2(I) \to \mathbb{R}$ defined by

$$\delta S \cdot \varepsilon = \left.\frac{d}{d\tau}\right|_{\tau=0} S(\mathbf{c}_0 + \tau\varepsilon)$$

Let us suppose that $\delta S = 0$ at c_0. Then we have

$$\begin{aligned}
0 &= \left.\frac{d}{d\tau}\right|_{\tau=0} S(\mathbf{c}_0'(t) + \tau\varepsilon) \\
&= \left.\frac{d}{d\tau}\right|_{\tau=0} \int_{t_1}^{t_2} \left(\frac{1}{2}m \left\|\mathbf{c}_0'(t) + \tau\varepsilon'(t)\right\|^2 - V(\mathbf{c}_0(t) + \tau\varepsilon(t))\right) dt \\
&= \int_{t_1}^{t_2} \left(m \left\langle \mathbf{c}_0'(t), \varepsilon'(t)\right\rangle - \frac{\partial V}{\partial x^i}(\mathbf{c}_0)\frac{d\varepsilon^i}{dt}(0)\right) dt \\
&= \int_{t_1}^{t_2} \left(m\mathbf{c}_0'(t) \cdot \frac{d}{dt}\varepsilon(t) - \operatorname{grad} V(\mathbf{c}_0(t)) \cdot \varepsilon(t)\right) dt \\
&\int_{t_1}^{t_2} (m\mathbf{c}_0''(t) - \operatorname{grad} V(\mathbf{c}_0(t))) \cdot \varepsilon(t)dt
\end{aligned}$$

Now since this is true for every choice of $\varepsilon \in C_0^2(I)$ we see that

$$m\mathbf{c}_0''(t) - \operatorname{grad} V(\mathbf{c}_0(t)) = 0$$

thus we see that $\mathbf{c}_0(t) = \mathbf{x}(t)$ is a critical point in $C(I, \mathbf{x}_1, \mathbf{x}_2)$, that is, a stationary path, iff **??** is satisfied.

22.1.1 Basic Variational Formalism for a Lagrangian

In general we consider a differentiable manifold Q as our state space and then a Lagrangian density function L is given on TQ. For example we can take a potential function $V : Q \to \mathbb{R}$, a Riemannian metric g on Q and define the action functional S on the space of smooth paths $I \to Q$ beginning and ending at a fixed points p_1 and p_2 given by

$$\begin{aligned}
S(c) = \int_{t_1}^{t_2} \mathrm{L}(c'(t))dt = \\
\int_{t_1}^{t_2} \frac{1}{2}m \left\langle c'(t), c'(t)\right\rangle - V(c(t))dt
\end{aligned}$$

The tangent space at a fixed c_0 is the Banach space $\Gamma_0^2(c_0^*TQ)$ of C^2 vector fields $\varepsilon : I \to TQ$ along c_0 which vanish at t_1 and t_2. A curve with tangent ε at c_0 is just a variation $v : (-\epsilon, \epsilon) \times I \to Q$ such that $\varepsilon(t) = \left.\frac{\partial}{\partial s}\right|_{s=0} v(s,t)$ is the variation vector field. Then we have

$$\begin{aligned}
\delta S \cdot \varepsilon &= \left.\frac{\partial}{\partial s}\right|_{s=0} \int_{t_1}^{t_2} \mathrm{L}\left(\frac{\partial v}{\partial t}(s,t)\right) dt \\
&= \int_{t_1}^{t_2} \left.\frac{\partial}{\partial s}\right|_{s=0} \mathrm{L}\left(\frac{\partial v}{\partial t}(s,t)\right) dt \\
&= etc.
\end{aligned}$$

Let us examine this in the case of $Q = U \subset \mathbb{R}^n$. With $\mathbf{q} = (q^1, ...q^n)$ being (general curvilinear) coordinates on U we have natural (tangent bundle chart) coordinates $\mathbf{q}, \dot{\mathbf{q}}$ on $TU = U \times \mathbb{R}^n$. Assume that the variation has the form $\mathbf{q}(s,t) = \mathbf{q}(t) + s\varepsilon(t)$. Then we have

$$\begin{aligned}
\delta S \cdot \varepsilon &= \left.\frac{\partial}{\partial s}\right|_{s=0} \int_{t_1}^{t_2} \mathrm{L}(\mathbf{q}(s,t), \dot{\mathbf{q}}(s,t)) dt \\
&\int_{t_1}^{t_2} \left.\frac{\partial}{\partial s}\right|_{s=0} \mathrm{L}(\mathbf{q} + s\varepsilon, \dot{\mathbf{q}} + s\dot{\varepsilon}) dt \\
&= \int_{t_1}^{t_2} \left(\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \varepsilon + \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\varepsilon}\right) dt \\
&= \int_{t_1}^{t_2} \left(\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \varepsilon - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \varepsilon\right) dt \\
&= \int_{t_1}^{t_2} \left(\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}})\right) \cdot \varepsilon dt
\end{aligned}$$

and since ε was arbitrary we get the Euler-Lagrange equations for the motion

$$\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) = 0$$

In general, a time-independent Lagrangian on a manifold Q (usually representing the position space of a particle system) is a smooth function on the tangent bundle (velocity space):

$$\mathrm{L} : TQ \to Q$$

and the associated action functional is a map from the space of smooth curves $C^\infty([a,b], Q)$ defined for $c : [a,b] \to Q$ by

$$\mathcal{S}_L(c) = \int_a^b \mathrm{L}(\dot{c}(t)) dt$$

where $\dot{c} : [a,b] \to TQ$ is the canonical lift (velocity). A time dependent Lagrangian is a smooth map

$$L : \mathbb{R} \times TQ \to Q$$

where the first factor $\mathbb{R}$ is the time t, and once again we have the associated action functional $\mathcal{S}_L(c) = \int_a^b L(t, \dot{c}(t)) dt$.

Let us limit ourselves initially to the time independent case.

Definition 22.1 *A smooth variation of a curve* $c : [a, b] \to Q$ *is a smooth map* $\nu : [a, b] \times (-\epsilon, \epsilon) \to Q$ *for small* ϵ *such that* $\nu(t, 0) = c(t)$*. We call the variation a* ***variation with fixed endpoints*** *if* $\nu(a, s) = c(a)$ *and* $\nu(b, s) = c(b)$ *for all* $s \in (-\epsilon, \epsilon)$*. Now we have a family of curves* $\nu_s = \nu(., s)$*. The infinitesimal variation at* ν_0 *is the vector field along* c *defined by* $V(t) = \frac{d\nu}{ds}(t, 0)$*. This* V *is called the* ***variation vector field*** *for the variation. The differential of the functional* δS_L *(classically called the first variation) is defined as*

$$\begin{aligned} \delta S_{\mathrm{L}}(c) \cdot V &= \left.\frac{d}{ds}\right|_{s=0} S_{\mathrm{L}}(\nu_s) \\ &= \left.\frac{d}{ds}\right|_{s=0} \int_a^b \mathrm{L}(\nu_s(t))dt \end{aligned}$$

Remark 22.1 *Every smooth vector field along* c *is the variational vector field coming from some variation of* c *and for any other variation* ν' *with* $V(t) = \frac{d\nu'}{ds}(t, 0)$ *the about computed quantity* $\delta S_L(c) \cdot V$ *will be the same.*

At any rate, if $\delta S_L(c) \cdot V = 0$ for all variations vector fields V along c and vanishing at the endpoints then we write $\delta S_L(c) = 0$ and call c **critical** (or **stationary**) for L.

Now consider the case where the image of c lies in some coordinate chart $U, \psi = q^1, q^2, ...q^n$ and denote by $TU, T\psi = (q^1, q^2, ..., q^n, \dot{q}^1, \dot{q}^2, ..., \dot{q}^n)$ the natural chart on $TU \subset TQ$. In other words, $T\psi(\xi) = (q^1 \circ \tau(\xi), q^2 \circ \tau(\xi), ..., q^n \circ \tau(\xi), dq^1(\xi), dq^2(\xi), ..., dq^n(\xi))$. Thus the curve has coordinates

$$(c, \dot{c}) = (q^1(t), q^2(t), ..., q^n(t), \dot{q}^1(t), \dot{q}^2(t), ..., \dot{q}^n(t))$$

where now the $\dot{q}^i(t)$ *really are* time derivatives. In this local situation we can choose our variation to have the form $q^i(t) + s\delta q^i(t)$ for some functions $\delta q^i(t)$ vanishing at a and b and some parameter s with respect to which we will differentiate. The lifted variation is $(\mathbf{q}(t) + s\delta(t), \dot{\mathbf{q}}(t) + s\delta\dot{\mathbf{q}}(t))$ which is the obvious abbreviation for a path in $T\psi(TU) \subset \mathbb{R}^n \times \mathbb{R}^n$. Now we have seen above that the path c will be critical if

$$\left.\frac{d}{ds}\right|_{s=0} \int \mathrm{L}(\mathbf{q}(t) + s\delta(t), \dot{\mathbf{q}}(t) + s\delta\dot{\mathbf{q}}(t)))dt = 0$$

for all such variations and the above calculations lead to the result that

$$\frac{\partial}{\partial \mathbf{q}} \mathrm{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) - \frac{d}{dt}\frac{\partial}{\partial \dot{\mathbf{q}}} \mathrm{L}(\mathbf{q}(t), \dot{\mathbf{q}}(\mathbf{t})) = \mathbf{0} \quad \text{Euler-Lagrange}$$

for any L-critical path (with image in this chart). Here $n = \dim(Q)$.

It can be show that even in the case that the image of c does not lie in the domain of a chart that c is L-critical path if it can be subdivided into sub-paths lying in charts and L-critical in each such chart.

22.1.2 Two examples of a Lagrangian

Example 22.1 *Suppose that we have a 1-form $\theta \in \mathfrak{X}^*(Q)$. A 1-form is just a map $\theta : TQ \to \mathbb{R}$ that happens to be linear on each fiber T_pQ. Thus we may examine the special case of $\mathrm{L} = \theta$. In canonical coordinates $(q, \dot{q})$ again,*

$$\mathrm{L} = \theta = \sum a_i(q)dq^i$$

for some functions $a_i(q)$. An easy calculation shows that the Euler-Lagrange equations become

$$\left(\frac{\partial a_i}{\partial q^k} - \frac{\partial a_k}{\partial q^i}\right)\dot{q}^i = 0$$

but on the other hand

$$d\theta = \frac{1}{2}\sum\left(\frac{\partial a_j}{\partial q^i} - \frac{\partial a_i}{\partial q^j}\right)\partial q^i \wedge \partial q^j$$

and one can conclude that if $c = (q^i(t))$ is critical for $L = \theta$ then for any vector field X defined on the image of c we have

$$\frac{1}{2}\sum\left(\frac{\partial a_j}{\partial q^i}(q^i(t)) - \frac{\partial a_i}{\partial q^j}(q^i(t))\right)\dot{q}^i(t)X^j$$

or $d\theta(\dot{c}(t), X) = 0$. This can be written succinctly as

$$\iota_{\dot{c}(t)}d\theta = 0.$$

Example 22.2 *Now let us take the case of a Riemannian manifold M, g and let $\mathrm{L}(v) = \frac{1}{2}\mathrm{g}(v, v)$. Thus the action functional is the "energy"*

$$S_g(c) = \int g(\dot{c}(t), \dot{c}(t))dt$$

In this case the critical paths are just geodesics.

22.2 Symmetry, Conservation and Noether's Theorem

Let G be a Lie group acting on a smooth manifold M.

$$\lambda : G \times M \to M$$

As usual we write $g \cdot x$ for $\lambda(g, x)$. We have a fundamental vector field $\xi^\natural$ associated to every $\xi \in \mathfrak{g}$ defined by the rule

$$\xi^\natural(p) = T_{(e,p)}\lambda \cdot (., 0)$$

or equivalently by the rule

$$\xi^\natural(p) = \left.\frac{d}{dt}\right|_0 \exp(t\xi) \cdot p$$

The map $\natural : \xi \mapsto \xi^\natural$ is a Lie algebra anti-homomorphism. Of course, here we are using the flow associated to ξ

$$\mathrm{Fl}^\xi(t,p) := \mathrm{Fl}^{\xi^\natural}(t,p) = \exp(t\xi) \cdot p$$

and it should be noted that $t \mapsto \exp(t\xi)$ is the one parameter subgroup associated to ξ and to get the corresponding left invariant vector field $X^\xi \in \mathfrak{X}^L(G)$ we act on the right:

$$X^\xi(g) = \left.\frac{d}{dt}\right|_0 g \cdot \exp(t\xi)$$

Now a diffeomorphism acts on a covariant k-tensor field contravariantly according to

$$(\phi^* K)(p)(v_1, ...v_k) = K(\phi(p))(T\phi v_1, ...T\phi v_k)$$

Suppose that we are given a covariant tensor field $\Upsilon \in \mathfrak{T}(M)$ on M. We think of Υ as defining some kind of extra structure on M. The two main examples for our purposes are

1. $\Upsilon = \langle .,. \rangle$ a nondegenerate covariant symmetric 2-tensor. Then $M, \langle .,. \rangle$ is a (semi-) Riemannian manifold.

2. $\Upsilon = \omega \in \Omega^2(M)$ a non-degenerate 2-form. Then M, ω is a **symplectic manifold**.

Then G acts on Υ since G acts on M as diffeomorphisms. We denote this natural (left) action by $g \cdot \Upsilon$. If $g \cdot \Upsilon = \Upsilon$ for all $g \in G$ we say that G acts by symmetries of the pair M, Υ.

Definition 22.2 *In general, a vector field X on M, Υ is called an **infinitesimal symmetry** of the pair M, Υ if $\mathcal{L}_X \Upsilon = 0$. Other terminology is that X is a $\Upsilon-$**Killing field** . The usual notion of a Killing field in (pseudo-) Riemannian geometry is the case when $\Upsilon = \langle , \rangle$ is the metric tensor.*

Example 22.3 *A group G is called a symmetry group of a symplectic manifold M, ω if G acts by symplectomorphisms so that $g \cdot \omega = \omega$ for all $g \in G$. In this case, each $\xi \in \mathfrak{g}$ is an infinitesimal symmetry of M, ω meaning that*

$$\mathcal{L}_\xi \omega = 0$$

where $\mathcal{L}_\xi$ is by definition the same as $\mathcal{L}_{\xi^\natural}$. This follows because if we let $g_t = \exp(t\xi)$ then each g_t is a symmetry so $g_t^* \omega = 0$ and

$$\mathcal{L}_\xi \omega = \left.\frac{d}{dt}\right|_0 g_t^* \omega = 0$$

22.2.1 Lagrangians with symmetries.

We need two definitions

Definition 22.3 *If $\phi : M \to M$ is a diffeomorphism then the induced tangent map $T\phi : TM \to TM$ is called the **canonical lift**.*

Definition 22.4 *Given a vector field $X \in \mathfrak{X}(M)$ there is a lifting of X to $\widetilde{X} \in \mathfrak{X}(TM) = \Gamma(TM, TTM)$*

$$\begin{array}{lll} \widetilde{X}: & TM \to & TTM \\ & \downarrow & \downarrow \\ X: & M \to & TM \end{array}$$

such that the flow $\mathrm{Fl}^{\widetilde{X}}$ is the canonical lift of Fl^{X}

$$\begin{array}{lll} \mathrm{Fl}_t^{\widetilde{X}}: & TM \to & TM \\ & \downarrow & \downarrow \\ \mathrm{Fl}_t^{X}: & M \to & M \end{array}.$$

In other words, $\mathrm{Fl}^{\widetilde{X}} = T\mathrm{Fl}_t^X$. We simply define $\widetilde{X}(v) = \frac{d}{dt}(T\mathrm{Fl}_t^X \cdot v)$.

Definition 22.5 *Let ω_{L} denote the unique 1-form on Q which in canonical coordinates is $\omega_{\mathrm{L}} = \sum_{i=1}^{n} \frac{\partial \mathrm{L}}{\partial \dot{q}^i} dq^i$.*

Theorem 22.1 (E. Noether) *If X is an infinitesimal symmetry of the Lagrangian then the function $\omega_{\mathrm{L}}(\widetilde{X})$ is constant along any path $c : I \subset \mathbb{R}$ that is stationary for the action associated to L.*

Let's prove this using local coordinates $(q^i, \dot{q}^i)$ for $TU_\alpha \subset TQ$. It turn out that locally,

$$\widetilde{X} = \sum_i \left(a^i \frac{\partial}{\partial q^i} + \sum_j \frac{\partial a^i}{\partial q^j} \frac{\partial}{\partial \dot{q}^i} \right)$$

where a^i is defined by $X = \sum a^i(q) \frac{\partial}{\partial q^i}$. Also, $\omega_L(\widetilde{X}) = \sum a^i \frac{\partial L}{\partial \dot{q}^i}$. Now suppose that $q^i(t), \dot{q}^i(t) = \frac{d}{dt} q^i(t)$ satisfies the Euler-Lagrange equations. Then

$$\begin{aligned}
&\frac{d}{dt}\omega_{\mathrm{L}}(\widetilde{X})(q^i(t), \dot{q}^i(t)) \\
&= \frac{d}{dt} \sum a^i(q^i(t)) \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) \\
&= \sum \frac{da}{dt}^i (q^i(t)) \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) + a^i(q^i(t)) \frac{d}{dt} \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) \\
&= \sum_i \left[\sum_j \frac{da}{dq^j}^i \dot{q}^j(t) \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) + a^i(q^i(t)) \frac{\partial \mathrm{L}}{\partial q^i}(q^i(t), \dot{q}^i(t)) \right] \\
&= d\mathrm{L}(X) = \mathcal{L}_X \mathrm{L} = 0
\end{aligned}$$

This theorem tells us one case when we get a conservation law. A conservation law is a function C on TQ (or T^*Q for the Hamiltonian flow) such that C is constant along solution paths. (i.e. stationary for the action or satisfying the E-L eqns.)

$$\mathrm{L} : TQ \to Q$$

let $X \in T(TQ)$.

22.2.2 Lie Groups and Left Invariants Lagrangians

Recall that G act on itself by left translation $l_g : G \to G$. The action lifts to the tangent bundle $Tl_g : TG \to TG$. Suppose that $\mathrm{L} : TG \to \mathbb{R}$ is invariant under this left action so that $\mathrm{L}(Tl_g X_h) = \mathrm{L}(X_p)$ for all $g, h \in G$. In particular, $\mathrm{L}(Tl_g X_e) = L(X_e)$ so L is completely determined by its restriction to $T_eG = \mathfrak{g}$. Define the restricted Lagrangian function by $\Lambda = \mathrm{L}|_{T_eG}$. We view the differential $d\Lambda$ as a map $d\Lambda : \mathfrak{g} \to \mathbb{R}$ and so in fact $d\lambda \in \mathfrak{g}^*$. Next, recall that for any $\xi \in \mathfrak{g}$ the map $\mathrm{ad}_\xi : \mathfrak{g} \to \mathfrak{g}$ is defined by $\mathrm{ad}_\xi \upsilon = [\xi, \upsilon]$ and we have the adjoint map $\mathrm{ad}^*_\xi : \mathfrak{g}^* \to \mathfrak{g}^*$. Now let $t \mapsto g(t)$ be a motion of the system and define the "body velocity" by $\nu_c(t) = Tl_{c(t)^{-1}} \cdot c'(t) = \omega_G(c'(t))$. Then we have

Theorem 22.2 *Assume L is invariant as above. The curve $c(.)$ satisfies the Euler-Lagrange equations for* L *iff*

$$\frac{d}{dt} d\Lambda(\nu_c(t)) = \mathrm{ad}^*_{\nu_c(t)} d\Lambda$$

22.3 The Hamiltonian Formalism

Let us now examine the change to a description in cotangent chart $\mathbf{q}, \mathbf{p}$ so that for a covector at $\mathbf{q}$ given by $\mathbf{a}(\mathbf{q}) \cdot d\mathbf{q}$ has coordinates $\mathbf{q}, \mathbf{a}$. Our method of transfer to the cotangent side is via the Legendre transformation induced by L. In fact, this is just the fiber derivative defined above. We must assume that the map $F : (\mathbf{q}, \dot{\mathbf{q}}) \mapsto (\mathbf{q}, \mathbf{p}) = (\mathbf{q}, \frac{\partial L}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}))$ is a diffeomorphism (this is written with respect to the two natural charts on TU and T^*U). Again this just means that the Lagrangian is nondegenerate. Now if $v(t) = (\mathbf{q}(t), \dot{\mathbf{q}}(t))$ is (a lift of) a solution curve then defining the **Hamiltonian** function

$$\widetilde{H}(\mathbf{q}, \dot{\mathbf{q}}) = \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}} - \mathrm{L}(\mathbf{q}, \dot{\mathbf{q}})$$

we compute with $\dot{\mathbf{q}} = \frac{d}{dt}\mathbf{q}$

$$\begin{aligned}
\frac{d}{dt}\widetilde{H}(\mathbf{q}, \dot{\mathbf{q}}) &= \frac{d}{dt}(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}}) - \frac{d}{dt}\mathrm{L}(\mathbf{q}, \dot{\mathbf{q}}) \\
&= \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \frac{d}{dt}\dot{\mathbf{q}} + \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}} - \frac{d}{dt}\mathrm{L}(\mathbf{q}, \dot{\mathbf{q}}) \\
&= 0
\end{aligned}$$

we have used that the Euler-Lagrange equations $\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q},\dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q},\dot{\mathbf{q}}) = 0$. Thus differential form $d\widetilde{H} = \frac{\partial \widetilde{H}}{\partial \mathbf{q}} d\mathbf{q} + \frac{\partial \widetilde{H}}{\partial \dot{\mathbf{q}}} d\dot{\mathbf{q}}$ is zero on the velocity $v'(t) = \frac{d}{dt}(\mathbf{q},\dot{\mathbf{q}})$

$$\begin{aligned} d\widetilde{H} \cdot v'(t) &= d\widetilde{H} \cdot \frac{d}{dt}(\mathbf{q},\dot{\mathbf{q}}) \\ &= \frac{\partial \widetilde{H}}{\partial \mathbf{q}} \frac{d\mathbf{q}}{dt} + \frac{\partial \widetilde{H}}{\partial \dot{\mathbf{q}}} \frac{d\dot{\mathbf{q}}}{dt} = 0 \end{aligned}$$

We then use the inverse of this diffeomorphism to transfer the Hamiltonian function to a function $H(\mathbf{q},\mathbf{p}) = F^{-1*}\widetilde{H}(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{p} \cdot \dot{\mathbf{q}}(\mathbf{q},\mathbf{p}) - \mathrm{L}(\mathbf{q},\dot{\mathbf{q}}(\mathbf{q},\mathbf{p}))$.. Now if $\mathbf{q}(t), \dot{\mathbf{q}}(t)$ is a solution curve then its image $b(t) = F \circ v(t) = (\mathbf{q}(t),\mathbf{p}(t))$ satisfies

$$\begin{aligned} dH(b'(t)) &= (dH \cdot TF.v'(t)) \\ &= (F^* dH) \cdot v'(t) \\ &= d(F^* H) \cdot v'(t) \\ &= d\widetilde{H} \cdot v'(t) = 0 \end{aligned}$$

so we have that

$$0 = dH(b'(t)) = \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{d\mathbf{q}}{dt} + \frac{\partial H}{\partial \mathbf{p}} \cdot \frac{d\mathbf{p}}{dt}$$

but also

$$\frac{\partial}{\partial \mathbf{p}} H(\mathbf{q},\mathbf{p}) = \dot{\mathbf{q}} + \mathbf{p} \cdot \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{p}} - \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{p}} = \dot{\mathbf{q}} = \frac{d\mathbf{q}}{dt}$$

solving these last two equations simultaneously we arrive at Hamilton's equations of motion:

$$\begin{aligned} \frac{d}{dt}\mathbf{q}(t) &= \frac{\partial H}{\partial \mathbf{p}}(\mathbf{q}(t),\mathbf{p}(t)) \\ \frac{d}{dt}\mathbf{p}(t) &= -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{q}(t),\mathbf{p}(t)) \end{aligned}$$

or

$$\frac{d}{dt}\begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{pmatrix} \frac{\partial H}{\partial \mathbf{q}} \\ \frac{\partial H}{\partial \mathbf{p}} \end{pmatrix}$$

Remark 22.2 *One can calculate directly that* $\frac{dH}{dt}(\mathbf{q}(t),\mathbf{p}(t)) = 0$ *for solutions these equations. If the Lagrangian was originally given by* $\mathrm{L} = \frac{1}{2}K - V$ *for some kinetic energy function and a potential energy function then this amounts to conservation of energy. We will see that this follows from a general principle below.*

Chapter 23

Symplectic Geometry

Equations are more important to me, because politics is for the present, but an equation is something for eternity

-Einstein

23.1 Symplectic Linear Algebra

A (real) **symplectic vector space** is a pair V, α where V is a (real) vector space and α is a nondegenerate alternating (skew-symmetric) bilinear form $\alpha : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$. The basic example is $\mathbb{R}^{2n}$ with

$$\alpha_0(x, y) = x^t J_n y$$

where

$$J_n = \begin{pmatrix} 0 & I_{n\times n} \\ -I_{n\times n} & 0 \end{pmatrix}.$$

The standard symplectic form on α_0 is typical. It is a standard fact from linear algebra that for any N dimensional symplectic vector space V, α there is a basis $e_1, ..., e_n, f^1, ..., f^n$ called a **symplectic basis** such that the matrix that represents α with respect to this basis is the matrix J_n. Thus we may write

$$\alpha = e^1 \wedge f_1 + ... + e^n \wedge f_n$$

where $e^1, ..., e^n, f_1, ..., f_n$ is the dual basis to $e_1, ..., e_n, f^1, ..., f^n$. If V, η is a vector space with a not necessarily nondegenerate alternating form η then we can define the null space

$$N_\eta = \{v \in \mathrm{V} : \eta(v, w) = 0 \text{ for all } w \in \mathrm{V}\}.$$

On the quotient space $\overline{\mathrm{V}} = \mathrm{V}/N_\eta$ we may define $\overline{\eta}(\overline{v}, \overline{w}) = \eta(v, w)$ where v and w represent the elements $\overline{v}, \overline{w} \in \overline{\mathrm{V}}$. Then $\overline{\mathrm{V}}, \overline{\eta}$ is a symplectic vector space called the symplectic reduction of V, η.

Proposition 23.1 *For any $\eta \in \bigwedge \mathrm{V}^*$ (regarded as a bilinear form) there is linearly independent set of elements $e^1, ..., e^k, f_1, ..., f_k$ from V^* such that*

$$\eta = e^1 \wedge f_1 + ... + e^k \wedge f_k$$

where $\dim(\mathrm{V}) - 2k \geq 0$ *is the dimension of N_η.*

Definition 23.1 *Note: The number k is called the* ***rank*** *of η. The matrix that represents η actually has rank $2k$ and so some might call k the* ***half rank*** *of η.*

Proof. Consider the symplectic reduction $\overline{\mathrm{V}}, \overline{\eta}$ of V, η and choose set of elements $e^1, ..., e^k, f_1, ..., f_k$ such that $\bar{e}^1, ..., \bar{e}^k, \bar{f}_1, ..., \bar{f}_k$ form a symplectic basis of $\overline{\mathrm{V}}, \overline{\eta}$. Add to this set a basis $b_1, ..., b_l$ a basis for N_η and verify that $e^1, ..., e^k, f_1, ..., f_k, b_1, ..., b_l$ must be a basis for V. Taking the dual basis one can check that

$$\eta = e^1 \wedge f_1 + ... + e^k \wedge f_k$$

by testing on the basis $e^1, ..., e^k, f_1, ..., f_k, b_1, ..., b_l$. ■

Now if W is a subspace of a symplectic vector space then we may define

$$\mathrm{W}^\perp = \{v \in \mathrm{V} : \eta(v, w) = 0 \text{ for all } w \in \mathrm{W}\}$$

and it is true that $\dim(\mathrm{W}) + \dim(\mathrm{W}^\perp) = \dim(\mathrm{V})$ but it is **not** necessarily the case that $\mathrm{W} \cap \mathrm{W}^\perp = 0$. In fact,we classify subspaces W by two numbers: $d = \dim(\mathrm{W})$ and $\nu = \dim(\mathrm{W} \cap \mathrm{W}^\perp)$. If $\nu = 0$ then $\eta|_{\mathrm{W}}$, W is a symplectic space and so we call W a **symplectic subspace** . At the opposite extreme, if $\nu = d$ then W is called a **Lagrangian subspace** . If $\mathrm{W} \subset \mathrm{W}^\perp$ we say that W is an **isotropic subspace**.

A linear transformation between symplectic vector spaces $\ell : \mathrm{V}_1, \eta_1 \to \mathrm{V}_2, \eta_2$ is called a **symplectic linear map** if $\eta_2(\ell(v), \ell(w)) = \eta_1(v, w)$ for all $v, w \in \mathrm{V}_1$; In other words, if $\ell^* \eta_2 = \eta_1$. The set of all symplectic linear isomorphisms from V, η to itself is called the **symplectic group** and denoted $Sp(\mathrm{V},\eta)$. With respect to a symplectic basis $\mathcal{B}$ a symplectic linear isomorphism ℓ is represented by a matrix $A = [\ell]_{\mathcal{B}}$ that satisfies

$$A^t J\, A = J$$

where $J = J_n$ is the matrix defined above and where $2n = \dim(\mathrm{V})$. Such a matrix is called a symplectic matrix and the group of all such is called the **symplectic matrix group** and denoted $Sp(n, \mathbb{R})$. Of course if $\dim(\mathrm{V}) = 2n$ then $Sp(\mathrm{V},\eta) \cong Sp(n, \mathbb{R})$ the isomorphism depending a choice of basis. If η is a symplectic from on V with $\dim(\mathrm{V}) = 2n$ then $\eta^n \in \wedge^{2n}\mathrm{V}$ is nonzero and so orients the vector space V.

Lemma 23.1 *If $A \in Sp(n, \mathbb{R})$ then* $\det(A) = 1$.

Proof. If we use A as a linear transformation $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ then $A^*\alpha_0 = \alpha_0$ and $A^*\alpha_0^n = \alpha_0^n$ where α_0 is the standard symplectic form on $\mathbb{R}^{2n}$ and $\alpha_0^n \in \wedge^{2n}\mathbb{R}^{2n}$ is top form. Thus $\det A = 1$. ■

Theorem 23.1 (Symplectic eigenvalue theorem) *If λ is a (complex) eigenvalue of a symplectic matrix A then so is $1/\lambda$, $\bar{\lambda}$ and $1/\bar{\lambda}$.*

Proof. Let $p(\lambda) = \det(A - \lambda I)$ be the characteristic polynomial. It is easy to see that $J^t = -J$ and $JAJ^{-1} = (A^{-1})^t$. Using these facts we have

$$\begin{aligned}
p(\lambda) &= \det(J(A - \lambda I)J^{-1}) = \det(A^{-1} - \lambda I) \\
&= \det(A^{-1}(I - \lambda A)) = \det(I - \lambda A) \\
&= \lambda^{2n} \det(\frac{1}{\lambda} I - A)) = \lambda^{2n} p(1/\lambda).
\end{aligned}$$

So we have $p(\lambda) = \lambda^{2n} p(1/\lambda)$. Using this and remembering that 0 is not an eigenvalue one concludes that $1/\lambda$ and $\bar{\lambda}$ are eigenvalues of A. ■

Exercise 23.1 *With respect to the last theorem, show that λ and $1/\lambda$ have the same multiplicity.*

23.2 Canonical Form (Linear case)

Suppose one has a vector space W with dual W*. We denote the pairing between W and W* by $\langle ., . \rangle$. There is a simple way to form a symplectic form on the space Z = W × W* which we will call the **canonical symplectic form**. This is defined by

$$\Omega((v_1, \alpha_1), (v_2, \alpha_2)) := \langle \alpha_2, v_1 \rangle - \langle \alpha_1, v_2 \rangle.$$

If W is an inner product space with inner product $\langle ., . \rangle$ then we may form the canonical symplectic from on Z = W × W by the same formula. As a special case we get the standard symplectic form on $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ given by

$$\Omega((x, y), (\widetilde{x}, \widetilde{y})) = \widetilde{y} \cdot x - y \cdot \widetilde{x}.$$

23.3 Symplectic manifolds

Definition 23.2 *A **symplectic form** on a manifold M is a* nondegenerate closed *2-form $\omega \in \Omega^2(M) = \Gamma(M, T^*M)$. A **symplectic manifold** is a pair (M, ω) where ω is a symplectic form on M. If there exists a symplectic form on M we say that M has a symplectic structure or admits a symplectic structure.*

A map of symplectic manifolds, say $f : (M, \omega) \to (N, \varpi)$ is called a **symplectic map** iff $f^*\varpi = \omega$. We will reserve the term **symplectomorphism** to refer to diffeomorphisms that are symplectic maps. Notice that since a symplectic form such as ω is nondegenerate, the $2n$ form $\omega^n = \omega \wedge \cdots \wedge \omega$ is nonzero and global. Hence a symplectic manifold is orientable (more precisely, it is oriented).

Definition 23.3 *The form $\Omega_\omega = \frac{(-1)^n}{(2n)!} \omega^n$ is called the **canonical volume form** or **Liouville volume**.*

We immediately have that if $f : (M, \omega) \to (M, \omega)$ is a symplectic diffeomorphism then $f^*\Omega_\omega = \Omega_\omega$.

Not every manifold admits a symplectic structure. Of course if M does admit a symplectic structure then it must have even dimension but there are other more subtle obstructions. For example, the fact that $H^2(S^4) = 0$ can be used to show that S^4 does not admit ant symplectic structure. To see this, suppose to the contrary that ω is a closed nondegenerate 2-form on S^4. The since $H^2(S^4) = 0$ there would be a 1-form θ with $d\theta = \omega$. But then since $d(\omega \wedge \theta) = \omega \wedge \omega$ the 4-form $\omega \wedge \omega$ would be exact also and Stokes' theorem would give $\int_{S^4} \omega \wedge \omega = \int_{S^4} d(\omega \wedge \theta) = \int_{\partial S^4 = \emptyset} \omega \wedge \theta = 0$. But as we have seen $\omega^2 = \omega \wedge \omega$ is a nonzero top form so we must really have $\int_{S^4} \omega \wedge \omega \neq 0$. So in fact, S^4 does not admit a symplectic structure. We will give a more careful examination to the question of obstructions to symplectic structures but let us now list some positive examples.

Example 23.1 (surfaces) *Any orientable surface with volume form (area form) qualifies since in this case the volume ω itself is a closed nondegenerate two form.*

Example 23.2 (standard) *The form $\omega_{can} = \sum_{i=1}^n dx^i \wedge dx^{i+n}$ on $\mathbb{R}^{2n}$ is the prototypical symplectic form for the theory and makes $\mathbb{R}^n$ a symplectic manifold. (See Darboux's theorem 23.2 below)*

Example 23.3 (cotangent bundle) *We will see in detail below that the cotangent bundle of any smooth manifold has a natural symplectic structure. The symplectic form in a natural bundle chart (q, p) has the form $\omega = \sum_{i=1}^n dq^i \wedge dp_i$. (warning: some authors use $-\sum_{i=1}^n dq^i \wedge dp_i = \sum_{i=1}^n dp_i \wedge dq^i$ instead).*

Example 23.4 (complex submanifolds) *The symplectic $\mathbb{R}^{2n}$ may be considered the realification of $\mathbb{C}^n$ and then multiplication by i is thought of as a map $J : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$. We have that $\omega_{can}(v, Jv) = -|v|^2$ so that ω_{can} is nondegenerate on any complex submanifold M of $\mathbb{R}^{2n}$ and so $M, \omega_{can}|_M$ is a symplectic manifold.*

Example 23.5 (coadjoint orbit) *Let G be a Lie group. Define the coadjoint map $\mathrm{Ad}^\dagger : G \to \mathrm{GL}(\mathfrak{g}^*)$ which takes g to $\mathrm{Ad}_g^\dagger$ by*

$$\mathrm{Ad}_g^\dagger(\xi)(x) = \xi(\mathrm{Ad}_{g^{-1}}(x)).$$

The action defined by $\mathrm{Ad}^\dagger$

$$g \to g \cdot \xi = \mathrm{Ad}_g^\dagger(\xi)$$

*is called the **coadjoint action**. Then we have an induced map $\mathrm{ad}^\dagger : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}^*)$ at the Lie algebra level;*

$$\mathrm{ad}^\dagger(x)(\xi)(y) = -\xi([x, y]).$$

The orbits of the action given by Ad^* are called coadjoint orbits and we will show in theorem below that each orbit is a symplectic manifold in a natural way.

23.4 Complex Structure and Kähler Manifolds

Recall that a complex manifold is a manifold modelled on $\mathbb{C}^n$ and such that the chart overlap functions are all biholomorphic. Every (real) tangent space T_pM of a complex manifold M has a complex structure $J_p : T_pM \to T_pM$ given in biholomorphic coordinates $z = x + iy$ by

$$J_p(\left.\frac{\partial}{\partial x^i}\right|_p) = \left.\frac{\partial}{\partial y^i}\right|_p$$
$$J_p(\left.\frac{\partial}{\partial y^i}\right|_p) = -\left.\frac{\partial}{\partial x^i}\right|_p$$

and for any (biholomorphic) overlap function $\Delta = \varphi \circ \psi^{-1}$ we have $T\Delta \circ J = J \circ T\Delta$.

Definition 23.4 *An* ***almost complex structure*** *on a smooth manifold M is a bundle map $J : TM \to TM$ covering the identity map such that $J^2 = -id$. If one can choose an atlas for M such that all the coordinate change functions (overlap functions) Δ satisfy $T\Delta \circ J = J \circ T\Delta$ then J is called a* ***complex structure*** *on M.*

Definition 23.5 *An* ***almost symplectic structure*** *on a manifold M is a nondegenerate smooth 2-form ω which is not necessarily closed.*

Theorem 23.2 *A smooth manifold M admits an almost complex structure if and only if it admits an almost symplectic structure.*

Proof. First suppose that M has an almost complex structure J and let g be any Riemannian metric on M. Define a quadratic form q_p on each tangent space by

$$q_p(v) = \mathrm{g}_p(v, v) + \mathrm{g}_p(Jv, Jv).$$

Then we have $q_p(Jv) = q_p(v)$. Now let h be the metric obtained from the quadratic form q by polarization. It follows that $\mathrm{h}(v, w) = \mathrm{h}(Jv, Jw)$ for all $v, w \in TM$. Now define a two form ω by

$$\omega(v, w) = \mathrm{h}(v, Jw).$$

This really is skew-symmetric since $\omega(v, w) = \mathrm{h}(v, Jw) = \mathrm{h}(Jv, J^2w) = -\mathrm{h}(Jv, w) = \omega(w, v)$. Also, ω is nondegenerate since if $v \neq 0$ then $\omega(v, Jv) = \mathrm{h}(v, v) > 0$.

Conversely, let ω be a nondegenerate two form on a manifold M. Once again choose a Riemannian metric g for M. There must be a vector bundle map $\Omega : TM \to TM$ such that

$$\omega(v, w) = \mathrm{g}(\Omega v, w) \text{ for all } v, w \in TM.$$

Since ω is nondegenerate the map Ω must be invertible. Furthermore, since Ω is clearly anti-symmetric with respect to g the map $-\Omega \circ \Omega = -\Omega^2$ must be symmetric and positive definite. From linear algebra applied fiberwise we know that there must be a positive symmetric square root for $-\Omega^2$. Denote this by $P = \sqrt{-\Omega^2}$. Finite dimensional spectral theory also tell us that $P\Omega = \Omega P$. Now let $J = \Omega P^{-1}$ and notice that

$$J^2 = (\Omega P^{-1})(\Omega P^{-1}) = \Omega^2 P^{-2} = -\Omega^2 \Omega^{-2} = -\operatorname{id}.$$

■

One consequence of this result is that there must be characteristic class obstructions to the existence of a symplectic structure on a manifolds. In fact, if M, ω is a symplectic manifold then it is certainly almost symplectic and so there is an almost complex structure J on M. The tangent bundle is then a complex vector bundle with J giving the action of multiplication by $\sqrt{-1}$ on each fiber T_pM. Denote the resulting complex vector bundle by TM^J and then consider the total Chern class

$$c(TM^J) = c_n(TM^J) + ... + c_1(TM^J) + 1.$$

Here $c_i(TM^J) \in H^{2i}(M, \mathbb{Z})$. Recall that with the orientation given by ω^n the top class $c_n(TM^J)$ is the Euler class $e(TM)$ of TM. Now for the real bundle TM we have the total Pontrijagin class

$$p(TM) = p_n(TM) + ... + p_1(TM) + 1$$

which are related to the Chern classes by the Whitney sum

$$\begin{aligned} p(TM) &= c(TM^J) \oplus c(TM^{-J}) \\ &= (c_n(TM^J) + ... + c_1(TM^J) + 1)((-1)^n c_n(TM^J) - +... + c_1(TM^J) + 1) \end{aligned}$$

where TM^{-J} is the complex bundle with $-J$ giving the multiplication by $\sqrt{-1}$. We have used the fact that

$$c_i(TM^{-J}) = (-1)^i c_i(TM^J).$$

Now the classes $p_k(TM)$ are invariants of the diffeomorphism class of M an so can be considered constant over all possible choices of J. In fact, from the above relations one can deduce a quadratic relation that must be satisfied:

$$p_k(TM) = c_k(TM^J)^2 - 2c_{k-1}(TM^J)c_{k+1}(TM^J) + \cdots + (-1)^k 2c_{2k}(TM^J).$$

Now this places a restriction on what manifolds might have almost complex structures and hence a restriction on having an almost symplectic structure. Of course some manifolds might have an almost symplectic structure but still have no symplectic structure.

Definition 23.6 *A positive definite real bilinear form h on an almost complex manifold M, J is will be called Hermitian metric or J-metric if h is J invariant. In this case h is the real part of a Hermitian form on the complex vector bundle TM, J given by*

$$\langle v, w \rangle = h(v, w) + ih(Jv, w)$$

Definition 23.7 *A diffeomorphism $\phi : M, J, h \to M, J, h$ is called a Hermitian isometry iff $T\phi \circ J = J \circ T\phi$ and*

$$h(T\phi v, T\phi w) = h(v, w).$$

A group action $\rho : G \times M \to M$ is called a Hermitian action if $\rho(g, .)$ is a Hermitian isometry for all g. In this case, we have for every $p \in M$ a the representation $d\rho_p : H_p \to \mathrm{Aut}(T_pM, J_p)$ of the isotropy subgroup H_p given by

$$d\rho_p(g)v = T_p\rho_g \cdot v.$$

Definition 23.8 *Let M, J be a complex manifold and ω a symplectic structure on M. The manifold is called a **Kähler manifold** if $h(v, w) := \omega(v, Jw)$ is positive definite.*

*Equivalently we can define a **Kähler manifold as** a complex manifold M, J with Hermitian metric h with the property that the nondegenerate 2-form $\omega(v, w) := h(v, Jw)$ is closed.*

Thus we have the following for a Kähler manifold:

1. A complex structure J,
2. A J-invariant positive definite bilinear form b,
3. A Hermitian form $\langle v, w\rangle = h(v, w) + ih(Jv, w)$.
4. A symplectic form ω with the property that $\omega(v, w) = h(v, Jw)$.

Of course if M, J is a complex manifold with Hermitian metric h then $\omega(v, w) := h(v, Jw)$ automatically gives a nondegenerate 2-form; the question is whether it is closed or not. Mumford's criterion is useful for this purpose:

Theorem 23.3 (Mumford) *Let $\rho : G\times M \to M$ be a smooth Lie group action by Hermitian isometries. For $p \in M$ let H_p be the isometry subgroup of the point p. If $J_p \in d\rho_p(H_p)$ for every p then we have that ω defined by $\omega(v, w) := h(v, Jw)$ is closed.*

Proof. It is easy to see that since ρ preserves both h and J it also preserves ω and $d\omega$. Thus for any given $p \in M$, we have

$$d\omega(d\rho_p(g)u, d\rho_p(g)v, d\rho_p(g)w) = d\omega(u, v, w)$$

for all $g \in H_p$ and all $u, v, w \in T_pM$. By assumption there is a $g_p \in H_p$ with $J_p = d\rho_p(g_p)$. Thus with this choice the previous equation applied twice gives

$$\begin{aligned} d\omega(u, v, w) &= d\omega(J_pu, J_pv, J_pw) \\ &= d\omega(J_p^2u, J_p^2v, J_p^2w) \\ &= d\omega(-u, -v, -w) = -d\omega(u, v, w) \end{aligned}$$

so $d\omega = 0$ at p which was an arbitrary point so $d\omega = 0$. ∎

There is also Riemannian condition for the closedness of ω. Since a Kähler manifold is a posteriori a Riemannian manifold it has associated with it the Levi-Civita connection ∇. In the following we view J as an element of $\mathfrak{X}(M)$.

Theorem 23.4 *For a Kähler manifold M, J, h with associated symplectic form ω we have that*

$$d\omega = 0 \quad \textit{iff} \quad \nabla J = 0.$$

23.5 Symplectic musical isomorphisms

Since a symplectic form ω on a manifold M is nondegenerate we have a map

$$\omega_\flat : TM \to T^*M$$

given by $\omega_\flat(X_p)(v_p) = \omega(X_p, v_p)$ and the inverse $\omega^\sharp$ is such that

$$\iota_{\omega^\sharp(\alpha)}\omega = \alpha$$

or

$$\omega(\omega^\sharp(\alpha_p), v_p) = \alpha_p(v_p)$$

Let check that $\omega^\sharp$ really is the inverse (one could easily be off by a sign in this business):

$$\omega_\flat(\omega^\sharp(\alpha_p))(v_p) = \omega(\omega^\sharp(\alpha_p), v_p) = \alpha_p(v_p) \text{ for all } v_p$$
$$\Longrightarrow \omega_\flat(\omega^\sharp(\alpha_p)) = \alpha_p.$$

Notice that $\omega^\sharp$ induces a map on sections also denoted by $\omega^\sharp$ with inverse $\omega_\flat : \mathfrak{X}(M) \to \mathfrak{X}^*(M)$.

Notation 23.1 *Let us abbreviate $\omega^\sharp(\alpha)$ to $\sharp\alpha$ and $\omega_\flat(v)$ to $\flat v$.*

23.6 Darboux's Theorem

Lemma 23.2 (Darboux's theorem) *On a $2n$- manifold (M, ω) with a closed 2-form ω with $\omega^n \neq 0$ (for instance if (M, ω) is symplectic) there exists a sub-atlas consisting of charts called symplectic charts (a.k.a. canonical coordinates) characterized by the property that the expression for ω in such a chart is*

$$\omega_U = \sum_{i=1}^{n} dx^i \wedge dx^{i+n}$$

and so in particular M must have even dimension $2n$.

Remark 23.1 *Let us agree that the canonical coordinates can be written (x^i, y_i) instead of (x^i, x^{i+n}) when convenient.*

Remark 23.2 *It should be noticed that if x^i, y_i is a symplectic chart then $\sharp dx^i$ must be such that*

$$\sum_{r=1}^{n} dx^r \wedge dy^r(\sharp dx^i, \frac{\partial}{\partial x^j}) = \delta^i_j$$

but also

$$\sum_{r=1}^{n} dx^r \wedge dy^r(\sharp dx^i, \frac{\partial}{\partial x^j}) = \sum_{r=1}^{n} \left(dx^r(\sharp dx)dy^r(\frac{\partial}{\partial x^j}) - dy^r(\sharp dx^i)dx^r(\frac{\partial}{\partial x^j}) \right)$$
$$= -dy^j(\sharp dx^i)$$

and so we conclude that $\sharp dx^i = -\frac{\partial}{\partial y^i}$ *and similarly* $\sharp dy^i = \frac{\partial}{\partial x^i}$.

Proof. We will use induction and follow closely the presentation in [**?**]. Assume the theorem is true for symplectic manifolds of dimension $2(n-1)$. Let $p \in M$. Choose a function y^1 on some open neighborhood of p such that $dy_1(p) \neq 0$. Let $X = \sharp dy_1$ and then X will not vanish at p. We can then choose another function x^1 such that $Xx^1 = 1$ and we let $Y = -\sharp dx^1$. Now since $d\omega = 0$ we can use Cartan's formula to get

$$\mathcal{L}_X\omega = \mathcal{L}_Y\omega = 0.$$

Next contract ω with the bracket of X and Y (using the notation $\langle X, \omega \rangle = \iota_X\omega$, see notation 10.1):

$$\langle [X,Y], \omega \rangle = \langle \mathcal{L}_X Y, \omega \rangle = \mathcal{L}_X \langle Y, \omega \rangle - \langle Y, \mathcal{L}_X\omega \rangle$$
$$= \mathcal{L}_X(-dx^1) = -d(X(x^1)) = -d1 = 0.$$

Now since ω is nondegenerate this implies that $[X,Y] = 0$ and so there must be a local coordinate system $(x^1, y_1, w^1, ..., w^{2n-2})$ with

$$\frac{\partial}{\partial y_1} = Y$$
$$\frac{\partial}{\partial x^1} = X.$$

In particular, the theorem is true if $n = 1$. Assume the theorem is true for symplectic manifolds of dimension $2(n-1)$. If we let $\omega' = \omega - dx^1 \wedge dy_1$ then since $d\omega' = 0$ and hence

$$\langle X, \omega' \rangle = \mathcal{L}_X\omega' = \langle Y, \omega' \rangle = \mathcal{L}_Y\omega' = 0$$

we conclude that ω' can be expressed as a 2-form in the $w^1, ..., w^{2n-2}$ variables alone. Furthermore,

$$0 \neq \omega^n = (\omega - dx^1 \wedge dy_1)^n$$
$$= \pm n dx^1 \wedge dy_1 \wedge (\omega')^n$$

from which it follows that ω' is the pullback of a form nondegenerate form ϖ on $\mathbb{R}^{2n-2}$. To be exact if we let the coordinate chart given by $(x^1, y_1, w^1, ..., w^{2n-2})$ by denoted by ψ and let pr be the projection $\mathbb{R}^{2n} = \mathbb{R}^2 \times \mathbb{R}^{2n-1} \to \mathbb{R}^{2n-1}$ then $\omega' = (pr \circ \psi)^*\varpi$. Thus the induction hypothesis says that ω' has the form

$\omega' = \sum_{i=2}^{n} dx^i \wedge dy_i$ for some functions x^i, y_i with $i = 2, ..., n$. It is easy to see that the construction implies that in some neighborhood of p the full set of functions x^i, y_i with $i = 1, ..., n$ form the desired symplectic chart. ■

An atlas $\mathcal{A}$ of symplectic charts is called a symplectic atlas. A chart (U, φ) is called compatible with the symplectic atlas $\mathcal{A}$ if for every $(\psi_\alpha, U_\alpha) \in \mathcal{A}$ we have

$$(\varphi \circ \psi^{-1})^* \omega_0 = \omega_0$$

for the canonical symplectic $\omega_{can} = \sum_{i=1}^{n} du^i \wedge du^{i+n}$ defined on $\psi_\alpha(U \cap U_\alpha) \subset \mathbb{R}^{2n}$ using standard rectangular coordinates u^i.

23.7 Poisson Brackets and Hamiltonian vector fields

Definition 23.9 (on forms) *The* ***Poisson bracket*** *of two 1-forms is defined to be*

$$\{\alpha, \beta\}_\pm = \mp \flat[\sharp\alpha, \sharp\beta]$$

where the musical symbols refer to the maps $\omega^\sharp$ and $\omega_\flat$. This puts a Lie algebra structure on the space of 1-forms $\Omega^1(M) = \mathfrak{X}^*(M)$.

Definition 23.10 (on functions) *The* ***Poisson bracket*** *of two smooth functions is defined to be*

$$\{f, g\}_\pm = \pm\omega(\sharp df, \sharp dg) = \pm\omega(X_f, X_g)$$

This puts a Lie algebra structure on the space $\mathcal{F}(M)$ of smooth function on the symplectic M. It is easily seen (using $dg = \iota_{X_g}\omega$) that $\{f, g\}_\pm = \pm L_{X_g} f = \mp L_{X_f} g$ which shows that $f \mapsto \{f, g\}$ is a derivation for fixed g. The connection between the two Poisson brackets is

$$d\{f, g\}_\pm = \{df, dg\}_\pm.$$

Let us take canonical coordinates so that $\omega = \sum_{i=1}^{n} dx^i \wedge dy_i$. If $X_p = \sum_{i=1}^{n} dx^i(X)\frac{\partial}{\partial x^i} + \sum_{i=1}^{n} dy_i(X)\frac{\partial}{\partial y_i}$ and $v_p = dx^i(v_p)\frac{\partial}{\partial x^i} + dy_i(v_p)\frac{\partial}{\partial y_i}$ then using the Einstein summation convention we have

$$\begin{aligned}
&\omega_\flat(X)(v_p) \\
&= \omega(dx^i(X)\frac{\partial}{\partial x^i} + dy_i(X)\frac{\partial}{\partial y_i}, dx^i(v_p)\frac{\partial}{\partial x^i} + dy_i(v_p)\frac{\partial}{\partial y_i}) \\
&= (dx^i(X)dy_i - dy_i(X)dx^i)(v_p)
\end{aligned}$$

so we have

Lemma 23.3 $\omega_\flat(X_p) = \sum_{i=1}^{n} dx^i(X)dy_i - dy_i(X)dx^i = \sum_{i=1}^{n}(-dy_i(X)dx^i + dx^i(X)dy_i)$

Corollary 23.1 *If* $\alpha = \sum_{i=1}^n \alpha(\frac{\partial}{\partial x^i})dx^i + \sum_{i=1}^n \alpha(\frac{\partial}{\partial y_i})dy^i$ *then* $\omega^\sharp(\alpha) = \sum_{i=1}^n \alpha(\frac{\partial}{\partial y_i})\frac{\partial}{\partial x^i} - \sum_{i=1}^n \alpha(\frac{\partial}{\partial x^i})\frac{\partial}{\partial y_i}$

An now for the local formula:

Corollary 23.2 $\{f,g\} = \sum_{i=1}^n (\frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial x^i})$

Proof. $df = \frac{\partial f}{\partial x^i}dx^i + \frac{\partial f}{\partial y_i}dy_i$ and $dg = \frac{\partial g}{\partial x^j}dx^j + \frac{\partial g}{\partial y_i}dy_i$ so $\sharp df = \frac{\partial f}{\partial y_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y_i}$ and similarly for dg. Thus (using the summation convention again);

$$\begin{aligned}
\{f,g\} &= \omega(\sharp df, \sharp dg) \\
&= \omega(\frac{\partial f}{\partial y_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y_i}, \frac{\partial g}{\partial y_i}\frac{\partial}{\partial x^j} - \frac{\partial g}{\partial x^j}\frac{\partial}{\partial y_i}) \\
&= \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial x^i}
\end{aligned}$$

■

A main point about Poison Brackets is

Theorem 23.5 *f is constant along the orbits of X_g iff $\{f,g\} = 0$. In fact,*

$$\tfrac{d}{dt} g \circ \mathrm{Fl}_t^{X_f} = 0 \Longleftrightarrow \quad \{f,g\} = 0 \quad \Longleftrightarrow \tfrac{d}{dt} f \circ \mathrm{Fl}_t^{X_g} = 0$$

Proof. $\frac{d}{dt} g \circ \mathrm{Fl}_t^{X_f} = (\mathrm{Fl}_t^{X_f})^* L_{X_f} g = (\mathrm{Fl}_t^{X_f})^* \{f,g\}$. Also use $\{f,g\} = -\{g,f\}$. ■

The equations of motion for a Hamiltonian H are

$$\frac{d}{dt} f \circ \mathrm{Fl}_t^{X_H} = \pm\{f \circ \mathrm{Fl}_t^{X_H}, H\}_\pm = \mp\{H, f \circ \mathrm{Fl}_t^{X_H}\}_\pm$$

which is true by the following simple computation

$$\begin{aligned}
\frac{d}{dt} f \circ \mathrm{Fl}_t^{X_H} &= \frac{d}{dt}(\mathrm{Fl}_t^{X_H})^* f = (\mathrm{Fl}_t^{X_H})^* L_{X_H} f \\
&= L_{X_H}(f \circ \mathrm{Fl}_t^{X_H}) = \{f \circ \mathrm{Fl}_t^{X_H}, H\}_\pm.
\end{aligned}$$

Notation 23.2 *From now on we will use only $\{.,.\}_+$ unless otherwise indicated and shall write $\{.,.\}$ for $\{.,.\}_+$.*

Definition 23.11 *A Hamiltonian system is a triple (M, ω, H) where M is a smooth manifold, ω is a symplectic form and H is a smooth function $H : M \to \mathbb{R}$.*

The main example, at least from the point of view of mechanics, is the cotangent bundle of a manifold which is discussed below. From a mechanical point of view the Hamiltonian function controls the dynamics and so is special.

Let us return to the general case of a symplectic manifold M, ω

Definition 23.12 *Now if $H : M \to \mathbb{R}$ is smooth then we define the* ***Hamiltonian vector field*** *X_H with energy function H to be $\omega^\sharp dH$ so that by definition $\iota_{X_H}\omega = dH$.*

Definition 23.13 *A vector field X on M, ω is called a* ***locally Hamiltonian vector field*** *or a* ***symplectic vector field*** *iff $L_X\omega = 0$.*

If a symplectic vector field is complete then we have that $(\mathrm{Fl}_t^X)^*\omega$ is defined for all $t \in \mathbb{R}$. Otherwise, for any relatively compact open set U the restriction Fl_t^X to U is well defined for all $t \le b(U)$ for some number depending only on U. Thus $(\mathrm{Fl}_t^X)^*\omega$ is defined on U for $t \le b(U)$. Since U can be chosen to contain any point of interest and since M can be covered by relatively compact sets, it will be of little harm to write $(\mathrm{Fl}_t^X)^*\omega$ even in the case that X is not complete.

Lemma 23.4 *The following are equivalent:*

1. *X is* ***symplectic vector field****, i.e. $L_X\omega = 0$*
2. *$\iota_X\omega$ is closed*
3. *$(\mathrm{Fl}_t^X)^*\omega = \omega$*
4. *X is locally a Hamiltonian vector field.*

Proof. (1)$\Longleftrightarrow$ (4) by the Poincaré lemma. Next, notice that $L_X\omega = d \circ \iota_X\omega + \iota_X \circ d\omega = d \circ \iota_X\omega$ so we have (2)$\Longleftrightarrow$(1). The implication (2)$\Longleftrightarrow$(3) follows from Theorem 7.8. ∎

Proposition 23.2 *We have the following easily deduced facts concerning Hamiltonian vector fields:*

1. The H is constant along integral curves of X_H
2. The flow of X_H is a local symplectomorphism. That is $\mathrm{Fl}_t^{X_H\,*}\omega = \omega$

Notation 23.3 *Denote the set of all Hamiltonian vector fields on M, ω by $\mathcal{H}(\omega)$ and the set of all symplectic vector fields by $\mathcal{SP}(\omega)$*

Proposition 23.3 *The set $\mathcal{SP}(\omega)$ is a Lie subalgebra of $\mathfrak{X}(M)$. In fact, we have $[\mathcal{SP}(\omega), \mathcal{SP}(\omega)] \subset \mathcal{H}(\omega) \subset \mathfrak{X}(M)$.*

Proof. Let $X, Y \in \mathcal{SP}(\omega)$. Then

$$
\begin{aligned}
[X,Y]\lrcorner\omega &= \mathcal{L}_X Y\lrcorner\omega = \mathcal{L}_X(Y\lrcorner\omega) - Y\lrcorner\mathcal{L}_X\omega \\
&= d(X\lrcorner Y\lrcorner\omega) + X\lrcorner d(Y\lrcorner\omega) - 0 \\
&= d(X\lrcorner Y\lrcorner\omega) + 0 + 0 \\
&= -d(\omega(X,Y)) = -X_{\omega(X,Y)}\lrcorner\omega
\end{aligned}
$$

and since ω in nondegenerate we have $[X,Y] = X_{-\omega(X,Y)} \in \mathcal{H}(\omega)$. ∎

23.8 Configuration space and Phase space

Consider the cotangent bundle of a manifold Q with projection map

$$\pi : T^*Q \to Q$$

and define the **canonical 1-form** $\theta \in T^*(T^*Q)$ by

$$\theta : v_{\alpha_p} \mapsto \alpha_p(T\pi \cdot v_{\alpha_p})$$

where $\alpha_p \in T_p^*Q$ and $v_{\alpha_p} \in T_{\alpha_p}(T_p^*Q)$. In local coordinates this reads

$$\theta_0 = \sum p_i dq^i.$$

Then $\omega_{T^*Q} = -d\theta$ is a symplectic form which in natural coordinates reads

$$\omega_{T^*Q} = \sum dq^i \wedge dp_i$$

Lemma 23.5 *θ is the unique 1-form such that for any $\beta \in \Omega^1(Q)$ we have*

$$\beta^*\theta = \beta$$

where we view β as $\beta : Q \to T^*Q$.

Proof: $\beta^*\theta(v_q) = \theta|_{\beta(q)}(T\beta \cdot v_q) = \beta(q)(T\pi \circ T\beta \cdot v_q) = \beta(q)(v_q)$ since $T\pi \circ T\beta = T(\pi \circ \beta) = T(\mathrm{id}) = \mathrm{id}$.

The cotangent lift T^*f of a diffeomorphism $f : Q_1 \to Q_2$ is defined by the commutative diagram

$$\begin{array}{ccc} T^*Q_1 & \xleftarrow{T^*f} & T^*Q_2 \\ \downarrow & & \downarrow \\ Q_1 & \xrightarrow{f} & Q_2 \end{array}$$

and is a symplectic map; i.e. $(T^*f)^*\omega_0 = \omega_0$. In fact, we even have $(T^*f)^*\theta_0 = \theta_0$.

The triple (T^*Q, ω_{T^*Q}, H) is a Hamiltonian system for any choice of smooth function. The most common form for H in this case is $\frac{1}{2}K + V$ where K is a Riemannian metric which is constructed using the mass distribution of the bodies modelled by the system and V is a smooth potential function which, in a conservative system, depends only on $\mathbf{q}$ when viewed in natural cotangent bundle coordinates q^i, p_i.

Now we have $\sharp dg = \frac{\partial g}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial}{\partial p_i}$ and introducing the $\pm$ notation one more time we have

$$\begin{aligned}
\{f, g\}_\pm &= \pm\omega_{T^*Q}(\sharp df, \sharp dg) = \pm df(\sharp dg) = \pm df(\frac{\partial g}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial}{\partial p_i}) \\
&= \pm(\frac{\partial g}{\partial p_i}\frac{\partial f}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial f}{\partial p_i}) \\
&= \pm(\frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i})
\end{aligned}$$

Thus letting

$$\mathrm{Fl}_t^{X_H}(q_0^1, ..., q_0^n, p_0^1, ..., p_0^n) = (q^1(t), ..., q^n(t), p^1(t), ..., p^n(t))$$

the equations of motions read

$$\begin{aligned}\frac{d}{dt} f(q(t), p(t)) &= \frac{d}{dt} f \circ \mathrm{Fl}_t^{X_H} = \{f \circ \mathrm{Fl}_t^{X_H}, H\} \\ &= \frac{\partial f}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q^i}.\end{aligned}$$

Where we have abbreviated $f \circ \mathrm{Fl}_t^{X_H}$ to just f. In particular, if $f = q^i$ and $f = p_i$ then

$$\begin{aligned}\dot{q}^i(t) &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i(t) &= -\frac{\partial H}{\partial q^i}\end{aligned}$$

which should be familiar.

23.9 Transfer of symplectic structure to the Tangent bundle

Case I: a (pseudo) Riemannian manifold

If Q, g is a (pseudo) Riemannian manifold then we have a map $\mathrm{g}^\flat : TQ \to T^*Q$ defined by

$$\mathrm{g}^\flat(v)(w) = \mathrm{g}(v, w)$$

and using this we can define a symplectic form ϖ_0 on TQ by

$$\varpi_0 = \left(\mathrm{g}^\flat\right)^* \omega$$

(Note that $d\varpi_0 = d(\mathrm{g}^{\flat *}\omega) = \mathrm{g}^{\flat *} d\omega = 0$.) In fact, ϖ_0 is exact since ω is exact:

$$\begin{aligned}\varpi_0 &= \left(\mathrm{g}^\flat\right)^* \omega \\ &= \left(\mathrm{g}^\flat\right)^* d\theta = d\left(\mathrm{g}^{\flat *}\theta\right).\end{aligned}$$

Let us write $\Theta_0 = \mathrm{g}^{\flat *}\theta$. Locally we have

$$\begin{aligned}\Theta_0(x, v)(v_1, v_2) &= \mathrm{g}_x(v, v_1) \text{ or} \\ \Theta_0 &= \sum \mathrm{g}_{ij}\dot{q}^i dq^j\end{aligned}$$

and also

$$\begin{aligned}&\varpi_0(x,v)((v_1,v_2),((w_1,w_2)))\\&=\mathrm{g}_x(w_2,v_1)-\mathrm{g}_x(v_2,w_1)+D_x\mathrm{g}_x(v,v_1)\cdot w_1-D_x\mathrm{g}_x(v,w_1)\cdot v_1\end{aligned}$$

which in classical notation (and for finite dimensions) looks like

$$\varpi_h=\mathrm{g}_{ij}dq^i\wedge d\dot{q}^j+\sum\frac{\partial \mathrm{g}_{ij}}{\partial q^k}\dot{q}^i dq^j\wedge dq^k$$

Case II: Transfer of symplectic structure by a Lagrangian function.

Definition 23.14 *Let $L : TQ \to Q$ be a Lagrangian on a manifold Q. We say that L is* ***regular*** *or* ***non-degenerate*** *at $\xi \in TQ$ if in any canonical coordinate system $(q, \dot{q})$ whose domain contains ξ, the matrix*

$$\left[\frac{\partial^2 L}{\partial \dot{q}^i\partial \dot{q}^j}(q(\xi),\dot{q}(\xi))\right]$$

is non-degenerate. L is called ***regular*** *or* ***nondegenerate*** *if it is regular at all points in TQ.*

We will need the following general concept:

Definition 23.15 *Let $\pi_E : E \to M$ and $\pi_F : F \to M$ be two vector bundles. A map $L : E \to F$ is called a* ***fiber preserving map*** *if the following diagram commutes*

$$\begin{array}{ccccc} E & & \stackrel{L}{\to} & & F \\ & & & & \\ \pi_E & \searrow & & \swarrow & \pi_F \\ & & M & & \end{array}.$$

We do not require that the map L be linear on the fibers and so in general L is not a vector bundle morphism.

Definition 23.16 *If $L : E \to F$ is a fiber preserving map then if we denote the restriction of L to a fiber E_p by L_p define the* ***fiber derivative***

$$\mathbf{F}L : E \to \mathrm{Hom}(E, F)$$

by $\mathbf{F}L : e_p \mapsto Df|_p(e_p)$ for $e_p \in E_p$.

In our application of this concept, we take F to be the trivial bundle $Q\times\mathbb{R}$ over Q so $\mathrm{Hom}(E, F) = \mathrm{Hom}(E, \mathbb{R}) = T^*Q$.

Lemma 23.6 *A Lagrangian function $L : TQ \to \mathbb{R}$ gives rise to a fiber derivative $\mathbf{F}L : TQ \to T^*Q$. The Lagrangian is nondegenerate iff $\mathbf{F}L$ is a diffeomorphism.*

Definition 23.17 *The form ϖ_L is defined by*

$$\varpi_L = (\mathbf{F}L)^* \omega$$

Lemma 23.7 *ω_L is a symplectic form on TQ iff L is nondegenerate (i.e. if $\mathbf{F}L$ is a diffeomorphism).*

Observe that we can also define $\theta_L = (\mathbf{F}L)^* \theta$ so that $d\theta_L = d(\mathbf{F}L)^* \theta = (\mathbf{F}L)^* d\theta = (\mathbf{F}L)^* \omega = \varpi_L$ so we see that ω_L is exact (and hence closed a required for a symplectic form).

Now in natural coordinates we have

$$\varpi_L = \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} dq^i \wedge dq^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} dq^i \wedge d\dot{q}^j$$

as can be verified using direct calculation.

The following connection between the transferred forms ϖ_L and ϖ_0 and occasionally not pointed out in some texts.

Theorem 23.6 *Let V be a smooth function on a Riemannian manifold M, h. If we define a Lagrangian by $L = \frac{1}{2}h - V$ then the Legendre transformation $\mathbf{F}L :: TQ \to T^*Q$ is just the map $g^\flat$ and hence $\varpi_L = \varpi_h$.*

Proof. We work locally. Then the Legendre transformation is given by

$$\begin{array}{c} q^i \mapsto q^i \\ \dot{q}^i \mapsto \frac{\partial L}{\partial \dot{q}^i} \end{array} .$$

But since $L(\dot{\mathbf{q}}, \dot{\mathbf{q}}) = \frac{1}{2}\mathrm{g}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) - V(q)$ we have $\frac{\partial L}{\partial \dot{q}^i} = \frac{\partial}{\partial \dot{q}^i} \frac{1}{2} \mathrm{g}_{kl} \dot{q}^l \dot{q}^k = \mathrm{g}_{il} \dot{q}^l$ which together with $q^i \mapsto q^i$ is the coordinate expression for $\mathrm{g}^\flat$:

$$\begin{array}{c} q^i \mapsto q^i \\ \dot{q}^i \mapsto \mathrm{g}_{il} \dot{q}^l \end{array}$$

■

23.10 Coadjoint Orbits

Let G be a Lie group and consider $\mathrm{Ad}^\dagger : G \to \mathrm{GL}(\mathfrak{g}^*)$ and the corresponding coadjoint action as in example 23.5. For every $\xi \in \mathfrak{g}^*$ we have a Left invariant 1-form on G defined by

$$\theta^\xi = \xi \circ \omega_G$$

where ω_G is the canonical $\mathfrak{g}$-valued 1-form (the Maurer Cartan form). Let the G_ξ be the isotropy subgroup of G for a point $\xi \in \mathfrak{g}^*$ under the coadjoint action. Then it is standard that orbit $G \cdot \xi$ is canonically diffeomorphic to the orbit space G/G_ξ and the map $\phi_\xi : g \mapsto g \cdot \xi$ is a submersion onto . Then we have

Theorem 23.7 *There is a unique symplectic form Ω^{ξ} on $G/G_{\xi} \cong G \cdot \xi$ such that $\phi_{\xi}^{*}\Omega^{\xi} = d\theta^{\xi}$.*

Proof: If such a form as Ω^{ξ} exists as stated then we must have

$$\Omega^{\xi}(T\phi_{\xi}.v, T\phi_{\xi}.w) = d\theta^{\xi}(v, w) \text{ for all } v, w \in T_g G$$

We will show that this in fact *defines* Ω^{ξ} as a symplectic form on the orbit $G \cdot \xi$. First of all notice that by the structure equations for the Maurer Cartan form we have for $v, w \in T_e G = \mathfrak{g}$

$$\begin{aligned} d\theta^{\xi}(v, w) &= \xi(d\omega_G(v, w)) = \xi(\omega_G([v, w])) \\ &= \xi(-[v, w]) = \mathrm{ad}^{\dagger}(v)(\xi)(w) \end{aligned}$$

From this we see that

$$\mathrm{ad}^{\dagger}(v)(\xi) = 0 \iff v \in \mathrm{Null}(\left. d\theta^{\xi}\right|_e)$$

where $\mathrm{Null}(\left. d\theta^{\xi}\right|_e) = \{v \in \mathfrak{g} : \left. d\theta^{\xi}\right|_e (v, w)$ for all $w \in \mathfrak{g}\}$. On the other hand, $G_{\xi} = \ker\{g \longmapsto \mathrm{Ad}_g^{\dagger}(\xi)\}$ so $\mathrm{ad}^{\dagger}(v)(\xi) = 0$ iff $v \in T_e G_{\xi} = \mathfrak{g}_{\xi}$.

Now notice that since $d\theta^{\xi}$ is left invariant we have that $\mathrm{Null}(\left. d\theta^{\xi}\right|_g) = TL_g(\mathfrak{g}_{\xi})$ which is the tangent space to the coset gG_{ξ} which is also $\ker \left. T\phi_{\xi}\right|_g$. Thus we conclude that

$$\mathrm{Null}(\left. d\theta^{\xi}\right|_g) = \ker \left. T\phi_{\xi}\right|_g .$$

It follows that we have a natural isomorphism

$$T_{g \cdot \xi}(G \cdot \xi) = \left. T\phi_{\xi}\right|_g (T_g G) \approx T_g G / (TL_g(\mathfrak{g}_{\xi}))$$

Another view: Let the vector field on $G \cdot \xi$ corresponding to $v, w \in \mathfrak{g}$ generated by the action be denoted by $v^{\dagger}$and $w^{\dagger}$. Then we have $\Omega^{\xi}(\xi)(v^{\dagger}, w^{\dagger}) := \xi(-[v, w])$ at $\xi \in G \cdot \xi$ and then extend to the rest of the points of the orbit by equivariance:

$$\Omega^{\xi}(g \cdot \xi)(v^{\dagger}, w^{\dagger}) = \overset{\dagger}{\underset{g}{\mathrm{Ad}}}(\xi(-[v, w]))$$

23.11 The Rigid Body

In what follows we will describe the rigid body rotating about one of its points in three different versions. The basic idea is that we can represent the configuration space as a subset of $\mathbb{R}^{3N}$ with a very natural kinetic energy function. But this space is also isomorphic to the rotation group $SO(3)$ and we can transfer the kinetic energy metric over to $SO(3)$ and then the evolution of the system is given by geodesics in $SO(3)$ with respect to this metric. Next we take advantage of the fact that the tangent bundle of $SO(3)$ is trivial to transfer the setup over to a trivial bundle. But there are two natural ways to do this and we explore the relation between the two.

23.11.1 The configuration in $\mathbb{R}^{3N}$

Let us consider a rigid body to consist of a set of point masses located in $\mathbb{R}^3$ at points with position vectors $\mathbf{r}_1(t), ...\mathbf{r}_N(t)$ at time t. Thus $\mathbf{r}_i = (x_1, x_2, x_3)$ is the coordinates of the i-th point mass. Let $m_1, ..., m_N$ denote the masses of the particles. To say that this set of point masses is rigid is to say that the distances $|\mathbf{r}_i - \mathbf{r}_j|$ are constant for each choice of i and j. Let us assume for simplicity that the body is in a state of uniform rectilinear motion so that by re-choosing our coordinate axes if necessary we can assume that the there is one of the point masses at the origin of our coordinate system at all times. Now the set of all possible configurations is some submanifold of $\mathbb{R}^{3N}$ which we denote by M. Let us also assume that at least 3 of the masses, say those located at $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2$ are situated so that the position vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2$ form a basis of $\mathbb{R}^3$. For convenience let $\mathbf{r}$ and $\dot{\mathbf{r}}$ be abbreviations for $(\mathbf{r}_1(t), ..., \mathbf{r}_N(t))$ and $(\dot{\mathbf{r}}_1(t), ..., \dot{\mathbf{r}}_N(t))$.The correct kinetic energy for the system of particles forming the rigid body is $\frac{1}{2}K(\dot{\mathbf{r}}, \dot{\mathbf{r}})$ where the kinetic energy metric K is

$$K(\mathbf{v}, \mathbf{w}) = m_1 \mathbf{v}_1 \cdot \mathbf{w}_1 + \cdots + m_N \mathbf{v}_N \cdot \mathbf{w}_N.$$

Since there are no other forces on the body other than those that constrain the body to be rigid the Lagrangian for M is just $\frac{1}{2}K(\dot{\mathbf{r}}, \dot{\mathbf{r}})$ and the evolution of the point in M representing the body is a geodesic when we use as Hamiltonian K and the symplectic form pulled over from T^*M as described previously.

23.11.2 Modelling the rigid body on $SO(3)$

Let $\mathbf{r}_1(0), ...\mathbf{r}(0)_N$ denote the initial positions of our point masses. Under these condition there is a unique matrix valued function $g(t)$ with values in $SO(3)$ such that $\mathbf{r}_i(t) = g(t)\mathbf{r}_i(0)$. Thus the motion of the body is determined by the curve in $SO(3)$ given by $t \mapsto g(t)$. In fact, we can map $SO(3)$ to the set of all possible configurations of the points making up the body in a 1-1 manner by letting $\mathbf{r}_1(0) = \xi_1, ...\mathbf{r}(0)_N = \xi_N$ and mapping $\Phi : g \mapsto (g\xi_1, ..., g\xi_N) \in M \subset \mathbb{R}^{3N}$. If we use the map Φ to transfer this over to $TSO(3)$ we get

$$k(\xi, \upsilon) = K(T\Phi \cdot \xi, T\Phi \cdot \upsilon)$$

for $\xi, \upsilon \in TSO(3)$. Now k is a Riemannian metric on $SO(3)$ and in fact, k is a left invariant metric:

$$k(\xi, \upsilon) = k(TL_g\xi, TL_g\upsilon) \text{ for all } \xi, \upsilon \in TSO(3).$$

Exercise 23.2 *Show that* k *really is left invariant. Hint: Consider the map* $\mu_{g_0} : (\mathbf{v_1}, \cdots, \mathbf{v_N}) \mapsto (g_0\mathbf{v_1}, \cdots, g_0\mathbf{v_N})$ *for* $g_0 \in SO(3)$ *and notice that* $\mu_{g_0} \circ \Phi = \Phi \circ L_{g_0}$ *and hence* $T\mu_{g_0} \circ T\Phi = T\Phi \circ TL_{g_0}$.

Now by construction, the Riemannian manifolds M, K and $SO(3)$, k are isometric. Thus the corresponding path $g(t)$ in $SO(3)$ is a geodesic with respect to the left invariant metric k. Our Hamiltonian system is now $(TSO(3), \Omega_k, k)$ where Ω_k is the Legendre transformation of the canonical symplectic form Ω on $T^*SO(3)$

23.11.3 The trivial bundle picture

Recall that we the Lie algebra of $SO(3)$ is the vector space of skew-symmetric matrices $\mathfrak{so}(3)$. We have the two trivializations of the tangent bundle $TSO(3)$ given by

$$\begin{aligned}\mathrm{triv}_L(v_g) &= (g, \omega_G(v_g)) = (g, g^{-1}v_g) \\ \mathrm{triv}_R(v_g) &= (g, \omega_G(v_g)) = (g, v_g g^{-1})\end{aligned}$$

with inverse maps $SO(3) \times \mathfrak{so}(3) \to TSO(3)$ given by

$$\begin{aligned}(g, B) &\mapsto TL_g B \\ (g, B) &\mapsto TR_g B\end{aligned}$$

Now we should be able to represent the system in the trivial bundle $SO(3) \times \mathfrak{so}(3)$ via the map $\mathrm{triv}_L(v_g) = (g, \omega_G(v_g)) = (g, g^{-1}v_g)$. Thus we let k_0 be the metric on $SO(3) \times \mathfrak{so}(3)$ coming from the metric k. Thus by definition

$$\mathrm{k}_0((g, v), (g, w)) = \mathrm{k}(TL_g v, TL_g w) = \mathrm{k}_e(v, w)$$

where $v, w \in \mathfrak{so}(3)$ are skew-symmetric matrices.

23.12 The momentum map and Hamiltonian actions

Remark 23.3 *In this section all Lie groups will be assumed to be connected.*

Suppose that (a connected Lie group) G acts on M, ω as a group of symplectomorphisms.

$$\sigma : G \times M \to M$$

Then we say that σ is a **symplectic G-action** . Since G acts on M we have for every $v \in \mathfrak{g}$ the fundamental vector field $X^v = v^\sigma$. The fundamental vector field will be symplectic (locally Hamiltonian). Thus every one-parameter group g^t of G induces a symplectic vector field on M. Actually, it is only the infinitesimal action that matters at first so we define

Definition 23.18 *Let M be a smooth manifold and let $\mathfrak{g}$ be the Lie algebra of a connected Lie group G. A linear map $\sigma' : v \mapsto X^v$ from $\mathfrak{g}$ into $\mathfrak{X}(M)$ is called a* $\mathfrak{g}$-***action*** *if*

$$\begin{aligned}[X^v, X^w] &= -X^{[v,w]} \text{ or} \\ [\sigma'(v), \sigma'(w)] &= -\sigma'([v, w]).\end{aligned}$$

*If M, ω is symplectic and the $\mathfrak{g}$-action is such that $\mathcal{L}_{X^v}\omega = 0$ for all $v \in \mathfrak{g}$ we say that the action is a **symplectic** g-**action**.*

Definition 23.19 *Every symplectic action $\sigma : G \times M \to M$ induces a $\mathfrak{g}$-action $d\sigma$ via*

$$d\sigma : \upsilon \mapsto X^{\upsilon}$$

$$\text{where } X^{\upsilon}(x) = \left.\frac{d}{dt}\right|_0 \sigma(\exp(t\upsilon), x).$$

In some cases, we may be able to show that for all υ the symplectic field X^{υ} is a full fledged Hamiltonian vector field. In this case associated to each $\upsilon \in \mathfrak{g}$ there is a Hamiltonian function $J_{\upsilon} = J_{X^{\upsilon}}$ with corresponding Hamiltonian vector field equal to X^{υ} and J_{υ} is determined up to a constant by $X^{\upsilon} = \sharp dJ_{X^{\upsilon}}$. Now $\iota_{X^{\upsilon}}\omega$ is always closed since $d\iota_{X^{\upsilon}}\omega = \mathcal{L}_{X^{\upsilon}}\omega$. When is it possible to define J_{υ} for every $\upsilon \in \mathfrak{g}$?

Lemma 23.8 *Given a symplectic $\mathfrak{g}$-action $\sigma' : \upsilon \mapsto X^{\upsilon}$ as above, there is a linear map $\upsilon \mapsto J_{\upsilon}$ such that $X^{\upsilon} = \sharp dJ_{\upsilon}$ for every $\upsilon \in \mathfrak{g}$ iff $\iota_{X^{\upsilon}}\omega$ is exact for all $\upsilon \in \mathfrak{g}$.*

Proof. If $H_{\upsilon} = H_{X^{\upsilon}}$ exists for all υ then $dJ_{X^{\upsilon}} = \omega(X^{\upsilon}, .) = \iota_{X^{\upsilon}}\omega$ for all υ so $\iota_{X^{\upsilon}}\omega$ is exact for all $\upsilon \in \mathfrak{g}$. Conversely, if for every $\upsilon \in \mathfrak{g}$ there is a smooth function h_{υ} with $dh_{\upsilon} = \iota_{X^{\upsilon}}\omega$ then $X^{\upsilon} = \sharp dh_{\upsilon}$ so h_{υ} is Hamiltonian for X^{υ}. Now let $\upsilon_1, ..., \upsilon_n$ be a basis for $\mathfrak{g}$ and define $J_{\upsilon_i} = h_{\upsilon_i}$ and extend linearly. ∎

Notice that the property that $\upsilon \mapsto J_{\upsilon}$ is linear means that we can define a map $J : M \to \mathfrak{g}^*$ by

$$J(x)(\upsilon) = J_{\upsilon}(x)$$

and this is called a **momentum map** .

Definition 23.20 *A symplectic G-action σ (resp. $\mathfrak{g}$-action σ') on M such that for every $\upsilon \in \mathfrak{g}$ the vector field X^{υ} is a Hamiltonian vector field on M is called a **Hamiltonian** G-**action** (resp. Hamiltonian $\mathfrak{g}$-action).*

We can thus associate to every Hamiltonian action at least one momentum map-this being unique up to an additive constant.

Example 23.6 *If G acts on a manifold Q by diffeomorphisms then G lifts to an action on the cotangent bundle T^*M which is automatically symplectic. In fact, because $\omega_0 = d\theta_0$ is exact the action is also a Hamiltonian action. The Hamiltonian function associated to an element $\upsilon \in \mathfrak{g}$ is given by*

$$J_{\upsilon}(x) = \theta_0\left(\left.\frac{d}{dt}\right|_0 \exp(t\upsilon) \cdot x \right).$$

Definition 23.21 *If G (resp. $\mathfrak{g}$) acts on M in a symplectic manner as above such that the action is Hamiltonian and such that we may choose a momentum map J such that*

$$J_{[\upsilon,w]} = \{J_{\upsilon}, J_w\}$$

*where $J_{\upsilon}(x) = J(x)(v)$ then we say that the action is a **strongly Hamiltonian** G-action (resp. $\mathfrak{g}$-action).*

Example 23.7 *The action of example 23.6 is strongly Hamiltonian.*

We would like to have a way to measure of whether a Hamiltonian action is strong or not. Essentially we are just going to be using the difference $J_{[v,w]} - \{J_v, J_w\}$ but it will be convenient to introduce another view which we postpone until the next section where we study "Poisson structures".

PUT IN THEOREM ABOUT MOMENTUM CONSERVATION!!!!

What is a momentum map in the cotangent case? Pick a fixed point $\alpha \in T^*Q$ and consider the map $\Phi_\alpha : G \to T^*Q$ given by $\Phi_\alpha(g) = g \cdot \alpha = g^{-1*}\alpha$. Now consider the pullback of the canonical 1-form $\Phi_\alpha^*\theta_0$.

Lemma 23.9 *The restriction* $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}$ *is an element of* $\mathfrak{g}^*$ *and the map* $\alpha \mapsto \Phi_\alpha^*\theta_0|_{\mathfrak{g}}$ *is the momentum map.*

Proof. We must show that $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v) = H_v(\alpha)$ for all $v \in \mathfrak{g}$. Does $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v)$ live in the right place? Let $g_v^t = \exp(vt)$. Then

$$
\begin{aligned}
(T_e\Phi_\alpha v) &= \left.\frac{d}{dt}\right|_0 \Phi_\alpha(\exp(vt)) \\
&= \left.\frac{d}{dt}\right|_0 (\exp(-vt))^*\alpha \\
&\quad \left.\frac{d}{dt}\right|_0 \exp(vt)\cdot\alpha
\end{aligned}
$$

We have

$$
\begin{aligned}
\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v) &= \theta_0|_{\mathfrak{g}}(T_e\Phi_\alpha v) \\
= \theta_0\Big(\left.\frac{d}{dt}\right|_0 \exp(vt)\cdot\alpha\Big) &= J_v(\alpha)
\end{aligned}
$$

■

Definition 23.22 *Let G act on a symplectic manifold M, ω and suppose that the action is Hamiltonian. A momentum map J for the action is said to be equivariant with respect to the coadjoint action if* $J(g\cdot x) = \mathrm{Ad}^*_{g^{-1}} J(x)$.

Chapter 24

Poisson Geometry

Life is good for only two things, discovering mathematics and teaching mathematics

–Siméon Poisson

24.1 Poisson Manifolds

In this chapter we generalize our study of symplectic geometry by approaching things from the side of a Poisson bracket.

Definition 24.1 *A* ***Poisson structure*** *on an associative algebra $\mathcal{A}$ is a Lie algebra structure with bracket denoted by $\{.,.\}$ such for a fixed $a \in \mathcal{A}$ that the map $x \mapsto \{a, x\}$ is a derivation of the algebra. An associative algebra with a Poisson structure is called a* ***Poisson algebra*** *and the bracket is called a* ***Poisson bracket*** *.*

We have already seen an example of a Poisson structure on the algebra $\mathfrak{F}(M)$ of smooth functions on a symplectic manifold. Namely,

$$\{f, g\} = \omega(\omega^\sharp df, \omega^\sharp dg).$$

By the Darboux theorem we know that we can choose local coordinates $(q^1, ..., q^n, p_1, ..., p_n)$ on a neighborhood of any given point in the manifold. Recall also that in such coordinates we have

$$\omega^\sharp df = \sum_{i=1}^{n} \frac{\partial f}{\partial p_i}\frac{\partial}{\partial q^i} - \sum_{i=1}^{n} \frac{\partial f}{\partial q^i}\frac{\partial}{\partial p_i}$$

sometimes called the symplectic gradient. It follows that

$$\sum_{i=1}^{n}\left(\frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}\right)$$

Definition 24.2 *A smooth manifold with a Poisson structure on is algebra of smooth functions is called a* ***Poisson manifold****.*

So every symplectic n-manifold gives rise to a Poisson structure. On the other hand, there are Poisson manifolds that are not so by virtue of being a symplectic manifold.

Now if our manifold is finite dimensional then every derivation of $\mathfrak{F}(M)$ is given by a vector field and since $g \mapsto \{f, g\}$ is a derivation there is a corresponding vector field X_f. Since the bracket is determined by these vector field and since vector fields can be defined locally (recall the presheaf $\mathfrak{X}_M$) we see that a Poisson structure is also a locally defined structure. In fact, $U \mapsto \mathfrak{F}_M(U)$ is a presheaf of Poisson algebras.

Now if we consider the map $w : \mathfrak{F}_M \to \mathfrak{X}_M$ defined by $\{f, g\} = w(f) \cdot g$ we see that $\{f, g\} = w(f) \cdot g = -w(g) \cdot f$ and so $\{f, g\}(p)$ depends only on the differentials df, dg of f and g. Thus we have a tensor $B(.,.) \in \Gamma \bigwedge^2 TM$ such that $B(df, dg) = \{f, g\}$. In other words, $B_p(.,.)$ is a symmetric bilinear map $T_p^*M \times T_p^*M \to \mathbb{R}$. Now any such tensor gives a bundle map $B^\sharp : T^*M \mapsto T^{**}M = TM$ by the rule $B^\sharp(\alpha)(\beta) = B(\beta, \alpha)$ for $\beta, \alpha \in T_p^*M$ and any $p \in M$. In other words, $B(\beta, \alpha) = \beta(B^\sharp(\alpha))$ for all $\beta \in T_p^*M$ and arbitrary $p \in M$. The 2-vector B is called the Poisson tensor for the given Poisson structure. B is also sometimes called a **co-symplectic structure** for reasons that we will now explain.

If M, ω is a symplectic manifold then the map $\omega_\flat : TM \to T^*M$ can be inverted to give a map $\omega^\sharp : T^*M \to TM$ and then a form $W \in \bigwedge^2 TM$ defined by $\omega^\sharp(\alpha)(\beta) = W(\beta, \alpha)$ (here again β, α must be in the same fiber). Now this form can be used to define a Poisson bracket by setting $\{f, g\} = W(df, dg)$ and so W is the corresponding Poisson tensor. But notice that

$$\begin{aligned}\{f, g\} &= W(df, dg) = \omega^\sharp(dg)(df) = df(\omega^\sharp(dg)) \\ &= \omega(\omega^\sharp df, \omega^\sharp dg)\end{aligned}$$

which is just the original Poisson bracket defined in the symplectic manifold M, ω.

Given a Poisson manifold $M, \{.,.\}$ we can always define $\{.,.\}_-$ by $\{f, g\}_- = \{g, f\}$. Since we some times refer to a Poisson manifold $M, \{.,.\}$ by referring just to the space we will denote M with the opposite Poisson structure by M^-.

A Poisson map is map $\phi : M, \{.,.\}_1 \to N, \{.,.\}_2$ is a smooth map such that $\phi^*\{f, g\} = \{\phi^* f, \phi^* g\}$ for all $f, g \in \mathfrak{F}(M)$.

For any subset S of a Poisson manifold let S_0 be the set of functions from $\mathfrak{F}(M)$ which vanish on S. A submanifold S of a Poisson manifold $M, \{.,.\}$ is called **coisotropic** if S_0 closed under the Poisson bracket. A Poisson manifold is called symplectic if the Poisson tensor B is non-degenerate since in this case we can use $B^\sharp$ to define a symplectic form on M. A Poisson manifold admits a (singular) foliation such that the leaves are symplectic. By a theorem of A.

Weinstien we can locally in a neighborhood of a point p find a coordinate system (q^i, p_i, w^i) centered at p and such that

$$B = \sum_{i=1}^{k} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i} + \frac{1}{2} \sum_{i,j} a^{ij}() \frac{\partial}{\partial w^i} \wedge \frac{\partial}{\partial w^j}$$

where the smooth functions depend only on the w's. vanish at p. Here k is the dimension of the leave through p. The rank of the map $B^\sharp$ on T_p^*M is k.

Now to give a typical example let $\mathfrak{g}$ be a Lie algebra with bracket $[.,.]$ and $\mathfrak{g}^*$ its dual. Choose a basis $e_1, ..., e_n$ of $\mathfrak{g}$ and the corresponding dual basis $\varepsilon^1, ..., \varepsilon^n$ for $\mathfrak{g}^*$. With respect to the basis $e_1, ..., e_n$ we have

$$[e_i, e_j] = \sum C_{ij}^k e_k$$

where C_{ij}^k are the structure constants.

For any functions $f, g \in \mathfrak{F}(\mathfrak{g}^*)$ we have that df_α, dg_α are linear maps $\mathfrak{g}^* \to \mathbb{R}$ where we identify $T_\alpha \mathfrak{g}^*$ with $\mathfrak{g}^*$. This means that df_α, dg_α can be considered to be in $\mathfrak{g}$ by the identification $\mathfrak{g}^{**} = \mathfrak{g}$. Now define the $\pm$ Poisson structure on $\mathfrak{g}^*$ by

$$\{f, g\}_\pm(\alpha) = \pm\alpha([df_\alpha, dg_\alpha])$$

Now the basis $e_1, ..., e_n$ is a coordinate system y on $\mathfrak{g}^*$ by $y_i(\alpha) = \alpha(e_i)$.

Proposition 24.1 *In terms of this coordinate system the Poisson bracket just defined is*

$$\{f, g\}_\pm = \pm \sum_{i=1}^{n} B_{ij} \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j}$$

where $B_{ij} = \sum C_{ij}^k y_k$.

Proof. We suppress the $\pm$ and compute:

$$\begin{aligned}
\{f, g\} &= [df, dg] = [\sum \frac{\partial f}{\partial y_i} dy_i, \sum \frac{\partial g}{\partial y_j} dy_j] \\
&= \sum \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} [dy_i, dy_j] = \sum \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} \sum C_{ij}^k y_k \\
&= \sum_{i=1}^{n} B_{ij} \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j}
\end{aligned}$$

■

Chapter 25

Quantization

25.1 Operators on a Hilbert Space

A bounded linear map between complex Hilbert spaces $A : \mathcal{H}_1 \to \mathcal{H}_2$ is said to be of finite rank iff the image $A(\mathcal{H}_1) \subset \mathcal{H}_2$ is finite dimensional. Let us denote the set of all such finite rank maps by $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$. If $\mathcal{H}_1 = \mathcal{H}_2$ we write $\mathcal{F}(\mathcal{H})$. The set of bounded linear maps is denoted $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and is a Banach space with norm given by $\|A\|_{\mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)} := \sup\{\|Av\| : \|v\| = 1\}$. Let $\mathcal{B}(\mathcal{H})$ be the algebra of bounded operators which is then a Banach space with the norm given by $\|A\|_{\mathcal{B}(\mathcal{H})} := \sup\{\|Av\| : \|v\| = 1\}$. The reader will recall that $\mathcal{B}(\mathcal{H})$ is in fact a Banach algebra meaning that in we have

$$\|AB\|_{\mathcal{B}(\mathcal{H})} \leq \|A\|_{\mathcal{B}(\mathcal{H})} \|B\|_{\mathcal{B}(\mathcal{H})} .$$

We will abbreviate $\|A\|_{\mathcal{B}(\mathcal{H})}$ to just $\|A\|$ when convenient. Recall that the adjoint of an element $A \in \mathcal{B}(\mathcal{H})$ is that unique element A^* defined by $\langle A^*v, w\rangle = \langle v, Aw\rangle$ for all $v, w \in \mathcal{H}$. One can prove that in fact we have

$$\|A^*A\| = \|A\|^2$$

One can define on $\mathcal{B}(\mathcal{H})$ two other important topologies. Namely, the **strong topology** is given by the family of seminorms $A \mapsto \|Av\|$ for various $v \in \mathcal{H}$ and the **weak topology** given by the family of seminorms $A \mapsto \langle Av, w\rangle$ for various $v, w \in \mathcal{H}$. Because of this we will use phrases like "norm-closed", "strongly closed" or "weakly closed" etc.

Recall that for any set $W \subset \mathcal{H}$ we define $W^{\perp} := \{v : \langle v, w\rangle = 0$ for all $w \in W\}$.The following fundamental lemma is straightforward to prove.

Lemma 25.1 *For any $A \in \mathcal{B}(\mathcal{H})$ we have*

$$\begin{aligned} \overline{\mathrm{img}(A^*)} &= \ker(A)^{\perp} \\ \ker(A^*) &= \mathrm{img}(A)^{\perp}. \end{aligned}$$

Lemma 25.2 *$\mathcal{F}(\mathcal{H})$ has the following properties:*
1) $F \in \mathcal{F}(\mathcal{H}) \Longrightarrow A \circ F \in \mathcal{F}(\mathcal{H})$ and $F \circ A \in \mathcal{F}(\mathcal{H})$ for all $A \in \mathcal{B}(\mathcal{H})$.
2) $F \in \mathcal{F}(\mathcal{H}) \Longrightarrow F^* \in \mathcal{F}(\mathcal{H})$.
*Thus $\mathcal{F}(\mathcal{H})$ is a two sided ideal of $\mathcal{B}(\mathcal{H})$ closed under *.*

Proof. Obvious. ■

Definition 25.1 *An linear map $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is called compact if $A(\overline{B(0,1)})$ is relatively compact (has compact closure) in $\mathcal{H}_2$. The set of all compact operators $\mathcal{H}_1 \to \mathcal{H}_2$, is denoted $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ or if $\mathcal{H}_1 = \mathcal{H}_2$ by $\mathcal{K}(\mathcal{H})$.*

Lemma 25.3 *$\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ is dense in $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ is closed in $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$.*

Proof. Let $A \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. For every small $\epsilon > 0$ there is a finite subset $\{y_1, y_2, .., y_k\} \subset \overline{B(0,1)}$ such that the $\epsilon-$balls with centers $y_1, y_2, .., y_k$ cover $A(B(0,1))$. Now let P be the projection onto the span of $\{Ay_1, Ay_2, .., Ay_k\}$. We leave it to the reader to show that $PA \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ and $\|PA - A\| < \epsilon$. It follows that $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ is dense in $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$.

Let $\{A_n \}_{n>0}$ be a sequence in $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ which converges to some $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. We want to show that $A \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. Let $\epsilon > 0$ be given and choose $n_0 > 0$ so that $\|A_n - A\| < \epsilon$. Now A_{n_0} is compact so $A_{n_0}(\overline{B(0,1)})$ is relatively compact. It follows that there is some finite subset $N \subset \overline{B(0,1)}$ such that

$$A_{n_0}(\overline{B(0,1)}) \subset \bigcup_{a \in N} B(a, \epsilon).$$

Then for any $y \in \overline{B(0,1)}$ there is an element $a \in N$ such that

$$\begin{aligned}
&\|A(y) - A(a)\| \\
&\leq \|A(y) - A_{n_0}(y)\| + \|A_{n_0}(y) - A_{n_0}(a)\| + \|A_{n_0}(a) - A(a)\| \\
&\leq 3\epsilon
\end{aligned}$$

and it follows that $A(\overline{B(0,1)})$ is relatively compact. Thus $A \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. ■

Corollary 25.1 *$\mathcal{K}(\mathcal{H})$ is a two sided ideal in $\mathcal{B}(\mathcal{H})$.*

We now give two examples (of types) of compact operators each of which is prototypical in its own way.

Example 25.1 *For every continuous function $K(.,.) \in L^2([0,1] \times [0,1])$ gives rise to a compact integral operator $\widetilde{K}$ on $L^2([0,1])$ defined by*

$$(A_K f)(x) := \int_{[0,1]\times[0,1]} K(x,y) f(y) dy.$$

Example 25.2 *If $\mathcal{H}$ is separable and we fix a Hilbert basis $\{e_i\}$ then we can define an operator by prescribing its action on the basis elements. For every sequence of numbers $\{\lambda_i\} \subset \mathbb{C}$ such that $\lambda_i \to 0$ we get an operator $A_{\{\lambda_i\}}$ defined by*

$$A_{\{\lambda_i\}}e_i = \lambda_i e_i.$$

This operator is the norm limit of the finite rank operators A_n defined by

$$A_n(e_i) = \begin{cases} \lambda_i e_i \text{ if } i \leq n \\ \\ 0 \text{ if } i > n \end{cases}.$$

*Thus $A_{\{y_i\}}$ is compact by lemma 25.3. This type of operator is called a **diagonal operator** (with respect to the basis $\{e_i\}$).*

25.2 C*-Algebras

Definition 25.2 *A **C* algebra of (bounded) operators** on $\mathcal{H}$ is a norm-closed subalgebra $\mathfrak{A}$ of $\mathcal{B}(\mathcal{H})$ such that*

$$A^* \in \mathfrak{A} \text{ for all } A \in \mathfrak{A}.$$

A trivial example is the space $\mathcal{B}(\mathcal{H})$ itself but beyond that we have the following simple but important example:

Example 25.3 *Consider the case where $\mathcal{H}$ is $L^2(X, \mu)$ for some compact space X and where the measure μ is a positive measure (i.e. non-negative) such that $\mu(U) > 0$ for all nonempty open sets U. Now for every function $f \in C(X)$ we have a multiplication operator $M_f : L^2(X, \mu) \to L^2(X, \mu)$ defined by $M_f(g) = fg$. The map $f \mapsto M_f$ is an algebra monomorphism from $C(X)$ into $\mathcal{B}(\mathcal{H}) := \mathcal{B}(L^2(X, \mu))$ such that $\overline{f} \mapsto M_f^* = M_{\overline{f}}$. Thus we identify $C(X)$ with a subspace of $\mathcal{B}(\mathcal{H})$ which is in fact a C^* algebra of operators on $\mathcal{H}$. This is an example of a **commutative** C^***-algebra.***

Example 25.4 (!) *Using corollary 25.1 one can easily see that $\mathcal{K}(\mathcal{H})$ is a C^* algebra of operators on $\mathcal{H}$. The fact that $\mathcal{K}(\mathcal{H})$ is self adjoint (closed under adjoint $A \mapsto A^*$) follows from the self adjointness of the algebra $\mathcal{F}(\mathcal{H})$.*

Definition 25.3 *A C^* algebra of operators on $\mathcal{H}$ is called separable if it has a countable dense subset.*

Proposition 25.1 *The C^* algebra $\mathcal{B}(\mathcal{H})$ itself is separable iff $\mathcal{H}$ is finite dimensional.*

Remark 25.1 *If one gives $\mathcal{B}(\mathcal{H})$ the* strong *topology then $\mathcal{B}(\mathcal{H})$ is separable iff $\mathcal{H}$ is separable.*

Proposition 25.2 *The algebra of multipliers $\mathfrak{M} \cong C(X)$ from example 25.3 is separable iff X is a separable compact space.*

In the case that $\mathcal{H}$ is finite dimensional we may as well assume that $\mathcal{H} = \mathbb{C}^n$ and then we can also identify $\mathcal{B}(\mathcal{H})$ with the algebra of matrices $\mathbb{M}_{n\times n}(\mathbb{C})$ where now A^* refers to the conjugate transpose of the matrix A. On can also verify that in this case the "operator" norm of a matrix A is given by the maximum of the eigenvalues of the self-adjoint matrix A^*A. Of course one also has

$$\|A\| = \|A^*A\|^{1/2}.$$

25.2.1 Matrix Algebras

It will be useful to study the finite dimensional case in order to gain experience and put the general case in perspective.

25.3 Jordan-Lie Algebras

In this section we will not assume that algebras are necessarily associative or commutative. We also do not always require an algebra to have a unity. If we consider the algebra $\mathbb{M}_{n\times n}(\mathbb{R})$ of $n \times n$ matrices and define a new product $\diamond$ by

$$A \diamond B := \frac{1}{2}(AB + BA)$$

then the resulting algebra $\mathbb{M}_{n\times n}(\mathbb{R}), \diamond$ is not associative and is an example of a so called Jordan algebra. This kind of algebra plays an important role in quantization theory. The general definition is as follows.

Definition 25.4 *A **Jordan algebra** $\mathfrak{A}$,$\diamond$ is a commutative algebra such that*

$$A \diamond (B \diamond A^2) = (A \diamond B) \diamond A^2.$$

In the most useful cases we have more structure which abstracts as the following:

Definition 25.5 *A Jordan Morphism $h : \mathfrak{A}, \diamond \to \mathfrak{B}, \diamond$ between Jordan algebras is a linear map satisfying $h(A \diamond B) = h(A) \diamond h(B)$ for all $A, B \in \mathfrak{A}$.*

Definition 25.6 *A **Jordan-Lie algebra** is a real vector space $\mathfrak{A}_{\mathbb{R}}$ which two bilinear products $\diamond$ and $\{.,.\}$ such that*

1)

$$A \diamond B = B \diamond A$$
$$\{A, B\} = -\{B, A\}$$

2) For each $A \in \mathfrak{A}_{\mathbb{R}}$ the map $B \mapsto \{A, B\}$ is a derivation for both $\mathfrak{A}_{\mathbb{R}}, \diamond$ and $\mathfrak{A}_{\mathbb{R}}, \{.,.\}$.

*3) There is a constant $\hbar$ such that we have the following **associator identity**:*

$$(A \diamond B) \diamond C - (A \diamond B) \diamond C = \frac{1}{4}\hbar^2\{\{A, , C\}, B\}.$$

Observe that this definition actually says that a **Jordan-Lie algebra** is actually two algebras coupled by the associator identity and the requirements concerning the derivation property. The algebra $\mathfrak{A}_{\mathbb{R}}, \diamond$ is a Jordan algebra and $\mathfrak{A}_{\mathbb{R}}, \{.,.\}$ is a Lie algebra.

Notice that if a Jordan-Lie algebra $\mathfrak{A}_{\mathbb{R}}$ is associative then by definition it is a Poisson algebra.

Because we have so much structure we will be interested in maps which preserve some or all of the structure and this leads us the following definition:

Definition 25.7 *A **Jordan** (resp. **Poisson**) morphism $h : \mathfrak{A}_{\mathbb{R}} \to \mathfrak{B}_{\mathbb{R}}$ between Jordan-Lie algebras is a Jordan morphism (resp. Lie algebra morphism) on the underlying algebras $\mathfrak{A}_{\mathbb{R}}, \diamond$ and $\mathfrak{B}_{\mathbb{R}}, \diamond$ (resp. underlying Lie algebras $\mathfrak{A}_{\mathbb{R}}, \{.,.\}$ and $\mathfrak{B}_{\mathbb{R}}, \{.,.\}$). A map $h : \mathfrak{A}_{\mathbb{R}} \to \mathfrak{B}_{\mathbb{R}}$ which is simultaneously a Jordan morphism and a Poisson morphism is called a (Jordan-Lie) **morphism**. In each case we also have the obvious definitions for isomorphism.*

Chapter 26

Appendices

26.1 A. Primer for Manifold Theory

After imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers. In fact, since a Euclidean space in this sense is an object of intuition (at least in 2d and 3d) some may insist that to be sure such a space of point really exists that we should in fact start with $\mathbb{R}^n$ and "forget" the origin and all the vector space structure while retaining the notion of point and distance. The coordinatization of Euclidean space is then just a "remembering" of this forgotten structure. Thus our coordinates arise from a map $x : E^n \to \mathbb{R}^n$ which is just the identity map. This approach has much to recommend it and we shall more or less follow this canonical path. There is at least one regrettable aspect to this approach which is the psychological effect that occurs when we impose other coordinates on our system an introduce differentiable manifolds as abstract geometric objects that support coordinate systems. It might seem that this is a big abstraction and when the definitions of charts and atlases and so on appear a certain notational fastidiousness sets in that somehow creates a psychological gap between open set in $\mathbb{R}^n$ and the abstract space that we coordinatize. But what is now lost from sight is that we have already been dealing with an abstract manifolds E^n which we have identified with $\mathbb{R}^n$ but could just as easily supports other coordinates such as spherical coordinates. What competent calculus student would waste time thinking of polar coordinates as given a map $E^2 \to \mathbb{R}^2$ (defined on a proper open subset of course) and then wonder whether something like $drd\phi d\theta$ lives on the original E^n or in the image of the coordinate map $E^n \to \mathbb{R}^n$? It has become an unfortunate consequence of the modern viewpoint that simple geometric ideas are lost from the notation. Ideas that allow one to think about "quantities and their variations" and then comfortably write things like

$$rdr \wedge d\theta = dx \wedge dy$$

without wondering if it shouldn't be a "pullback" $\psi_{12}^*(rdr \wedge d\theta) = dx \wedge dy$ where ψ_{12} is the change of coordinate map $\mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\begin{aligned} x(r,\theta) &= r\cos\theta \\ y(r,\theta) &= r\cos\theta \end{aligned}$$

Of course, the experienced differential geometer understands the various meaning and the contextual understanding which removes ambiguity. The student, on the other hand, is faced with a pedagogy that teaches notation, trains one to examine each, equation for logical self consistence, but fails to teach geometric intuition. Having made this complain the author must confess that he too will use the modern notation and will not stray far from standard practice. These remarks are meant to encourage the student to stop and seek the simplest most intuitive viewpoint whenever feeling overwhelmed by notation. The student is encouraged to experiment with abbreviated personal notation when checking calculations and to draw diagrams and schematics that encode the geometric ideas whenever possible. "The picture writes the equations".

So, as we said, after imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers. We will envision there to be a copy $\mathbb{R}^n_p$ of $\mathbb{R}^n$ at each of its points $p \in \mathbb{R}^n$. The elements of $\mathbb{R}^n_p$ are to be though of as the vectors based at p, that is, the "tangent vectors". These tangent spaces are related to each other by the obvious notion of vectors being parallel (this is exactly what is not generally possible for tangents spaces of a manifold). For the standard basis vectors e_j (relative to the coordinates x_i) taken as being based at p we often write $\left.\frac{\partial}{\partial x_i}\right|_p$ and this has the convenient second interpretation as a differential operator acting on smooth functions defined near $p \in \mathbb{R}^n$. Namely,

$$\left.\frac{\partial}{\partial x_i}\right|_p f = \frac{\partial f}{\partial x_i}(p).$$

An n-tuple of smooth functions $X^1, ..., X^n$ defines a smooth vector field $X = \sum X^i \frac{\partial}{\partial x_i}$ whose value at p is $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$. Thus a vector field assigns to each p in its domain, an open set U, a vector $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$ at p. We may also think of vector field as a differential operator via

$$\begin{aligned} f &\mapsto Xf \in C^\infty(U) \\ (Xf)(p) &:= \sum X^i(p) \frac{\partial f}{\partial x_i}(p) \end{aligned}$$

Example 26.1 $X = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}$ *is a vector field defined on* $U = \mathbb{R}^2 - \{0\}$ *and* $(Xf)(x,y) = y\frac{\partial f}{\partial x}(x,y) - x\frac{\partial f}{\partial y}(x,y)$.

Notice that we may certainly add vector fields defined over the same open set as well as multiply by functions defined there:

$$(fX + gY)(p) = f(p)X(p) + g(p)X(p)$$

The familiar expression $df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n$ has the intuitive interpretation expressing how small changes in the variables of a function give rise to small changes in the value of the function. Two questions should come to mind. First, "what does 'small' mean and how small is small enough?" Second, "which direction are we moving in the coordinate" space? The answer to these questions lead to the more sophisticated interpretation of df as being a linear functional on each tangent space. Thus we must choose a direction v_p at $p \in \mathbb{R}^n$ and then $df(v_p)$ is a number depending linearly on our choice of vector v_p. The definition is determined by $dx_i(e_j) = \delta_{ij}$. In fact, this shall be the basis of our definition of df at p. We want

$$Df|_p \left(\left. \frac{\partial}{\partial x_i} \right|_p \right) := \frac{\partial f}{\partial x_i}(p).$$

Now any vector at p may be written $v_p = \sum_{i=1}^n v^i \left. \frac{\partial}{\partial x_i} \right|_p$ which invites us to use v_p as a differential operator (at p):

$$v_p f := \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p) \in \mathbb{R}$$

This consistent with our previous statement about a vector field being a differential operator simply because $X(p) = X_p$ is a vector at p for every $p \in U$. This is just the directional derivative. In fact we also see that

$$\begin{aligned} Df|_p (v_p) &= \sum_j \frac{\partial f}{\partial x_j}(p) dx_j \left(\sum_{i=1}^n v^i \left. \frac{\partial}{\partial x_i} \right|_p \right) \\ &= \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p) = v_p f \end{aligned}$$

so that our choices lead to the following definition:

Definition 26.1 *Let f be a smooth function on an open subset U of $\mathbb{R}^n$. By the symbol df we mean a family of maps $Df|_p$ with p varying over the domain U of f and where each such map is a linear functional of tangent vectors based at p given by $Df|_p (v_p) = v_p f = \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p)$.*

Definition 26.2 *More generally, a smooth 1-form α on U is a family of linear functionals $\alpha_p : T_p\mathbb{R}^n \to \mathbb{R}$ with $p \in U$ which is smooth is the sense that $\alpha_p(\left. \frac{\partial}{\partial x_i} \right|_p)$ is a smooth function of p for all i.*

From this last definition it follows that if $X = X^i \frac{\partial}{\partial x_i}$ is a smooth vector field then $\alpha(X)(p) := \alpha_p(X_p)$ defines a smooth function of p. Thus an alternative way to view a 1−form is as a map $\alpha : X \mapsto \alpha(X)$ which is defined on vector fields and linear over the algebra of smooth functions $C^\infty(U)$:

$$\alpha(fX + gY) = f\alpha(X) + g\alpha(Y).$$

26.1.1 Fixing a problem

Now it is at this point that we want to destroy the privilege of the rectangular coordinates and express our objects in an arbitrary coordinate system smoothly related to the existing coordinates. This means that for any two such coordinate systems, say $u^1, ..., u^n$ and $y^1, ..., y^n$ we want to have the ability to express fields and forms in either system and have for instance

$$X^i_{(y)} \frac{\partial}{\partial y_i} = X = X^i_{(u)} \frac{\partial}{\partial u_i}$$

for appropriate functions $X^i_{(y)}$, $X^i_{(u)}$. This equation only makes sense on the overlap of the domains of the coordinate systems. To be consistent with the chain rule we must have

$$\frac{\partial}{\partial y^i} = \frac{\partial u^j}{\partial y^i} \frac{\partial}{\partial u^j}$$

which then forces the familiar transformation law:

$$\sum \frac{\partial u^j}{\partial y^i} X^i_{(y)} = X^i_{(u)}$$

We think of $X^i_{(y)}$ and $X^i_{(u)}$ as referring to or representing the same geometric reality from two different coordinate systems. No big deal right? We how about the fact, that there is this underlying abstract space that we are coordinatizing? That too is no big deal. We were always doing it in calculus anyway. What about the fact that the coordinate systems aren't defined as a 1-1 correspondence with the points of the space unless we leave out some point in some coordinates like we leave out the origin to avoid ambiguity in θ and have a nice open domain. Well if this is all fine then we may as well imagine other abstract spaces that support coordinates in this way. In fact, we don't have to look far for an example. Any surface such as the sphere will do. We can talk about 1-forms like say $\alpha = \theta d\phi + \phi \sin(\theta) d\theta$, or a vector field tangent to the sphere $\theta \sin(\phi) \frac{\partial}{\partial \theta} + \theta^2 \frac{\partial}{\partial \phi}$ and so on (just pulling things out of a hat). We just have to be clear about how these arise and most of all how to change to a new coordinate expression for the same object. This is the approach of tensor analysis. An object called a 2-tensor T is represented in two different coordinate systems as for instance

$$\sum T^{ij}_{(y)} \frac{\partial}{\partial y^i} \otimes \frac{\partial}{\partial y^j} = \sum T^{ij}_{(u)} \frac{\partial}{\partial u^i} \otimes \frac{\partial}{\partial u^j}$$

where all we really need to know for many purposes the transformation law

$$T^{ij}_{(y)} = \sum_{r,s} T^{rs}_{(u)} \frac{\partial y^i}{\partial u^r} \frac{\partial y^i}{\partial u^s}.$$

Then either expression is referring to the same abstract tensor T. This is just a preview but it highlight the approach wherein a transformation laws play a defining role.

26.2 B. Topological Spaces

In this section we briefly introduce the basic notions from point set topology together with some basic examples. We include this section only as a review and a reference since we expect that the reader should already have a reasonable knowledge of point set topology. In the Euclidean coordinate plane $\mathbb{R}^n$ consisting of all n-tuples of real numbers $(x_1, x_2, ...x_n)$ we have a natural notion of distance between two points. The formula for the distance between two points $p_1 = (x_1, x_2, ...x_n)$ and $p_2 = (y_1, y_2, ..., y_n)$ is simply

$$d(p_1, p_2) = \sqrt{\sum (x_i - y_i)^2}. \tag{26.1}$$

The set of all points of distance less than ϵ from a given point p_0 in the plain is denoted $B(p_0, \epsilon)$, i.e.

$$B(p_0, \epsilon) = \{p \in \mathbb{R}^n : d(p, p_0) < \epsilon\}. \tag{26.2}$$

The set $B(p_0, \epsilon)$ is call the **open ball** of radius ϵ and center p_0. A subset S of $\mathbb{R}^2$ is called **open** if every one of its points is the center of an open ball completely contained inside S. The set of all open subsets of the plane has the property that the union of any number of open sets is still open and the intersection of any finite number of open sets is still open. The abstraction of this situation leads to the concept of a topological space.

Definition 26.3 *A set X together with a family $\mathfrak{T}$ of subsets of X is called a topological space if the family $\mathfrak{T}$ has the following three properties.*

1. *$X \in \mathfrak{T}$ and $\emptyset \in \mathfrak{T}$.*
2. *If U_1 and U_2 are both in $\mathfrak{T}$ then $U_1 \cap U_2 \in \mathfrak{T}$ also.*
3. *If $\{U_\alpha\}_{\alpha \in A}$ is any sub-family of $\mathfrak{T}$ indexed by a set A then the union $\bigcup_{\alpha \in A} U_\alpha$ is also in $\mathfrak{T}$.*

In the definition above the family of subsets $\mathfrak{T}$ is called a **topology** on X and the sets in $\mathfrak{T}$ are called **open** sets. The compliment $U^c := X \setminus U$ of an open set U is called **closed** set. The reader is warned that a generic set may be neither open nor closed. Also, some subsets of X might be both open and closed (consider X itself and the empty set). A topology $\mathfrak{T}_2$ is said to be **finer** than a topology $\mathfrak{T}_1$ if $\mathfrak{T}_1 \subset \mathfrak{T}_2$ and in this case we also say that $\mathfrak{T}_1$ is **coarser** than $\mathfrak{T}_2$. We also say that the topology $\mathfrak{T}_1$ is **weaker** than $\mathfrak{T}_2$ and that $\mathfrak{T}_2$ is **stronger** than $\mathfrak{T}_1$.

Neither one of these topologies is generally very interesting but we shall soon introduce much richer topologies. A fact worthy of note in this context is the fact that if $X, \mathfrak{T}$ is a topological space and $S \subset X$ then S inherits a topological structure from X. Namely, a topology on S (called the **relative topology**) is given by

$$\mathfrak{T}_S = \{\text{all sets of the form } S \cap T \text{ where } T \in \mathfrak{T}\} \tag{26.3}$$

In this case we say that S is a topological subspace of X.

Definition 26.4 *A map between topological spaces $f : X \to Y$ is said to be* ***continuous at*** $p \in X$ *if for any open set O containing $f(p)$ there is an open set U containing $p \in X$ such that $f(U) \subset O$. A map $f : X \to Y$ is said to be* ***continuous*** *if it is continuous at each point $p \in X$.*

Proposition 26.1 *$f : X \to Y$ is continuous iff $f^{-1}(O)$ is open for every open set $O \subset Y$.*

Definition 26.5 *A subset of a topological space is called* ***closed*** *if it is the compliment of an open set.*

Closed sets enjoy properties complimentary to those of open sets:

1. The whole space X and the empty set $\emptyset$ are both closed.
2. The intersection of any family of closed sets is a closed set.
3. The union of a finite number of closed sets is closed.

Since the intersection of closed sets is closed every set $S \subset X$ is contained in a closed set which is the smallest of all closed sets containing S which is called the closure of S and is denoted by $\overline{S}$. The **closure** $\overline{S}$ is the intersection of all closed subsets containing S:

$$\overline{S} = \bigcap_{S \subset F} F.$$

Similarly, the **interior** of a set S is the largest open set contained in S and is denoted by $\overset{\circ}{S}$. A point $p \in S \subset X$ is called an **interior point** of S if there is an open set containing p and contained in S. The interior of S is just the set of all its interior points. It may be shown that $\overset{\circ}{S} = (\overline{S^c})^c$

Definition 26.6 *The (topological)* ***boundary*** *of a set $S \subset X$ is $\partial S := \overline{S} \cap \overline{S^c}$ and*

We say that a set $S \subset X$ is **dense** in X if $\overline{S} = X$.

Definition 26.7 *A subset of a topological space $X, \mathfrak{T}$ is called* ***cloven*** *if it is both open and closed.*

Definition 26.8 *A topological space X is called* ***connected*** *if it is not the union of two proper cloven set. Here, proper means not X or $\emptyset$. A topological space X is called* ***path connected*** *if for every pair of points $p, q \in X$ there is a continuous map $c : [a, b] \to X$ (a path) such that $c(a) = q$ and $c(b) = p$. (Here $[a, b] \subset \mathbb{R}$ is endowed with the relative topology inherited from the topology on $\mathbb{R}$.)*

Example 26.2 *The unit sphere S^2 is a topological subspace of the Euclidean space $\mathbb{R}^3$.*

Let X be a set and $\{\mathfrak{T}_\alpha\}_{\alpha\in A}$ any family of topologies on X indexed by some set A. The the intersection

$$\mathfrak{T} = \bigcap_{\alpha\in A} \mathfrak{T}_\alpha$$

is a topology on X. Furthermore, $\mathfrak{T}$ is coarser that every $\mathfrak{T}_\alpha$.

Given any family $\mathfrak{F}$ of subsets of X there exists a weakest (coarsest) topology containing all sets of $\mathfrak{F}$. We will denote this topology by $\mathfrak{T}(\mathfrak{F})$.

One interesting application of this is the following; Given a family of maps $\{f_\alpha\}$ from a set S to a topological space $Y, \mathfrak{T}_Y$ there is a coarsest topology on S such that all of the maps f_α are continuous. This topology will be denoted $\mathfrak{T}_{\{f_\alpha\}}$ and is called the topology generated by the family of maps $\{f_\alpha\}$.

Definition 26.9 *If X and Y are topological spaces then we define the **product topology** on $X\times Y$ as the topology generated by the projections $pr_1 : X\times Y \to X$ and $pr_2 : X \times Y \to Y$.*

Definition 26.10 *If $\pi : X \to Y$ is a surjective map where X is a topological space but Y is just a set. Then the **quotient topology** is the topology generated by the map π. In particular, if $A \subset X$ we may form the set of equivalence classes X/A where $x \sim y$ if both are in A or they are equal. The the map $x \mapsto [x]$ is surjective and so we may form the quotient topology on X/A.*

Let X be a topological space and $x \in X$. A family of subsets $\mathcal{B}_x$ all of which contain x is called an **open neighborhood base** at x if every open set containing x contains (as a superset) an set from $\mathcal{B}_x$. If X has a countable open base at each $x \in X$ we call X **first countable**.

A subfamily $\mathcal{B}$ is called a **base** for a topology $\mathfrak{T}$ on X if the topology $\mathfrak{T}$ is exactly the set of all unions of elements of $\mathcal{B}$. If X has a countable base for its given topology we say that X is a **second countable** topological space.

By considering balls of rational radii and rational centers one can see that $\mathbb{R}^n$ is first and second countable.

26.2.1 Separation Axioms

Another way to classify topological spaces is according to the following scheme:

(**Separation Axioms**)

A topological space $X, \mathfrak{T}$ is called a T_0 space if given $x, y \in X$, $x \neq y$, there exists either an open set containing x, but not y or the other way around (We don't get to choose which one).

A topological space $X, \mathfrak{T}$ is called T_1 if whenever given any x,$y \in X$ there is an open set containing x but not y (and the other way around;we get to do it either way).

A topological space $X, \mathfrak{T}$ is called T_2 or **Hausdorff** if whenever given any two points x, $y \in X$ there are disjoint open sets U_1 and U_2 with $x \in U_1$ and $y \in U_2$.

A topological space $X, \mathfrak{T}$ is called T_3 or **regular** if whenever given a closed set $F \subset X$ and a point $x \in X \backslash F$ there are disjoint open sets U_1 and U_2 with $x \in U_1$ and $F \subset U_2$

A topological space $X, \mathfrak{T}$ is called T_4 or **normal** if given any two disjoint closed subsets of X, say F_1 and F_2, there are two disjoint open sets U_1 and U_2 with $F_1 \subset U_1$ and $F_2 \subset U_2$.

Lemma 26.1 (Urysohn) *Let X be normal and $F, G \subset X$ closed subsets with $F \cap G = \emptyset$. Then there exists a continuous function $f : X \to [0,1] \subset \mathbb{R}$ such that $f(F) = 0$ and $f(G) = 1$.*

A open cover of topological space X (resp. subset $S \subset X$) a collection of open subsets of X, say$\{U_\alpha\}$, such that $X = \bigcup U_\alpha$ (resp. $S \subset \bigcup U_\alpha$). For example the set of all open disks of radius $\epsilon > 0$ in the plane covers the plane. A finite cover consists of only a finite number of open sets.

Definition 26.11 *A topological space X is called **compact** if every open cover of X can be reduced to a finite open cover by eliminating some (possibly an infinite number) of the open sets of the cover. A subset $S \subset X$ is called **compact** if it is compact as a topological subspace (i.e. with the relative topology).*

Proposition 26.2 *The continuous image of a compact set is compact.*

26.2.2 Metric Spaces

If the set X has a notion of distance attached to it then we can get an associated topology. This leads to the notion of a metric space.

A set X together with a function $d : X \times X \to \mathbb{R}$ is called a **metric space** if

$d(x, x) \geq 0$ for all $x \in X$

$d(x, y) = 0$ iff $x = y$

$d(x, z) \leq d(x, y) + d(y, z)$ for any $x, y, z \in X$ (*this is called the triangle inequality*).

The function d is called a **metric** or a **distance function**.

Imitating the situation in the plane we can define the notion of an open ball $B(p_0, \epsilon)$ with center p_0 and radius ϵ. Now once we have the metric then we have a topology; we define a subset of a metric space X, d to be open if every point of S is an interior point where a point $p \in S$ is called an **interior point** of S if there is some ball $B(p, \epsilon)$ with center p and (sufficiently small) radius $\epsilon > 0$ completely contained in S. The family of all of these metrically defined open sets forms a topology on X which we will denote by $\mathfrak{T}_d$. It is easy to see that any $B(p, \epsilon)$ is open according to our definition.

If $f : X, d \to Y, \rho$ is a map of metric spaces then f is continuous at $x \in X$ if and only if for every $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ such that if $d(x', x) < \delta(\epsilon)$ then $\rho(f(x'), f(x)) < \epsilon$.

Definition 26.12 *A sequence of elements $x_1, x_2, ...$ of a metric space X, d is said to* ***converge*** *to p if for every $\epsilon > 0$ there is an $N(\epsilon) > 0$ such that if $k > N(\epsilon)$ then $x_k \in B(p, \epsilon)$. A sequence $x_1, x_2, ...$ is called a* ***Cauchy sequence*** *is for every $\epsilon > 0$ there is an $N(\epsilon) > 0$ such that if $k, l > N(\epsilon)$ then $d(x_k, x_l) < \epsilon$. A metric space X, d is said to be* ***complete*** *if every Cauchy sequence also converges.*

A map $f : X, d \to Y, \rho$ of metric spaces is continuous at $x \in X$ if and only if for every sequence x_i converging to x, the sequence $y_i := f(x_i)$ converges to $f(x)$.

26.3 C. Topological Vector Spaces

We shall outline some of the basic definitions and theorems concerning topological vector spaces.

Definition 26.13 *A* ***topological vector space (TVS)*** *is a vector space V with a Hausdorff topology such that the addition and scalar multiplication operations are (jointly) continuous.*

Definition 26.14 *Recall that a neighborhood of a point p in a topological space is a subset which has a nonempty interior containing p. The set of all neighborhoods that contain a point x in a topological vector space V is denoted $\mathcal{N}(x)$.*

The families $\mathcal{N}(x)$ for various x satisfy the following neighborhood axioms

1. Every set which contains a set from $\mathcal{N}(x)$ is also a set from $\mathcal{N}(x)$
2. If N_i is a family of sets from $\mathcal{N}(x)$ then $\bigcap_i N_i \in \mathcal{N}(x)$
3. Every $N \in \mathcal{N}(x)$ contains x
4. If $V \in \mathcal{N}(x)$ then there exists $W \in \mathcal{N}(x)$ such that for all $y \in W$, $V \in \mathcal{N}(y)$.

Conversely, let X be some set. If for each $x \in X$ there is a family $\mathcal{N}(x)$ of subsets of X that satisfy the above neighborhood axioms then there is a uniquely determined topology on X for which the families $\mathcal{N}(x)$ are the neighborhoods of the points x. For this a subset $U \subset X$ is open iff for each $x \in U$ we have $U \in \mathcal{N}(x)$.

Definition 26.15 *A sequence x_n in a TVS is call a Cauchy sequence iff for every neighborhood U of 0 there is a number N_U such that $x_l - x_k \in U$ for all $k, l \geq N_U$.*

Definition 26.16 *A relatively nice situation is when V has a norm which induces the topology. Recall that a* ***norm*** *is a function $\|\| : v \mapsto \|v\| \in \mathbb{R}$ defined on V such that for all $v, w \in \mathsf{V}$ we have*

1. $\|v\| \geq 0$ and $\|v\| = 0$ iff $v = 0$,
2. $\|v + w\| \leq \|v\| + \|w\|$,
3. $\|\alpha v\| = |\alpha| \, \|v\|$ for all $\alpha \in \mathbb{R}$.

   In this case we have a metric on V given by $\text{dist}(v, w) := \|v - w\|$. A **seminorm** is a function $\|\| : v \mapsto \|v\| \in \mathbb{R}$ such that 2) and 3) hold but instead of 1) we require only that $\|v\| \geq 0$.

Definition 26.17 *A **normed space** V is a TVS which has a metric topology given by a norm. That is the topology is generated by the family of all open balls*

$$B_{\mathsf{V}}(\mathsf{x}, r) := \{\mathsf{x} \in \mathsf{V} : \|\mathsf{x}\| < 0\}.$$

Definition 26.18 *A linear map $\ell : \mathsf{V} \to \mathsf{W}$ between normed spaces is called bounded iff there is a constant C such that for all $\mathsf{v} \in \mathsf{V}$ we have $\|\ell \mathsf{v}\|_{\mathsf{W}} \leq C \|\mathsf{v}\|_{\mathsf{V}}$. If ℓ is bounded then the smallest such constant C is*

$$\|\ell\| := \sup \frac{\|\ell \mathsf{v}\|_{\mathsf{W}}}{\|\mathsf{v}\|_{\mathsf{V}}} = \sup\{\|\ell \mathsf{v}\|_{\mathsf{W}} : \|\mathsf{v}\|_{\mathsf{V}} \leq 1\}$$

The set of all bounded linear maps $\mathsf{V} \to \mathsf{W}$ is denoted $\mathcal{B}(\mathsf{V}, \mathsf{W})$. The vector space $\mathcal{B}(\mathsf{V}, \mathsf{W})$ is itself a normed space with the norm given as above.

Definition 26.19 *A **locally convex topological vector space** V is a TVS such that it's topology is generated by a family of seminorms $\{\|.\|_\alpha\}_\alpha$. This means that we give V the weakest topology such that all $\|.\|_\alpha$ are continuous. Since we have taken a TVS to be Hausdorff we require that the family of seminorms is **sufficient** in the sense that for each $\mathsf{x} \in \mathsf{V}$ we have $\bigcap\{x : \|x\|_\alpha = 0\} = \emptyset$. A locally convex topological vector space is sometimes called a locally convex space and so we abbreviate the latter to **LCS**.*

Example 26.3 *Let Ω be an open set in $\mathbb{R}^n$ or any manifold. For each $x \in \Omega$ define a seminorm ρ_x on $C(\Omega)$ by $\rho_x(f) = f(x)$. This family of seminorms makes $C(\Omega)$ a topological vector space. In this topology convergence is pointwise convergence. Also, $C(\Omega)$ is not complete with this TVS structure.*

Definition 26.20 *An LCS which is complete (every Cauchy sequence converges) is called a **Frechet space**.*

Definition 26.21 *A complete normed space is called a **Banach space**.*

Example 26.4 *Suppose that X, μ is a σ-finite measure space and let $p \geq 1$. The set $L^p(X, \mu)$ of all with respect to measurable functions $f : X \to \mathbb{C}$ such that $\int |f|^p \, d\mu \leq \infty$ is a Banach space with the norm $\|f\| := \left(\int |f|^p \, d\mu\right)^{1/p}$. Technically functions equal almost everywhere $d\mu$ must be identified.*

Example 26.5 *The space $C_b(\Omega)$ of bounded continuous functions on Ω is a Banach space with norm given by $\|f\|_\infty := \sup_{x \in \Omega} |f(x)|$.*

Example 26.6 *Once again let Ω be an open subset of $\mathbb{R}^n$. For each compact $K \subset\subset \Omega$ we have a seminorm on $C(\Omega)$ defined by $f \mapsto \|f\|_K := \sup_{x\in K} |f(x)|$. The corresponding convergence is the uniform convergence on compact subsets of Ω. It is often useful to notice that the same topology can be obtained by using $\|f\|_{K_i}$ obtained from a countable sequence of nested compact sets $K_1 \subset K_2 \subset ...$ such that*

$$\bigcup K_n = \Omega.$$

Such a sequence is called an exhaustion of Ω.

If we have topological vector space V and a closed subspace S, then we can form the quotient V/S. The quotient can be turned in to a normed space by introducing as norm

$$\|[x]\|_{\mathrm{V/S}} := \inf_{v\in[x]} \|v\|.$$

If S is not closed then this only defines a seminorm.

Theorem 26.1 *If V is Banach space and a closed subspace* S *a closed (linear) subspace then* V/S *is a Banach space with the above defined norm.*

Proof. Let x_n be a sequence in V such that $[x_n]$ is a Cauchy sequence in V/S. Choose a subsequence such that $\|[x_n] - [x_{n+1}]\| \leq 1/2^n$ for $n = 1, 2, ...$ Setting s_1 equal to zero we find $s_2 \in \mathrm{S}$ such that $\|x_1 - (x_2 + s_2)\|$ and continuing inductively define a sequence s_i such that such that $\{x_n + s_n\}$ is a Cauchy sequence in V. Thus there is an element $y \in \mathrm{V}$ with $x_n + s_n \to y$. But since the quotient map is norm decreasing the sequence $[x_n + s_n] = [x_n]$ must also converge;

$$[x_n] \to [y].$$

■

Remark 26.1 *It is also true that if S is closed and* V/S *is a Banach space then so is* V*.*

26.3.1 Hilbert Spaces

Definition 26.22 *A Hilbert space $\mathcal{H}$ is a complex vector space with a Hermitian inner product $\langle .,. \rangle$. A Hermitian inner product is a bilinear form with the following properties:*

1) $\langle v, w\rangle = \overline{\langle v, w\rangle}$
2) $\langle v, \alpha w_1 + \beta w_2\rangle = \alpha\langle v, w_1\rangle + \beta\langle v, w_2\rangle$
3) $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ only if $v = 0$.

One of the most fundamental properties of a Hilbert space is the projection property.

Theorem 26.2 *If K is a convex, closed subset of a Hilbert space $\mathcal{H}$, then for any given $x \in \mathcal{H}$ there is a unique element $p_K(x) \in \mathcal{H}$ which minimizes the distance $\|x - y\|$ over $y \in K$. That is*

$$\|x - p_K(x)\| = \inf_{y \in K} \|x - y\| .$$

If K is a closed linear subspace then the map $x \mapsto p_K(x)$ is a bounded linear operator with the projection property $p_K^2 = p_K$.

Definition 26.23 *For any subset $S \in \mathcal{H}$ we have the orthogonal compliment $S^\perp$ defined by*

$$S^\perp = \{x \in \mathcal{H} : \langle x, s\rangle = 0 \text{ for all } s \in S\}.$$

$S^\perp$ is easily seen to be a linear subspace of $\mathcal{H}$. Since $\ell_s : x \mapsto \langle x, s\rangle$ is continuous for all s and since

$$S^\perp = \cap_s \ell_s^{-1}(0)$$

we see that $S^\perp$ is closed. Now notice that since by definition

$$\|x - P_s x\|^2 \leq \|x - P_s x - \lambda s\|^2$$

for any $s \in S$ and any real λ we have $\|x - P_s x\|^2 \leq \|x - P_s x\|^2 - 2\lambda\langle x - P_s x, s\rangle + \lambda^2 \|s\|^2$. Thus we see that $p(\lambda) := \|x - P_s x\|^2 - 2\lambda\langle x - P_s x, s\rangle + \lambda^2 \|s\|^2$ is a polynomial in λ with a minimum at $\lambda = 0$. This forces $\langle x - P_s x, s\rangle = 0$ and so we see that $x - P_s x$. From this we see that any $x \in \mathcal{H}$ can be written as $x = x - P_s x + P_s x = s + s^\perp$. On the other hand it is easy to show that $S^\perp \cap S = 0$. Thus we have $\mathcal{H} = S \oplus S^\perp$ for any closed linear subspace $S \subset \mathcal{H}$. In particular the decomposition of any x as $s + s^\perp \in S \oplus S^\perp$ is unique.

26.3.2 Orthonormal sets

26.4 D. Overview of Classical Physics

26.4.1 Units of measurement

In classical mechanics we need units for measurements of length, time and mass. These are called elementary units. WE need to add a measure of electrical current to the list if we want to study electromagnetic phenomenon. Other relevant units in mechanics are derived from these alone. For example, speed has units of length$\times$time^{-1}, volume has units of length$\times$length$\times$length kinetic energy has units of mass$\times$length$\times$length$\times$length$\times$time$^{-1}\times$time^{-1} and so on. A common system, called the SI system uses meters (m), kilograms (km) and seconds (sec) for length, mass and time respectively. In this system, the unit of energy kg$\times$m^{2}sec^{-2} is called a joule. The unit of force in this system is Newton's and decomposes into elementary units as kg$\times$m$\times$sec^{-2}.

26.4.2 Newton's equations

The basic assumptions of Newtonian mechanics can be summarized by saying that the set of all mechanical events M taking place in ordinary three dimensional space is such that we can impose on this set of events a coordinate system called an inertial coordinate system. An inertial coordinate system is first of all a 1-1 correspondence between events and the vector space $\mathbb{R} \times \mathbb{R}^3$ consisting of 4-tuples (t, x, y, z). The laws of mechanics are then described by equations and expressions involving the variables (t, x, y, z) written $(t, \mathbf{x})$ where $\mathbf{x} = (x, y, z)$. There will be many correspondences between the event set and $\mathbb{R} \times \mathbb{R}^3$ but not all are inertial. An inertial coordinate system is picked out by the fact that the equations of physics take on a particularly simple form in such coordinates. Intuitively, the x, y, z variables locate an event in space while t specifies the time of an event. Also, x, y, z should be visualized as determined by measuring against a mutually perpendicular set of three axes and t is measured with respect to some sort of clock with $t = 0$ being chosen arbitrarily according to the demands of the experimental situation. Now we expect that the laws of physics should not prefer any particular such choice of mutually perpendicular axes or choice of starting time. Also, the units of measurement of length and time are conventionally determined by human beings and so the equations of the laws of physics should in some way not depend on this choice in any significant way. Careful consideration along these lines leads to a particular set of "coordinate changes" or transformations which translate among the different inertial coordinate systems. The group of transformations which is chosen for classical (non-relativistic) mechanics is the so called Galilean group Gal.

Definition 26.24 *A map $g : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R} \times \mathbb{R}^3$ is called a Galilean transformation iff it can be decomposed as a composition of transformations of the following type:*

1. Translation of the origin:
$$(t, \mathbf{x}) \mapsto (t + t_0, \mathbf{x} + \mathbf{x}_0)$$

2. Uniform motion with velocity $\mathbf{v}$:
$$(t, \mathbf{x}) \mapsto (t, \mathbf{x} + t\mathbf{v})$$

3. Rotation of the spatial axes:
$$(t, \mathbf{x}) \mapsto (t, R\mathbf{x})$$
where $R \in O(3)$.

If $(t, \mathbf{x})$ are inertial coordinates then so will $(T, \mathbf{X})$ be inertial coordinates iff $(T, \mathbf{X}) = g(t, \mathbf{x})$ for some Galilean transformation. We will take this as given.

The motion of a idealized point mass moving in space is described in an inertial frame $(t, \mathbf{x})$ as a curve $t \mapsto c(t) \in \mathbb{R}^3$ with the corresponding curve $t \mapsto$

$(t, c(t))$ in the (coordinatized) event space $\mathbb{R} \times \mathbb{R}^3$. We often write $\mathbf{x}(t)$ instead of $c(t)$. If we have a system of n particles then we may formally treat this as a single particle moving in an $3n-$dimensional space and so we have a single curve in $\mathbb{R}^{3n}$. Essentially we are concatenating the spatial part of inertial coordinates $\mathbb{R}^{3n} = \mathbb{R}^3 \times \cdots \mathbb{R}^3$ taking each factor as describing a single particle in the system so we take $\mathbf{x} = (x_1, y_1, z_1, ..., x_n, y_n, z_n)$. Thus our new inertial coordinates may be thought of as $\mathbb{R} \times \mathbb{R}^{3n}$. If we have a system of particles it will be convenient to define the momentum vector $\mathbf{p} = (m_1 x_1, m_1 y_1, m_1 z_1, ..., m_n x_n, m_n y_n, m_n z_n) \in \mathbb{R}^{3n}$. In such coordinates, Newton's law for n particles of masses $m_1, ..., m_n$ reads

$$\frac{d^2\mathbf{p}}{dt^2} = \mathbf{F}(\mathbf{x}(t), t)$$

where $t \mapsto \mathbf{x}(t)$ describes the motion of a system of n particles in space as a smooth path in $\mathbb{R}^{3n}$ parameterized by t representing time. The equation has units of force (Newton's in the SI system). If all bodies involved are taken into account then the force $\mathbf{F}$ cannot depend explicitly on time as can be deduced by the assumption that the form taken by $\mathbf{F}$ must be the same in any inertial coordinate system. We may not always be able to include explicitly all involved bodies and so it may be that our mathematical model will involve a changing force $\mathbf{F}$ exerted on the system from without as it were. As an example consider the effect of the tidal forces on sensitive objects on earth. Also, the example of earths gravity shows that if the earth is not taken into account as one of the particles in the system then the form of $\mathbf{F}$ will not be invariant under all spatial rotations of coordinate axes since now there is a preferred direction (up-down).

26.4.3 Classical particle motion in a conservative field

There are special systems for making measurements that can only be identified in actual practice by interaction with the physical environment. In classical mechanics, a point mass will move in a straight line unless a force is being applied to it. The coordinates in which the mathematical equations describing motion are the simplest are called inertial coordinates (x, y, z, t). If we consider a *single* particle of mass m then Newton's law simplifies to

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F}(\mathbf{x}(t), t)$$

The force $\mathbf{F}$ is **conservative** if it doesn't depend on time and there is a potential function $V : \mathbb{R}^3 \to \mathbb{R}$ such that $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$. Assume this is the case. Then Newton's law becomes

$$m\frac{d^2}{dt^2}\mathbf{x}(t) + \operatorname{grad} V(\mathbf{x}(t)) = 0.$$

Newton's equations are often written

$$\mathbf{F}(\mathbf{x}(t)) = m\mathbf{a}(t)$$

$\mathbf{F}:\mathbb{R}^3 \to \mathbb{R}^3$ is the force function and we have taken it to not depend explicitly on time t. The force will be conservative so $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$ for some scalar function $V(\mathbf{x})$. The total energy or Hamiltonian function is a function of two vector variables $\mathbf{x}$ and $\mathbf{v}$ given (in this simple situation) by

$$H(\mathbf{x},\mathbf{v}) = \frac{1}{2}m\left\|\mathbf{v}\right\|^2 + V(\mathbf{x})$$

so that if we plug in $\mathbf{x} = \mathbf{x}(t)$ and $\mathbf{v} = \mathbf{x}'(t)$ for the motion of a particle then we get the energy of the particle. Since this is a conservative situation $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$ we discover by differentiating and using equation **??** that $\frac{d}{dt}H(\mathbf{x}(t),\mathbf{x}'(t)) = 0$. This says that the total energy is conserved along any path which is a solution to equation **??** as long as $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$.

There is a lot of structure that can be discovered by translating the equations of motion into an arbitrary coordinate system (q^1, q^2, q^3) and then extending that to a coordinate system $(q^1, q^2, q^3, \dot{q}^1, \dot{q}^2, \dot{q}^3)$ for velocity space $\mathbb{R}^3 \times \mathbb{R}^3$. Here, $\dot{q}^1, \dot{q}^2, \dot{q}^3$ are not derivatives until we compose with a curve $\mathbb{R} \to \mathbb{R}^3$ to get functions of t. Then (and only then) we will take $\dot{q}^1(t), \dot{q}^2(t), \dot{q}^3(t)$ to be the derivatives. Sometimes $(\dot{q}^1(t), \dot{q}^2(t), \dot{q}^3(t))$ is called the **generalized velocity** vector. Its physical meaning depends on the particular form of the generalized coordinates.

In such a coordinate system we have a function $L(\mathbf{q}, \dot{\mathbf{q}})$ called the Lagrangian of the system. Now there is a variational principle that states that if $\mathbf{q}(t)$ is a path which solve the equations of motion and defined from time t_1 to time t_2 then out of all the paths which connect the same points in space at the same times t_1 and t_2 , the one that makes the following action the smallest will be the solution:

$$S(\mathbf{q}(t)) = \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}})dt$$

Now this means that if we add a small variation to $\mathbf{q}$ get another path $\mathbf{q} + \delta\mathbf{q}$ then we calculate formally (and rigorously later in the notes):

$$\begin{aligned}
\delta S(\mathbf{q}(t)) &= \delta \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}})dt \\
&\int_{t_1}^{t_2} \left[\delta\mathbf{q}\cdot\frac{\partial}{\partial\mathbf{q}}L(\mathbf{q}, \dot{\mathbf{q}}) + \delta\dot{\mathbf{q}}\cdot\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right] dt \\
&= \int_{t_1}^{t_2} \delta\mathbf{q}\cdot\left(\frac{\partial}{\partial\mathbf{q}}L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right) dt + \left[\delta\mathbf{q}\cdot\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right]_{t_1}^{t_2}
\end{aligned}$$

If our variation is among those that start and end at the same space-time locations then $\delta\mathbf{q} = \mathbf{0}$ is the end points so the last term vanishes. Now if the path $\mathbf{q}(t)$ is stationary for such variations then $\delta S(\mathbf{q}(t)) = 0$ so

$$\int_{t_1}^{t_2} \delta\mathbf{q}\cdot\left(\frac{\partial}{\partial\mathbf{q}}L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right) dt = 0$$

and since this is true for all such paths we conclude that

$$\frac{\partial}{\partial \mathbf{q}}L(\mathbf{q},\dot{\mathbf{q}})-\frac{d}{dt}\frac{\partial}{\partial \dot{\mathbf{q}}}L(\mathbf{q},\dot{\mathbf{q}})=\mathbf{0}$$

or in indexed scalar form

$$\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}=0 \text{ for } 1\le i\le 3$$

on a stationary path. This is (these are) the **Euler-Lagrange equation**(s). If $\mathbf{q}$ were just rectangular coordinates and if L were $\frac{1}{2}m\left\|\mathbf{v}\right\|^2-V(\mathbf{x})$ this turns out to be Newton's equation. Notice, the minus sign in front of the V.

Definition 26.25 *For a Lagrangian L we can associate the quantity $E=\sum\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i-L(\mathbf{q},\dot{\mathbf{q}})$.*

Let us differentiate E. We get

$$\begin{aligned}\frac{d}{dt}E&=\frac{d}{dt}\sum\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i-L(\mathbf{q},\dot{\mathbf{q}})\\&=\frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i-\dot{q}^i\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}-\frac{d}{dt}L(\mathbf{q},\dot{\mathbf{q}})\\&=\frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i-\dot{q}^i\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}-\frac{\partial L}{\partial q^i}\dot{q}^i-\frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i\\&=0 \text{ by the Euler Lagrange equations.}\end{aligned}\tag{26.4}$$

Conclusion 26.1 *If L does not depend explicitly on time; $\frac{\partial L}{\partial t}=0$, then the **energy E is conserved ;** $\frac{dE}{dt}=0$ along any solution of the Euler-Lagrange equations..*

But what about spatial symmetries? Suppose that $\frac{\partial}{\partial q^i}L=0$ for one of the coordinates q^i. Then if we define $p_i=\frac{\partial L}{\partial \dot{q}^i}$ we have

$$\frac{d}{dt}p_i=\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}=-\frac{\partial}{\partial q^i}L=0$$

so p_i is constant along the trajectories of Euler's equations of motion. The quantity $p_i=\frac{\partial L}{\partial \dot{q}^i}$ is called a generalized momentum and we have reached the following

Conclusion 26.2 *If $\frac{\partial}{\partial q^i}L=0$ then p_i is a conserved quantity. This also applies if $\frac{\partial}{\partial \mathbf{q}}L=(\frac{\partial L}{\partial q^1},...,\frac{\partial L}{\partial q^n})=0$ with the conclusion that the vector $\mathbf{p}=\frac{\partial}{\partial \mathbf{q}}L=(\frac{\partial L}{\partial \dot{q}^1},...,\frac{\partial L}{\partial \dot{q}^n})$ is conserved (each component separately).*

Now let us apply this to the case a free particle. The Lagrangian in rectangular inertial coordinates are

$$L(\mathbf{x},\dot{\mathbf{x}})=\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2$$

and this Lagrangian is symmetric with respect to translations $\mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$

$$L(\mathbf{x} + \mathbf{c}, \dot{\mathbf{x}}) = L(\mathbf{x}, \dot{\mathbf{x}})$$

and so the generalized momentum vector for this is $\mathbf{p} = m\dot{\mathbf{x}}$ each component of which is conserved. This last quantity is actually the usual momentum vector.

Now let us examine the case where the Lagrangian is invariant with respect to rotations about some fixed point which we will take to be the origin of an inertial coordinate system. For instance suppose the potential function $V(\mathbf{x})$ is invariant in the sense that $V(\mathbf{x}) = V(O\mathbf{x})$ for any orthogonal matrix O. The we can take an antisymmetric matrix A and form the family of orthogonal matrices e^{sA}. The for the Lagrangian

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}m\,|\dot{\mathbf{x}}|^2 - V(\mathbf{x})$$

we have

$$\begin{aligned}\frac{d}{ds}L(e^{sA}\mathbf{x}, e^{sA}\dot{\mathbf{x}}) &= \frac{d}{dt}(\frac{1}{2}m\left|e^{sA}\dot{\mathbf{x}}\right|^2 - V(e^{sA}\mathbf{x})) \\ &= \frac{d}{dt}(\frac{1}{2}m\,|\dot{\mathbf{x}}|^2 - V(\mathbf{x})) = 0\end{aligned}$$

On the other hand, recall the result of a variation $\delta\mathbf{q}$

$$\int_{t_1}^{t_2} \delta\mathbf{q}\cdot\left(\frac{\partial}{\partial\mathbf{q}}L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right)dt + \left[\delta\mathbf{q}\cdot\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right]_{t_1}^{t_2}$$

what we have done is to let $\delta\mathbf{q} = A\mathbf{q}$ since to first order we have $e^{sA}\mathbf{q} = I + sA\mathbf{q}$. But if $\mathbf{q}(t)$ satisfies Euler's equation then the integral above is zero and yet the whole variation is zero too. We are led to conclude that

$$\left[\delta\mathbf{q}\cdot\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q}, \dot{\mathbf{q}})\right]_{t_1}^{t_2} = 0$$

which in the present case is

$$\begin{aligned}\left[A\mathbf{x}\cdot\frac{\partial}{\partial\dot{\mathbf{x}}}(\frac{1}{2}m\,|\dot{\mathbf{x}}|^2 - V(\mathbf{x}))\right]_{t_1}^{t_2} &= 0 \\ [mA\mathbf{x}\cdot\dot{\mathbf{x}}]_{t_1}^{t_2} &= 0\end{aligned}$$

for all t_2 and t_1. Thus the quantity $mA\mathbf{x}\cdot\dot{\mathbf{x}}$ is conserved. Let us apply this with A equal to the following in turn

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

then we get $mA\mathbf{x}\cdot\dot{\mathbf{x}} = m(-x^2, x^1, 0)\cdot(\dot{x}^1, \dot{x}^2, \dot{x}^3) = m(x^1\dot{x}^2 - \dot{x}^1x^2)$ which is the same as $m\dot{\mathbf{x}}\times\mathbf{k} = \mathbf{p}\times\mathbf{k}$ which is called the angular momentum about the $\mathbf{k}$

axis ($\mathbf{k} = (0,0,1)$ so this is the z-axis) and is a conserved quantity. To see the point here notice that

$$e^{tA} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the rotation about the z axis. We can do the same thing for the other two coordinate axes and in fact it turns out that for any unit vector $\mathbf{u}$ the angular momentum about that axis defined by $\mathbf{p} \times \mathbf{u}$ is conserved.

Remark 26.2 *We started with the assumption that L was invariant under all rotations O but if it had only been invariant under counterclockwise rotations about an axis given by a unit vector $\mathbf{u}$ then we could still conclude that at least $\mathbf{p} \times \mathbf{u}$ is conserved.*

Remark 26.3 *Let begin to use the index notation (like q^i, p_i and x^i etc.) a little more since it will make the transition to fields more natural.*

Now we define the Hamiltonian function derived from a given Lagrangian via the formulas

$$H(\mathbf{q},\mathbf{p}) = \sum p_i \dot{q}^i - L(\mathbf{q},\dot{\mathbf{q}})$$
$$p_i = \frac{\partial L}{\partial \dot{q}^i}$$

where we think of $\dot{\mathbf{q}}$ as depending on $\mathbf{q}$ and $\mathbf{p}$ via the inversion of $p_i = \frac{\partial L}{\partial \dot{q}^i}$. Now it turns out that if $\mathbf{q}(t), \dot{\mathbf{q}}(t)$ satisfy the Euler Lagrange equations for L then $\mathbf{q}(t)$ and $\mathbf{p}(t)$ satisfy the Hamiltonian equations of motion

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}$$
$$\frac{dp^i}{dt} = -\frac{\partial H}{\partial q^i}$$

One of the beauties of this formulation is that if $Q^i = Q^i(q^j)$ are any other coordinates on $\mathbb{R}^3$ and we define $P^i = p^j \frac{\partial Q^i}{\partial q^j}$ then taking $H(..q^i.,..p^i..) = \widetilde{H}(..Q^i..,..P_i..)$ the equations of motion have the same form in the new coordinates. More generally, if Q, P are related to q, p in such a way that the Jacobian matrix J of the coordinate change (on $\mathbb{R}^3 \times \mathbb{R}^3$) is **symplectic**

$$J^t \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

then the equations 26.4.3 will hold in the new coordinates. These kind of coordinate changes on the q, p space $\mathbb{R}^3 \times \mathbb{R}^3$ (momentum space) are called canonical transformations. Mechanics is, in the above sense, invariant under canonical transformations.

Next, take any smooth function $f(q,p)$ on momentum space (also called phase space). Such a function is called an observable. Then along any solution curve $(q(t),p(t))$ to Hamilton's equations we get

$$\begin{aligned}\frac{df}{dt} &= \frac{\partial f}{\partial q}\frac{dq}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} \\ &= \frac{\partial f}{\partial q^i}\frac{\partial H}{\partial p^i} + \frac{\partial f}{\partial p^i}\frac{\partial H}{\partial q^i} \\ &= [f,H]\end{aligned}$$

where we have introduced the Poisson bracket $[f,H]$ defined by the last equality above. So we also have the equations of motion in the form $\frac{df}{dt} = [f,H]$ for any function f not just the coordinate functions q and p. Later, we shall study a geometry hiding here; **Symplectic geometry**.

Remark 26.4 *For any coordinate $t, \mathbf{x}$ we will often consider the curve $(\mathbf{x}(t), \mathbf{x}'(t)) \in \mathbb{R}^{3n} \times \mathbb{R}^{3n}$ the latter product space being a simple example of a **velocity phase space**.*

26.4.4 Some simple mechanical systems

1. As every student of basic physics know the equations of motion for a particle falling freely through a region of space near the earths surface where the force of gravity is (nearly) constant is $\mathbf{x}''(t) = -g\mathbf{k}$ where $\mathbf{k}$ is the usual vertical unit vector corresponding to a vertical z-axis. Integrating twice gives the form of any solution $\mathbf{x}(t) = -\frac{1}{2}gt^2\mathbf{k}+t\mathbf{v}_0 + \mathbf{x}_0$ for constant vectors $\mathbf{x}_0, \mathbf{v}_0 \in \mathbb{R}^3$. We get different motions depending on the initial conditions $(\mathbf{x}_0, \mathbf{v}_0)$. If the initial conditions are right, for example if $\mathbf{v}_0 = 0$ then this is reduced to the one dimensional equation $x''(t) = -g$. The path of a solution with initial conditions (x_0, v_0) is given in phase space as
$$t \mapsto (-\frac{1}{2}gt^2 + tv_0 + x_0, -gt + v_0)$$
and we have shown the phase trajectories for a few initial conditions.

Phase portrait-falling object

2. A somewhat general 1-dimensional system is given by a Lagrangian of the form
$$L = \frac{1}{2}a(q)\dot{q}^2 - V(q) \tag{26.5}$$
and example of which is the motion of a particle of mass m along a 1-dimensional continuum and subject to a potential $V(x)$. Then the Lagrangian is $L = \frac{1}{2}m\dot{x}^2 - V(x)$. Instead of writing down the Euler-Lagrange equations we can use the fact that $E = \frac{\partial L}{\partial \dot{x}^i}\dot{x}^i - L(x,\dot{x}) = m\dot{x}^2 - (\frac{1}{2}m\dot{x}^2 - V(x)) = \frac{1}{2}m\dot{x}^2 + V(x)$ is conserved. This is the total energy which is traditionally divided into kinetic energy $\frac{1}{2}m\dot{x}^2$ and potential energy $V(x)$. We have $E = \frac{1}{2}m\dot{x}^2 + V(x)$ for some constant. Then
$$\frac{dx}{dt} = \sqrt{\frac{2E - 2V(x)}{m}}$$
and so
$$t = \sqrt{m/2}\int \frac{1}{\sqrt{E - V(x)}} + c.$$
Notice that we must always have $E - V(x) \geq 0$. This means that if $V(x)$ has a single local minimum between some points $x = a$ and $x = b$ where $E - V = 0$, then the particle must stay between $x = a$ and $x = b$ moving back and forth with some time period. What is the time period?.

3. **Central Field.** A central field is typically given by a potential of the form $V(\mathbf{x}) = -\frac{k}{|\mathbf{x}|}$. Thus the Lagrangian of a particle of mass m in this central field is
$$\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2 + \frac{k}{|\mathbf{x}|}$$
where we have centered inertial coordinates at the point where the potential has a singularity $\lim_{\mathbf{x}\to 0} V(\mathbf{x}) = \pm\infty$. In cylindrical coordinates (r,θ,z) the Lagrangian becomes
$$\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2) + \frac{k}{(r^2 + z^2)^{1/2}}.$$
We are taking $q^1 = r$, $q^2 = \theta$ and $q^3 = \theta$. But if initially $z = \dot{z} = 0$ then by conservation of angular momentum discussed above the particle stays in the $z = 0$ plane. Thus we are reduced to a study of the two dimensional case:
$$\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + \frac{k}{r}.$$
What are Lagrange's equations? Answer:
$$\begin{aligned} 0 &= \frac{\partial L}{\partial q^1} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^1} \\ &= mr\dot{\theta}^2 - \frac{k}{r^2} - m\dot{r}\ddot{r} \end{aligned}$$

and

$$\begin{aligned} 0 &= \frac{\partial L}{\partial q^2} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^2} \\ &= -mr^2\dot{\theta}\ddot{\theta}. \end{aligned}$$

The last equation reaffirms that $\dot{\theta} = \omega_0$ is constant. Then the first equation becomes $mr\omega_0^2 - \frac{k}{r^2} - m\dot{r}\ddot{r} = 0$. On the other hand conservation of energy becomes

4.

$$\begin{aligned} \frac{1}{2}m(\dot{r}^2 + r^2\omega_0^2) + \frac{k}{r} = E_0 = \frac{1}{2}m(\dot{r}_0^2 + r_0^2\omega_0^2) + \frac{k}{r_0} \qquad \text{or} \\ \dot{r}^2 + r^2\omega_0^2 + \frac{2k}{mr} = \frac{2E_0}{m} \end{aligned}$$

5. A simple oscillating system is given by $\frac{d^2x}{dt^2} = -x$ which has solutions of the form $x(t) = C_1 \cos t + C_2 \sin t$. This is equivalent to the system

$$\begin{aligned} x' &= v \\ v' &= -x \end{aligned}$$

6. Consider a single particle of mass m which for some reason is viewed with respect to rotating frame and an inertial frame (taken to be stationary). The rotating frame $(\mathbf{E}_1(t), \mathbf{E}_2(t), \mathbf{E}_3(t)) = \mathtt{E}$ (centered at the origin of R^3) is related to stationary frame $(e_1, e_2, e_3) = \mathtt{e}$ by an orthogonal matrix O :

$$\mathtt{E}(t) = \mathrm{O}(t)\mathtt{e}$$

and the rectangular coordinates relative to these frames are related by

$$\mathbf{x}(t) = \mathrm{O}(t)\mathbf{X}(t)$$

We then have

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathrm{O}(t)\dot{\mathbf{X}} + \dot{\mathrm{O}}(t)\mathbf{X} \\ &= \mathrm{O}(t)(\dot{\mathbf{X}} + \Omega(t)\mathbf{X})\end{aligned}$$

where $\Omega(t) = \mathrm{O}^t(t)\dot{\mathrm{O}}(t)$ is an angular velocity. The reason we have chosen to work with $\Omega(t)$ rather than directly with $\dot{\mathrm{O}}(t)$ will become clearer later in the book. Let us define the operator D_t by $D_t\mathbf{X}{=}\dot{\mathbf{X}}{+}\Omega(t)\mathbf{X}$. This is sometimes called the "total derivative". At any rate the equations of motion in the inertial frame is of the form $m\frac{d\mathbf{x}}{dt} = \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x})$. In the moving frame this becomes an equation of the form

$$m\frac{d}{dt}(\mathrm{O}(t)(\dot{\mathbf{X}} + \Omega(t)\mathbf{X})) = \mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)(\dot{\mathbf{X}} + \Omega(t)\mathbf{X}))$$

and in turn

$$\mathrm{O}(t)\frac{d}{dt}(\dot{\mathbf{X}}+\Omega(t)\mathbf{X})+\dot{\mathrm{O}}(t)(\dot{\mathbf{X}}+\Omega(t)\mathbf{X}) = m^{-1}\mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)(\dot{\mathbf{X}}+\Omega(t)\mathbf{X})).$$

Now recall the definition of D_t we get

$$\mathrm{O}(t)(\frac{d}{dt}D_t\mathbf{X} + \Omega(t)D_t\mathbf{X}) = m^{-1}\mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)\mathbf{V})$$

and finally

$$mD_t^2\mathbf{X} = \mathbf{F}(\mathbf{X}, \mathbf{V}) \tag{26.6}$$

where we have defined the **relative velocity** $\mathbf{V} = \dot{\mathbf{X}}{+}\Omega(t)\mathbf{X}$ and $\mathbf{F}(\mathbf{X}, \mathbf{V})$ is by definition the transformed force $\mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)\mathbf{V})$. The equation we have derived would look the same in any moving frame: It is a covariant expression.

5. Rigid Body We will use this example to demonstrate how to work with the rotation group and it's Lie algebra. The advantage of this approach is that it generalizes to motions in other Lie groups and their algebra's. Let us denote the group of orthogonal matrices of determinant one by $SO(3)$. This is the rotation group. If the Lagrangian of a particle as in the last example is invariant under actions of the orthogonal group so that $L(\mathbf{x}, \dot{\mathbf{x}}) = L(Qx, Q\dot{x})$ for $Q \in SO(3)$ then the quantity $\ell = \mathbf{x} \times m\dot{\mathbf{x}}$ is constant for the motion of the particle $\mathbf{x} = \mathbf{x}(t)$ satisfying the equations of motion in the inertial frame. The matrix group $SO(3)$ is an example of a Lie group which we study intensively in later chapters. Associated with every Lie group is its Lie algebra which in this case is the set of all anti-symmetric 3x3 matrices denoted $\mathfrak{so}(3)$. There is an interesting correspondence between and $\mathbb{R}^3$ given by

$$\begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \leftrightarrows (\omega_1, \omega_2, \omega_3) = \omega$$

Furthermore if we define the bracket for matrices A and B in $\mathfrak{so}(3)$ by $[A, B] = AB - BA$ then under the above correspondence $[A, B]$ corresponds to the cross product. Let us make the temporary convention that if x is an element of $\mathbb{R}^3$ then the corresponding matrix in $\mathfrak{so}(3)$ will be denoted by using the same letter but a new font while lower case refers to the inertial frame and upper to the moving frame:

$$\mathbf{x} \leftrightarrows \mathsf{x} \in \mathfrak{so}(3) \text{ and}$$
$$\mathbf{X} \leftrightarrows \mathsf{X} \in \mathfrak{so}(3) \text{ etc.}$$

	$\mathbb{R}^3$		$\mathfrak{so}(3)$
Inertial frame	$\mathbf{x}$	$\leftrightarrows$	x
Moving frame	$\mathbf{X}$	$\leftrightarrows$	X

Then we have the following chart showing how various operations match up:

$\mathbf{x}=\mathrm{O}\mathbf{X}$	$\leftrightarrows$	$\mathsf{x}= \mathsf{O}\mathsf{X}\mathsf{O}^t$
$\mathbf{v}_1 \times \mathbf{v}_2$	$\leftrightarrows$	$[\mathsf{v}_1, \mathsf{v}_2]$
$\mathbf{v} = \dot{\mathbf{x}}$	$\leftrightarrows$	$\mathsf{v}=\dot{\mathsf{x}}$
$\mathbf{V} =D_t\mathbf{X}=\dot{\mathbf{X}}+\Omega(t)\mathbf{X}$	$\leftrightarrows$	$\mathsf{V}=D_t\mathsf{X} = \dot{\mathsf{X}}+[\Omega(t), \mathsf{X}]$
$\ell = \mathbf{x} \times m\dot{\mathbf{x}}$	$\leftrightarrows$	$\mathsf{l} = [\mathsf{x}, m\dot{\mathsf{x}}]$
$\ell = \mathrm{O}\mathbf{L}$	$\leftrightarrows$	$\mathsf{l} = \mathsf{O}\mathsf{L}\mathsf{O}^t = [\mathsf{V},\Omega(t)]$
$D_t\mathbf{L} = \dot{\mathbf{L}}+\Omega(t) \times \mathbf{L}$	$\leftrightarrows$	$D_t\mathsf{L} = \dot{\mathsf{L}}+[\Omega(t), \mathsf{L}]$

and so on. Some of the quantities are actually defined by their position in this chart. In any case, let us differentiate $\mathbf{l} = \mathbf{x} \times m\dot{\mathbf{x}}$ and use the equations of motion to get

$$\begin{aligned} \frac{d\mathbf{l}}{dt} &= \mathbf{x} \times m\dot{\mathbf{x}} \\ &= \mathbf{x} \times m\ddot{\mathbf{x}} + \mathbf{0} \\ &= \mathbf{x} \times \mathbf{f}. \end{aligned}$$

But we have seen that if the Lagrangian (and hence the force $\mathbf{f}$) is invariant under rotations that $\frac{d\mathbf{l}}{dt} = 0$ along any solution curve. Let us examine this case. We have $\frac{d\mathbf{l}}{dt} = 0$ and in the moving frame $D_t\mathbf{L}=\dot{\mathbf{L}}+\Omega(t)\mathbf{L}$. Transferring the equations over to our $\mathfrak{so}(3)$ representation we have $D_t\mathsf{L}=\dot{\mathsf{L}}+[\Omega(t), \mathsf{L}] = 0$. Now if our particle is rigidly attached to the rotating frame, that is, if $\dot{\mathbf{x}} = 0$ then $\dot{\mathsf{X}} = 0$ and $\mathsf{V} =[\Omega(t), \mathsf{X}]$ so

$$\mathsf{L} = m[\mathsf{X}, [\Omega(t), \mathsf{X}]].$$

In Lie algebra theory the map $\mathsf{v} \mapsto [\mathsf{x}, \mathsf{v}] = -[\mathsf{v}, \mathsf{x}]$ is denoted $\mathrm{ad}(\mathsf{x})$ and is linear. With this notation the above becomes

$$\mathsf{L} = -m\,\mathrm{ad}(\mathsf{X})\Omega(t).$$

The map $I : \mathsf{X} \mapsto -m\,\mathrm{ad}(\mathsf{X})\Omega(t) = I(\mathsf{X})$ is called the momentum operator. Suppose now that we have k particles of masses $m_1, m_2, ...m_2$ each at rigidly attached to the rotating frame and each giving quantities $\mathsf{x}_i, \mathsf{X}_i$ etc. Then to total angular momentum is $\sum I(\mathsf{X}_i)$. Now if we have a continuum of mass with mass density ρ in a moving region B_t (a rigid body) then letting $\mathsf{X}_{\mathbf{u}}(t)$ denote path in $\mathfrak{so}(3)$ of the point of initially at $\mathbf{u} \in B_0 \in \mathbb{R}^3$ then we can integrate to get the total angular momentum at time t;

$$\mathsf{L}_{tot}(t) = -\int_B \mathrm{ad}(\mathsf{X}_{\mathbf{u}}(t))\Omega(t)d\rho(\mathbf{u})$$

which is a conserved quantity.

26.4.5 The Basic Ideas of Relativity

We will draw an analogy with the geometry of the Euclidean plane. Recall that the abstract plane $\mathcal{P}$ is *not* the same thing as coordinate space $\mathbb{R}^2$ but rather there are many "good" bijections $\Psi : \mathcal{P} \to \mathbb{R}^2$ called coordinatizations such that points $p \in \mathcal{P}$ corresponding under Ψ to coordinates $(x(p), y(p))$ are temporarily identified with the pair $(x(p), y(p))$ is such a way that the distance between points is given by $dist(p,q) = \sqrt{(x(p) - x(q))^2 + (y(p) - y(q))^2}$ or

$$d^2 = \Delta x^2 + \Delta y^2$$

for short. Now the numbers Δx and Δy separately have no absolute meaning since a different good-coordinatization $\Phi : \mathcal{P} \to \mathbb{R}^2$ would give something like $(X(p), Y(p))$ and then for the same two points p, q we expect that in general $\Delta x \neq \Delta X$ and $\Delta y \neq \Delta Y$. On the other hand, the notion of distance is a geometric reality that should be independent of coordinate choices and so we always have $\Delta x^2 + \Delta y^2 = \Delta X^2 + \Delta Y^2$. But what is a "good coordinatization"? Well, one thing can be said for sure and that is if x, y are good then X, Y will be good also iff

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \cos\theta & \pm\sin\theta \\ \mp\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

for some θ and numbers T_1, T_2. The set of all such transformations form a group under composition is called the Euclidean motion group. Now the idea of points on an abstract plane is easy to imagine but it is really just a "set" of objects with some logical relations; an idealization of certain aspects of our experience. Similarly, we now encourage the reader to make the following idealization. Imagine the set of all ideal local events or possible events as a sort of 4-dimensional plane. Imagine that when we speak of an event happening at location (x, y, z) in rectangular coordinates and at time t it is only because we have imposed some sort of coordinate system on the set of events that is implicit in our norms regarding measuring procedures etc. What if some other system were used to describe the same set of events, say, two explosions $e1$ and

$e2$. You would not be surprised to find out that the spatial separations for the two events

$$\Delta X, \Delta Y, \Delta Z$$

would not be absolute and would not individually equal the numbers

$$\Delta y, \Delta y, \Delta y.$$

But how about ΔT and Δt. Is the time separation, in fixed units of seconds say, a real thing?

The answer is actually NO according to the special theory of relativity. In fact, not even the quantities $\Delta X^2 + \Delta Y^2 + \Delta Z^2$ will agree with $\Delta x^2 + \Delta y^2 + \Delta y^2$ under certain circumstances! Namely, if two observers are moving relative to each other at constant speed, there will be objectively unresolvable disagreements. Who is right? There simply is no fact of the matter. The objective or absolute quantity is rather

$$\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta y^2$$

which always equals $\Delta T^2 - \Delta X^2 - \Delta Y^2 - \Delta Y^2$ for good coordinates systems. But what is a good coordinate system? It is one in which the equations of physics take on their simplest form. Find one, and then all others are related by the Poincaré group of linear transformations given by

$$\begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ t_0 \end{pmatrix}$$

where the matrix A is a member of the Lorentz group. The Lorentz group is characterized as that set $O(1,3)$ of matrices A such that

$$A^T \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} := \Lambda.$$

This is exactly what makes the following true.

Fact If $(t, \tilde{\mathbf{x}})$ and $(T, \tilde{\mathbf{X}})$ are related by $(t, \tilde{\mathbf{x}})^t = A(T, \tilde{\mathbf{X}})^t + (t_0, \tilde{\mathbf{x}}_0)^t$ for $A \in O(1,3)$ then $t^2 - |\tilde{\mathbf{x}}|^2 = T^2 - \left|\tilde{\mathbf{X}}\right|^2$.

A vector quantity described relative to an inertial coordinates $(t, \vec{x})$ by a 4-tuple $\mathsf{v} = (v^0, v^1, v^2, v^3)$ such that its description relative to $(T, \tilde{\mathbf{X}})$ as above is given by

$$\mathsf{V}^t = A\mathsf{v}^t \text{ (contravariant).}$$

Notice that we are using superscripts to index the components of a vector (and are not to be confused with exponents). This due to the following convention:

vectors written with components up are called contravariant vectors while those with indices down are called covariant. Contravariant and covariant vectors transform differently and in such a way that the contraction of a contravariant with a covariant vector produces a quantity that is the same in any inertial coordinate system. To change a contravariant vector to its associated covariant form one uses the matrix Λ introduced above which is called the Lorentz metric tensor. Thus $(v^0, v^1, v^2, v^3)\Lambda = (v^0, -v^1, -v^2, -v^3) := (v_0, v_1, v_2, v_3)$ and thus the pseudo-length $v_i v^i = (v^0)^2 - (v^1)^2 - (v^2)^2 - (v^3)^2$ is an invariant with respect to coordinate changes via the Lorentz group or even the Poincaré group. Notice also that $v_i v^i$ actually means $\sum_{i=0}^{3} v_i v^i$ which is in turn the same thing as

$$\sum \Lambda_{ij} v^i v^j$$

The so called **Einstein summation convention** say that when an index is repeated once up and once down as in $v_i v^i$, then the summation is implied.

Minkowski Space

One can see from the above that lurking in the background is an inner product space structure: If we fix the origin of space time then we have a vector space with inner product $\langle \mathsf{v}, \mathsf{v} \rangle = v_i v^i = (v^0)^2 - (v^1)^2 - (v^2)^2 - (v^3)^2$. This inner product space (indefinite inner product!) is called Minkowski space. The inner product just defined is the called the **Minkowski metric** or **the Lorentz metric.**.

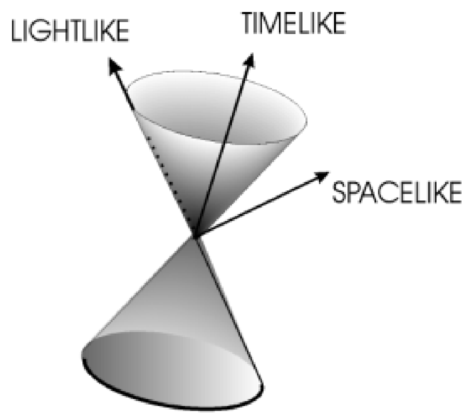
Definition 26.26 *A 4-vector* v *is called* ***space-like*** *iff* $\langle \mathsf{v}, \mathsf{v} \rangle < 0$, ***time-like*** *iff* $\langle \mathsf{v}, \mathsf{v} \rangle > 0$ *and* ***light-like*** *iff* $\langle \mathsf{v}, \mathsf{v} \rangle = 0$. *The set of all light-like vectors at a point in Minkowski space form a double cone in* $\mathbb{R}^4$ *referred to as the* ***light cone***.

Remark 26.5 (Warning) *Sometimes the definition of the Lorentz metric given is opposite in sign from the one we use here. Both choices of sign are popular. One consequence of the other choice is that time-like vectors become those for which* $\langle \mathsf{v}, \mathsf{v} \rangle < 0$.

Definition 26.27 *At each point of* $x \in \mathbb{R}^4$ *there is a set of vectors parallel to the 4-axes of* $\mathbb{R}^4$. *We will denote these by* $\partial_0, \partial_1, \partial_2$,*and* ∂_3 *(suppressing the point at which they are based).*

Definition 26.28 *A vector* v *based at a point in* $\mathbb{R}^4$ *such that* $\langle \partial_0, \mathsf{v} \rangle > 0$ *will be called* ***future pointing*** *and the set of all such forms the interior of the "future" light-cone.*

One example of a 4-vector is the momentum 4-vector written $\mathsf{p} = (E, \overrightarrow{\mathbf{p}})$ which we will define below. We describe the motion of a particle by referring to its career in space-time. This is called its world-line and if we write c for the speed of light and use coordinates $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ then a world line is a curve $\gamma(s) = (x^0(s), x^1(s), x^2(s), x^3(s))$ for some parameter. The momentum 4-vector is then $\mathsf{p} = mc\mathsf{u}$ where u is the unite vector in the direction



of the 4-velocity $\gamma'(s)$. The action functional for a free particle in Relativistic mechanics is invariant with respect to Lorentz transformations described above. In the case of a free particle of mass m it is

$$A_L = -\int_{s_1}^{s_2} mc\langle \dot{\gamma}(s), \dot{\gamma}(s)\rangle^{1/2} ds.$$

The quantity c is a constant equal to the speed of light in any inertial coordinate system. Here we see the need to assume that $\langle \dot{\gamma}(s), \dot{\gamma}(s)\rangle \geq 0$ ($\dot{\gamma}(s)$ is timelike) and then there is an obvious analogy with the length of a curve. Define

$$\tau(s) = \int_{s_1}^{s_2} \langle \dot{\gamma}(s), \dot{\gamma}(s)\rangle^{1/2} ds$$

Then $\frac{d}{ds}\tau(s) = cm\langle \dot{\gamma}(s), \dot{\gamma}(s)\rangle^{1/2} \geq 0$ so we can reparameterize by τ:

$$\gamma(\tau) = (x^0(\tau), x^1(\tau), x^2(\tau), x^3(\tau)).$$

The parameter τ is called the proper time of the particle in question. A stationary curve will be a geodesic or straight line in $\mathbb{R}^4$.

Let us return to the Lorentz group. The group of rotations generated by rotations around the spatial axes x,y,z are a copy of $SO(3)$ sitting inside $SO(1,3)$ and consists of matrices of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}$$

where $R \in SO(3)$. Now a rotation of the x,y plane about the z,t-plane[1] for example has the form

$$\begin{pmatrix} cT \\ X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

where as a Lorentz "rotation" of the t,x plane about the y,z-plane has the form

$$\begin{pmatrix} cT \\ X \\ Y \\ Z \end{pmatrix} = \left[\begin{array}{cccc} \cosh(\beta) & \sinh(\beta) & 0 & 0 \\ \sinh(\beta) & \cosh(\beta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \right]$$

$= [cT, X, Y, Z] = [(\cosh\beta)\, ct + (\sinh\beta)\, x, (\sinh\beta)\, ct + (\cosh\beta)\, x, y, z]$ Here the parameter β is usually taken to be the real number given by

$$\tanh\beta = v/c$$

where v is a velocity indicating that in the new coordinates the observer in travelling at a velocity of magnitude v in the x-direction as compared to an observer in the original before the transformation. Indeed we have for an observer motionless at the spatial origin the T, X, Y, Z system the observers path is given in T, X, Y, Z coordinates as $T \mapsto (T, 0, 0, 0)$

$$\begin{aligned} \frac{dX}{dt} &= c\sinh\beta \\ \frac{dX}{dx} &= \cosh\beta \\ v = \frac{dx}{dt} &= \frac{\sinh\beta}{\cosh\beta} = c\tanh\beta. \end{aligned}$$

Calculating similarly, and using $\frac{dY}{dt} = \frac{dZ}{dt} = 0$ we are lead to $\frac{dy}{dt} = \frac{dz}{dt} = 0$. So the t, x, y, z observer sees the other observer (and hence his frame) as moving in the x- direction at speed $v = c\tanh\beta$. The transformation above is called a **Lorentz boost** in the x-direction.

Now as a curve parameterized by the parameter τ (the proper time) the 4-momentum is the vector

$$\mathsf{p} = mc\frac{d}{d\tau}\mathsf{x}(\tau).$$

In a specific inertial (Lorentz) frame

$$\mathsf{p}(t) = \frac{dt}{d\tau}\frac{d}{dt}mc\mathsf{x}(t) = \left(\frac{mc^2}{c\sqrt{1-(v/c)^2}}, \frac{mc\dot{x}}{\sqrt{1-(v/c)^2}}, \frac{mc\dot{y}}{\sqrt{1-(v/c)^2}}, \frac{mc\dot{z}}{\sqrt{1-(v/c)^2}} \right)$$

[1] It is the z,t-plane rather than just the z-axis since we are in four dimensions and both the z-axis and the t-axis would remain fixed.

which we abbreviate to $\mathsf{p}(t) = (E/c, \overrightarrow{\mathbf{p}})$ where $\overrightarrow{\mathbf{p}}$ is the 3-vector given by the last there components above. Notice that

$$m^2c^2 = \langle \mathsf{p}(t), \mathsf{p}(t)\rangle = E^2/c^2 + \left|\overrightarrow{\mathbf{p}}\right|^2$$

is an invariant quantity but the pieces E^2/c^2 and $\left|\overrightarrow{\mathbf{p}}\right|^2$ are dependent on the choice of inertial frame.

What is the energy of a moving particle (or tiny observer?) in this theory? We claim it is the quantity E just introduced. Well, if the Lagrangian is any guide we should have from the point of view of inertial coordinates and for a particle moving at speed v in the positive $x-$direction

$$\begin{aligned} E &= v\frac{\partial L}{\partial v} = v\frac{\partial}{\partial v}\left(-mc^2\sqrt{1-(v/c)^2}\right) - (-mc^2\sqrt{1-(v/c)^2}) \\ &= m\frac{c^3}{\sqrt{(c^2-v^2)}} = m\frac{c^2}{\sqrt{(1-(v/c)^2)}}. \end{aligned}$$

Expanding in powers of the dimensionless quantity v/c we have $E = mc^2 + \frac{1}{2}mv^2 + O\left((v/c)^4\right)$. Now the term $\frac{1}{2}mv^2$ is just the nonrelativistic expression for kinetic energy. What about the mc^2? If we take the Lagrangian approach seriously, this must be included as some sort of energy. Now if v had been zero then we would still have a "rest energy" of mc^2! This is interpreted as the energy possessed by the particle by virtue of its mass. A sort of energy of being as it were. Thus we have the famous equation for the equivalence of mass and energy (we have $v = 0$ here):

$$E = mc^2.$$

If the particle is moving then mc^2 is only part of the energy but we can define $E_0 = mc^2$ as the "rest energy". Notice however, that although the length of the momentum 4-vector $m\dot{\mathsf{x}}$=p is always m;

$$\langle \mathsf{p}, \mathsf{p}\rangle^{1/2} = m\{(dt/d\tau)^2 - (dx/d\tau)^2 - (dy/d\tau)^2 - (dz/d\tau)^2\}^{1/2} = m$$

and is therefore conserved in the sense of being constant one must be sure to remember that the mass of a body consisting of many particles is *not* the sum of the individual particle masses.

26.4.6 Variational Analysis of Classical Field Theory

In field theory we study functions $\phi : \mathbb{R}\times\mathbb{R}^3 \to \mathbb{R}^k$. We use variables $\phi(x^0, x^1, x^2, x^3) = \phi(t, x, y, z)$ A Lagrangian density is a function $\mathcal{L}(\phi, \partial\phi)$ and then the Lagrangian would be

$$L(\phi, \partial\phi) = \int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi, \partial\phi)d^3x$$

and the action is

$$S = \int\int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi, \partial\phi)d^3x dt = \int_{V\times I\subset\mathbb{R}^4} \mathcal{L}(\phi, \partial\phi)d^4x$$

What has happened is that the index i is replaced by the space variable $\vec{x} = (x^1, x^2, x^3)$ and we have the following translation

$$\begin{array}{lll}
i & \rightarrowtail\rightarrowtail\rightarrowtail & \vec{x} \\
q & \rightarrowtail\rightarrowtail\rightarrowtail & \phi \\
q^i & \rightarrowtail\rightarrowtail\rightarrowtail & \phi(., \vec{x}) \\
q^i(t) & \rightarrowtail\rightarrowtail\rightarrowtail & \phi(t, \vec{x}) = \phi(x) \\
p^i(t) & \rightarrowtail\rightarrowtail\rightarrowtail & \partial_t \phi(t, \vec{x}) + \nabla_{\vec{x}} \phi(t, \vec{x}) = \partial\phi(x) \\
L(q, p) & \rightarrowtail\rightarrowtail\rightarrowtail & \int_{V \subset \mathbb{R}^3} \mathcal{L}(\phi, \partial\phi) d^3x \\
S = \int L(\mathbf{q}, \dot{\mathbf{q}}) dt & \rightarrowtail\rightarrowtail\rightarrowtail & S = \int \int \mathcal{L}(\phi, \partial\phi) d^3x dt
\end{array}$$

where $\partial\phi = (\partial_0\phi, \partial_1\phi, \partial_2\phi, \partial_3\phi)$. So in a way, the mechanics of classical massive particles is classical field theory on the space with three points which is the set $\{1, 2, 3\}$. Or we can view field theory as infinitely many particle systems indexed by points of space-i.e. infinite degrees of freedom.

Actually, we have only set up the formalism of scalar fields and have not, for instance, set things up to cover internal degrees of freedom like spin. However, we will discuss spin later in this text. Let us look at the formal variational calculus of field theory. We let $\delta\phi$ be a variation which we might later assume to vanish on the boundary of some region in space-time $U = I \times V \subset \mathbb{R} \times \mathbb{R}^3 = \mathbb{R}^4$. In general, we have

$$\begin{aligned}
\delta S &= \int_U \left(\delta\phi \frac{\partial \mathcal{L}}{\partial \phi} + \partial_\mu \delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x \\
&= \int_U \partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x + \int_U \delta\phi \left(\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x
\end{aligned}$$

Now the first term would vanish by the divergence theorem if $\delta\phi$ vanished on the boundary ∂U. If ϕ were a field that were stationary under such variations then

$$\delta S = \int_U \delta\phi \left(\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x = 0$$

for all $\delta\phi$ vanishing on ∂U so we can conclude that Lagrange's equation holds for ϕ stationary in this sense and visa versa:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0$$

These are the field equations.

26.4.7 Symmetry and Noether's theorem for field theory

Now an interesting thing happens if the Lagrangian density is invariant under some set of transformations. Suppose that $\delta\phi$ is an infinitesimal "internal" symmetry of the Lagrangian density so that $\delta S(\delta\phi) = 0$ even though $\delta\phi$ does not vanish on the boundary. Then if ϕ is already a solution of the field equations then

$$0 = \delta S = \int_U \partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x$$

for all regions U. This means that $\partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0$ so if we define $j^\mu = \delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}$ we get

$$\partial_\mu j^\mu = 0$$

or

$$\frac{\partial}{\partial t} j^0 = -\nabla \cdot \overrightarrow{\mathbf{j}}$$

where $\overrightarrow{\mathbf{j}}=(j^1, j^2, j^3)$ and $\nabla \cdot \overrightarrow{\mathbf{j}} = \operatorname{div}(\overrightarrow{\mathbf{j}})$ is the spatial divergence. This looks like some sort of conservation.. Indeed, if we define the total charge at any time t by

$$Q(t) = \int j^0 d^3x$$

the assuming $\overrightarrow{\mathbf{j}}$ shrinks to zero at infinity then the divergence theorem gives

$$\begin{aligned} \frac{d}{dt} Q(t) &= \int \frac{\partial}{\partial t} j^0 d^3x \\ &= -\int \nabla \cdot \overrightarrow{\mathbf{j}} d^3x = 0 \end{aligned}$$

so the charge $Q(t)$ is a conserved quantity. Let $Q(U,t)$ denote the total charge inside a region U. The charge inside any region U can only change via a flux through the boundary:

$$\begin{aligned} \frac{d}{dt} Q(U,t) &= \int_U \frac{\partial}{\partial t} j^0 d^3x \\ &= \int_{\partial U} \overrightarrow{\mathbf{j}} \cdot \mathbf{n} dS \end{aligned}$$

which is a kind of "local conservation law". To be honest the above discussion only takes into account so called internal symmetries. An example of an internal symmetry is given by considering a curve of linear transformations of $\mathbb{R}^k$ given as matrices $C(s)$ with $C(0) = I$. Then we vary ϕ by $C(s)\phi$ so that $\delta\phi = \frac{d}{ds}\big|_0 C(s)\phi = C'(0)\phi$. Another possibility is to vary the underlying space so that $C(s,.)$ is now a curve of transformations of $\mathbb{R}^4$ so that if $\phi_s(x) = \phi(C(s,x))$ is a variation of fields then we must take into account the fact that the domain of integration is also varying:

$$L(\phi_s, \partial\phi_s) = \int_{U_s \subset \mathbb{R}^4} \mathcal{L}(\phi_s, \partial\phi_s) d^4x$$

We will make sense of this later.

26.4.8 Electricity and Magnetism

Up until now it has been mysterious how any object of matter could influence any other. It turns out that most of the forces we experience as middle sized objects pushing and pulling on each other is due to a single electromagnetic force. Without the help of special relativity there appears to be two forces; electric and magnetic. Elementary particles that carry electric charges such as electrons or protons, exert forces on each other by means of a field. In a particular Lorentz frame, the electromagnetic field is described by a skew-symmetric matrix of functions called the electromagnetic field tensor:

$$(F_{\mu\nu}) = \begin{bmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{bmatrix}.$$

Where we also have the forms $F_\mu^\nu = \Lambda^{s\nu} F_{\mu s}$ and $F^{\mu\nu} = \Lambda^{s\mu} F_s^\nu$. This tensor can be derived from a potential $\mathsf{A} = (A_0, A_1, A_2, A_3)$ by $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. The contravariant form of the potential is $(A_0, -A_1, -A_2, -A_3)$ is a four vector often written as

$$\mathsf{A} = (\phi, \overrightarrow{\mathbf{A}}).$$

The action for a charged particle in an electromagnetic field is written in terms of A in a manifestly invariant way as

$$\int_a^b -mcd\tau - \frac{e}{c} A_\mu dx^\mu$$

so writing $\mathsf{A} = (\phi, \overrightarrow{\mathbf{A}})$ we have

$$S = \int_a^b (-mc\frac{d\tau}{dt} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt})dt$$

so in a given frame the Lagrangian is

$$L(\tilde{\mathbf{x}}, \frac{d\tilde{\mathbf{x}}}{dt}, t) = -mc^2\sqrt{1-(v/c)^2} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt}.$$

Remark 26.6 *The system under study is that of a particle in a field and does not describe the dynamics of the field itself. For that we would need more terms in the Lagrangian.*

This is a time dependent Lagrangian because of the $\phi(t)$ term but it turns out that one can re-choose A so that the new $\phi(t)$ is zero and yet still have $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. This is called change of gauge. Unfortunately, if we wish to express things in such a way that a constant field is given by a constant potential then we cannot make this choice. In any case, we have

$$L(\overrightarrow{\mathbf{x}}, \frac{d\overrightarrow{\mathbf{x}}}{dt}, t) = -mc^2\sqrt{1-(v/c)^2} - e\phi + \overrightarrow{\mathbf{A}} \cdot \frac{d\overrightarrow{\mathbf{x}}}{dt}$$

and setting $\overrightarrow{\mathbf{v}} = \frac{d\tilde{\mathbf{x}}}{dt}$ and $\left|\overrightarrow{\mathbf{v}}\right| = v$ we get the follow form for energy

$$\overrightarrow{\mathbf{v}} \cdot \frac{\partial}{\partial \overrightarrow{\mathbf{v}}} L(\tilde{\mathbf{x}}, \overrightarrow{\mathbf{v}}, t) - L(\tilde{\mathbf{x}}, \overrightarrow{\mathbf{v}}, t) = \frac{mc^2}{\sqrt{1-(v/c)^2}} + e\phi.$$

Now this is not constant with respect to time because $\frac{\partial L}{\partial t}$ is not identically zero. On the other hand, this make sense from another point of view; the particle is interacting with the field and may be picking up energy from the field.

The Euler-Lagrange equations of motion turn out to be

$$\frac{d\tilde{\mathbf{p}}}{dt} = e\tilde{\mathbf{E}} + \frac{e}{c}\overrightarrow{\mathbf{v}} \times \tilde{\mathbf{B}}$$

where $\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \tilde{\mathbf{A}}}{\partial t} - \operatorname{grad}\phi$ and $\tilde{\mathbf{B}} = curl \tilde{\mathbf{A}}$ are the electric and magnetic parts of the field respectively. This decomposition into electric and magnetic parts is an artifact of the choice of inertial frame and may be different in a different frame. Now the momentum $\tilde{\mathbf{p}}$ is $\frac{m\overrightarrow{\mathbf{v}}}{\sqrt{1-(v/c)^2}}$ but a speeds $v << c$ this becomes nearly equal to $m\mathbf{v}$ so the equations of motion of a charged particle reduce to

$$m\frac{d\overrightarrow{\mathbf{v}}}{dt} = e\tilde{\mathbf{E}} + \frac{e}{c}\overrightarrow{\mathbf{v}} \times \tilde{\mathbf{B}}.$$

Notice that is the particle is not moving, or if it is moving parallel the magnetic field $\tilde{\mathbf{B}}$ then the second term on the right vanishes.

The electromagnetic field equations.

We have defined the 3-vectors $\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \tilde{\mathbf{A}}}{\partial t} - \operatorname{grad}\phi$ and $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ but since the curl of a gradient is zero it is easy to see that $\operatorname{curl}\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \tilde{\mathbf{B}}}{\partial t}$. Also, from $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ we get $\operatorname{div}\tilde{\mathbf{B}} = \mathbf{0}$. This easily derived pair of equations is the first two of the four famous Maxwell's equations. Later we will see that the electromagnetic field tensor is really a differential 2-form F and these two equations reduce to the statement that the (exterior) derivative of F is zero:

$$dF = 0$$

Exercise 26.1 *Apply Gauss's theorem and stokes theorem to the first two Maxwell's equations to get the integral forms. What do these equations say physically?*

One thing to notice is that these two equations do not determine $\frac{\partial}{\partial t}\tilde{\mathbf{E}}$.

Now we have not really written down a action or Lagrangian that includes terms that represent the field itself. When that part of the action is added in we get

$$S = \int_a^b (-mc - \frac{e}{c}A_\mu \frac{dx^\mu}{d\tau})d\tau + a\int_V F^{\nu\mu}F_{\nu\mu}dx^4$$

where in so called Gaussian system of units the constant a turns out to be $\frac{-1}{16\pi c}$. Now in a particular Lorentz frame and recalling **??** we get $= a\int_V F^{\nu\mu}F_{\nu\mu}dx^4 = \frac{1}{8\pi}\int_V \left|\tilde{\mathbf{E}}\right|^2 - \left|\tilde{\mathbf{B}}\right|^2 dtdxdydz$.

In order to get a better picture in mind let us now assume that there is a continuum of charged particle moving through space and that volume density of charge at any given moment in space-time is ρ so that if $dxdydz = dV$ then ρdV is the charge in the volume dV. Now we introduce the four vector $\rho \mathsf{u} = \rho(d\mathsf{x}/d\tau)$ where u is the velocity 4-vector of the charge at (t, x, y, z). Now recall that $\rho d\mathsf{x}/d\tau = \frac{d\tau}{dt}(\rho, \rho\overrightarrow{\mathbf{v}}) = \frac{d\tau}{dt}(\rho, \tilde{\mathbf{j}}) = \mathsf{j}$. Here $\tilde{\mathbf{j}} = \rho\overrightarrow{\mathbf{v}}$ is the charge current density as viewed in the given frame a vector field varying smoothly from point to point. Write $\mathsf{j} = (j^0, j^1, j^2, j^3)$.

Assuming now that the particle motion is determined and replacing the discrete charge e be the density we have applying the variational principle with the region $U = [a, b] \times V$ says

$$
\begin{aligned}
0 &= -\delta\left(\int_V \int_a^b \frac{\rho dV}{c} dV A_\mu \frac{dx^\mu}{d\tau} d\tau + a\int_U F^{\nu\mu}F_{\nu\mu}dx^4\right) \\
&= -\delta\left(\frac{1}{c}\int_U j^\mu A_\mu + aF^{\nu\mu}F_{\nu\mu}dx^4\right)
\end{aligned}
$$

Now the Euler-Lagrange equations become

$$
\frac{\partial\mathcal{L}}{\partial A_\nu} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\nu)} = 0
$$

where $\mathcal{L}(A_\mu, \partial_\mu A_\eta) = \frac{\rho}{c}A_\mu \frac{dx^\mu}{dt} + aF^{\nu\mu}F_{\nu\mu}$ and $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. If one is careful to remember that $\partial_\mu A_\nu = \frac{\partial A_\nu}{\partial x^\mu}$ is to be treated as an independent variable one cane arrive at some complicated looking equations and then looking at the matrix **??** we can convert the equations into statements about the fields $\tilde{\mathbf{E}}$, $\tilde{\mathbf{B}}$,and $(\rho, \tilde{\mathbf{j}})$. We will not carry this out since we later discover a much more efficient formalism for dealing with the electromagnetic field. Namely, we will use differential forms and the Hodge star operator. At any rate the last two of Maxwell's equations read

$$
\begin{aligned}
\operatorname{curl}\tilde{\mathbf{B}} &= 0 \\
\operatorname{div}\tilde{\mathbf{E}} &= 4\pi\rho.
\end{aligned}
$$

26.4.9 Quantum Mechanics

26.5 E. Calculus on Banach Spaces

> Mathematics is not only real, but it is the only reality. That is that the entire universe is made of matter is obvious. And matter is made of particles. It's made of electrons and neutrons and protons. So the entire universe is made out of particles. Now what are the particles made out of? They're not made out of anything. The only thing you can say about the reality of an electron is to cite its mathematical properties. So there's

> a sense in which matter has completely dissolved and what is left is just a mathematical structure.
>
> Gardner on Gardner: JPBM Communications Award Presentation. Focus-The Newsletter of the Mathematical Association of America v. 14, no. 6, December 1994.

26.6 Categories

We hope the reader has at least some familiarity with the notion of a category. We will not use category theory in any deep way except perhaps when discussing homological algebra. Nevertheless the language of categories is extremely convenient and so we will review a few of the salient features. It is not necessary for the reader to be completely at home with category theory before going further into the book and so he or she should not be discouraged if this section seems overly abstract. In particular, physics students may not be used to this kind of abstraction and should simply try to slowly get used to the language of categories.

Definition 26.29 *A* ***category*** $\mathfrak{C}$ *is a collection of objects* $Obj(\mathfrak{C}) = \{X, Y, Z, ...\}$ *and for every pair of objects* X, Y *a set* $\mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *called the set of* ***morphisms*** *from* X *to* Y*. The family of all such morphisms will be denoted* $\mathrm{Mor}(\mathfrak{C})$*. In addition a category is required to have a* ***composition*** *law which is a map* $\circ : \mathrm{Hom}_{\mathfrak{C}}(X, Y) \times \mathrm{Hom}_{\mathfrak{C}}(Y, Z) \to \mathrm{Hom}_{\mathfrak{C}}(X, Z)$ *such that for every three objects* $X, Y, Z \in Obj(\mathfrak{C})$ *the following axioms hold:*

Axiom 26.1 (Cat1) $\mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *and* $\mathrm{Hom}_{\mathfrak{C}}(Z, W)$ *are disjoint unless* $X = Z$ *and* $Y = W$ *in which case* $\mathrm{Hom}_{\mathfrak{C}}(X, Y) = \mathrm{Hom}_{\mathfrak{C}}(Z, W)$*.*

Axiom 26.2 (Cat2) *The composition law is associative:* $f \circ (g \circ h) = (f \circ g) \circ h$*.*

Axiom 26.3 (Cat3) *Each set of morphisms of the form* $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ *must contain an identity element* id_X *such that* $f \circ \mathrm{id}_X = f$ *for any* $f \in \mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *(and any* Y*), and* $\mathrm{id}_X \circ f = f$ *for any* $f \in \mathrm{Hom}_{\mathfrak{C}}(Y, X)$*.*

Notation 26.1 *A morphism is sometimes written using an arrow. For example, if* $f \in \mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *we would indicate this by writing* $f : X \to Y$ *or also by* $X \xrightarrow{f} Y$*. Also, the set* $\mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *of all morphisms from* X *to* Y *is also denoted by* $Mor_{\mathfrak{C}}(X, Y)$ *or abbreviated to* $\mathrm{Hom}(X, Y)$*.*

The notion of category is typified by the case where the objects are sets and the morphisms are maps between the sets. In fact, subject to putting restrictions on the sets and the maps, this will be almost the only type of category we shall need. On the other hand there are plenty of categories of this type:

1. **Grp**: The objects are groups and the morphisms are group homomorphisms.

2. **Rng** : The objects are rings and the morphisms are ring homomorphisms.

3. **Lin** : The objects are vector spaces and the morphisms are linear maps.

4. **Top**: The objects are topological spaces and the morphisms are continuous maps.

5. **Man^r: The category of C^r-differentiable manifolds** (and C^r-maps): One of the main categories discussed in this book. This is also called the smooth or differentiable category especially when $r = \infty$.

Notation 26.2 *If for some morphisms $f_i : X_i \to Y_i$, $(i = 1, 2)$, $g_X : X_1 \to X_2$ and $g_Y : Y_1 \to Y_2$ we have $g_Y \circ f_1 = f_2 \circ g_X$ then we express this by saying that the following diagram "commutes":*

$$\begin{array}{ccccc} & & f_1 & & \\ & X_1 & \to & Y_1 & \\ g_X & \downarrow & & \downarrow & g_Y \\ & X_2 & \to & Y_2 & \\ & & f_2 & & \end{array}$$

Similarly, if $h \circ f = g$ we say that the diagram

$$\begin{array}{cccc} & f & & \\ X & \to & Y & \\ & \searrow & \downarrow & h \\ g & & Z & \end{array}$$

commutes. More generally, tracing out a path of arrows in a diagram corresponds to composition of morphisms and to say that such a diagram commutes is to say that the compositions arising from two paths of arrows which begin and end at the same object are equal.

Definition 26.30 *Suppose that $f : X \to Y$ is a morphism from some category $\mathfrak{C}$. If f has the property that for any two (parallel) morphisms $g_1, g_2 : Z \to X$ we always have that $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$, i.e. if f is "left cancellable", then we call f a* ***$\mathfrak{C}$-monomorphism****. Similarly, if $f : X \to Y$ is "right cancellable" we call f a* ***$\mathfrak{C}$-epimorphism****. A morphism that is both a monomorphism and an epimorphism is called an* ***isomorphism*** *(or* ***$\mathfrak{C}$-isomorphism*** *if the category needs to be specified).*

In some cases we will use other terminology. For example, an isomorphism in the differentiable category is called a **diffeomorphism**. In the linear category, we speak of linear maps and linear isomorphisms. Morphisms from $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ are also called **endomorphisms** and so we also write $\mathrm{End}_{\mathfrak{C}}(X) := \mathrm{Hom}_{\mathfrak{C}}(X, X)$. The set of all $\mathfrak{C}$-isomorphisms in $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ is sometimes denoted by $\mathrm{Aut}_{\mathfrak{C}}(X)$ and are called **automorphisms**.

We single out the following: In many categories like the above we can form a sort of derived category that uses the notion of pointed space and pointed map. For example, in the topological category a pointed topological space is an topological space X together with a distinguished point p. Thus a typical object would be written (X, p). A morphism $f : (X, p) \to (W, q)$ would be a continuous map such that $f(p) = q$.

A **functor** $\digamma$ is a pair of maps both denoted by the same letter $\digamma$ which map objects and morphisms from one category to those of another

$$\digamma : Ob(\mathfrak{C}_1) \to Ob(\mathfrak{C}_2)$$
$$\digamma : Mor(\mathfrak{C}_1) \to Mor(\mathfrak{C}_2)$$

such that composition and identity morphisms are respected: If $f : X \to Y$ then

$$\digamma(f) : \digamma(X) \to \digamma(Y)$$

is a morphism in the second category and we must have

1. $\digamma(\mathrm{id}_{\mathfrak{C}_1}) = \mathrm{id}_{\mathfrak{C}_2}$
2. If $f : X \to Y$ and $g : Y \to Z$ then $\digamma(f) : \digamma(X) \to \digamma(Y)$, $\digamma(g) : \digamma(Y) \to \digamma(Z)$ and finally
$$\digamma(g \circ f) = \digamma(g) \circ \digamma(f).$$

This last property is perhaps the salient feature of a functor.

Example 26.7 *Let* $\mathbf{Lin}_{\mathbb{R}}$ *be the category whose objects are real vector spaces and whose morphisms are real linear maps. Similarly, let* $\mathbf{Lin}_{\mathbb{C}}$ *be the category of complex vector spaces with complex linear maps. To each real vector space* V *we can associate the complex vector space* $\mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$ *called the complexification of* V *and to each linear map of real vector spaces* $\ell : \mathrm{V} \to \mathrm{W}$ *we associate the complex extension* $\ell_{\mathbb{C}} : \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V} \to \mathbb{C} \otimes_{\mathbb{R}} \mathrm{W}$*. Here,* $\mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$ *is easily thought of as the vector space* V *where now complex scalars are allowed. Elements of* $\mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$ *are generated by elements of the form* $c \otimes v$ *where* $c \in \mathbb{C}$*,* $v \in \mathrm{V}$ *and we have* $i(c \otimes v) = ic \otimes v$ *where* $i = \sqrt{-1}$*. The map* $\ell_{\mathbb{C}} : \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V} \to \mathbb{C} \otimes_{\mathbb{R}} \mathrm{W}$ *is defined by the requirement* $\ell_{\mathbb{C}}(c \otimes v) = c \otimes \ell v$*. Now the assignments*

$$\ell \mapsto \ell_{\mathbb{C}}$$
$$\mathrm{V} \mapsto \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$$

define a functor from $\mathbf{Lin}_{\mathbb{R}}$ *to* $\mathbf{Lin}_{\mathbb{C}}$*.*

Remark 26.7 *In practice, complexification amounts to simply allowing complex scalars. For instance, we might just write cv instead of* $c \otimes v$*.*

Actually we have defined what is called a **covariant functor**. A **contravariant functor** is defined similarly except that the order of composition is reversed and instead of Funct2 above we would have $\digamma(g \circ f) = \digamma(f) \circ \digamma(g)$.

An example of a functor contravariant is the dual vector space functor which is a functor from the category of vector spaces $\mathbf{Lin}_{\mathbb{R}}$ to itself and sends each space to its dual and each linear map to its dual (adjoint):

$$\begin{array}{ccc} \mathrm{V} & \stackrel{L}{\rightarrow} & \mathrm{W} \\ \downarrow & & \downarrow \\ \mathrm{V}^* & \stackrel{L^*}{\leftarrow} & \mathrm{W}^* \end{array}$$

Notice the arrow reversal.

Remark 26.8 *One of the most important functors for our purposes is the* ***tangent functor*** *defined in section 3. Roughly speaking this functor replaces differentiable maps and spaces by their linear parts.*

Example 26.8 *Consider the category of real vector spaces and linear maps. To every vector space V we can associate the dual of the dual V^{**}. This is a* covariant *functor which is the composition of the dual functor with itself:*

$$\begin{array}{ccccc} V & & W^* & & V^{**} \\ A\downarrow & \mapsto & A^*\downarrow & \mapsto & A^{**}\downarrow \\ W & & V^* & & W^{**} \end{array}$$

Now suppose we have two functors

$$\mathcal{F}_1 : Ob(\mathfrak{C}_1) \to Ob(\mathfrak{C}_2)$$
$$\mathcal{F}_1 : Mor(\mathfrak{C}_1) \to Mor(\mathfrak{C}_2)$$

and

$$\mathcal{F}_2 : Ob(\mathfrak{C}_1) \to Ob(\mathfrak{C}_2)$$
$$\mathcal{F}_2 : Mor(\mathfrak{C}_1) \to Mor(\mathfrak{C}_2)$$

A **natural transformation** $\mathcal{T}$ from $\mathcal{F}_1$ to $\mathcal{F}_2$ is an assignment to each object X of $\mathfrak{C}_1$ a morphism $\mathcal{T}(X) : \mathcal{F}_1(X) \to \mathcal{F}_2(X)$ such that for every morphism $f : X \to Y$ of $\mathfrak{C}_1$ we have that the following diagram commutes:

$$\begin{array}{ccccc} & & \mathcal{T}(X) & & \\ & \mathcal{F}_1(X) & \rightarrow & \mathcal{F}_2(X) & \\ \mathcal{F}_1(f) & \downarrow & & \downarrow & \mathcal{F}_2(f) \\ & \mathcal{F}_1(Y) & \rightarrow & \mathcal{F}_2(Y) & \\ & & \mathcal{T}(Y) & & \end{array}$$

A common first example is the natural transformation ι between the identity functor $I : \mathbf{Lin}_{\mathbb{R}} \to \mathbf{Lin}_{\mathbb{R}}$ and the double dual functor $\iota : \mathbf{Lin}_{\mathbb{R}} \to \mathbf{Lin}_{\mathbb{R}}$:

$$\begin{array}{ccccc} & & \iota(\mathrm{V}) & & \\ & \mathrm{V} & \rightarrow & \mathrm{V}^{**} & \\ f & \downarrow & & \downarrow & f^{**} \\ & \mathrm{W} & \rightarrow & \mathrm{W}^{**} & \\ & & \iota(\mathrm{W}) & & \end{array} .$$

Here V^{**} is the dual space of the dual space V^*. The map $V \to V^{**}$ sends a vector to a linear function $\widetilde{v} : V^* \to \mathbb{R}$ defined by $\widetilde{v}(\alpha) := \alpha(v)$ (the hunter becomes the hunted so to speak). If there is an inverse natural transformation $\mathcal{T}^{-1}$ in the obvious sense, then we say that $\mathcal{T}$ is a natural isomorphism and for any object $X \in \mathfrak{C}_1$ we say that $\digamma_1(X)$ is naturally isomorphic to $\digamma_2(X)$. The natural transformation just defined is easily checked to have an inverse so is a natural isomorphism. The point here is not just that V is isomorphic to V^{**} in the category $\mathbf{Lin}_{\mathbb{R}}$ but that the isomorphism exhibited is natural. It works for all the spaces V in a uniform way that involves no special choices. This is to be contrasted with the fact that V is isomorphic to V^* but the construction of such an isomorphism involve an arbitrary choice of a basis.

26.7 Differentiability

For simplicity and definiteness all Banach spaces in this section will be real Banach spaces. First, the reader will recall that a linear map on a normed space, say $A : \mathsf{V}_1 \to \mathsf{V}_2$, is bounded if and only it is continuous at one and therefore any point in V_1. Given two Banach spaces V_1 and V_2 with norms $\|.\|_1$ and $\|.\|_2$ we can form a Banach space from the Cartesian product $\mathsf{V}_1 \times \mathsf{V}_2$ by using the norm $\|(\mathsf{v},\mathsf{u})\| := \max\{\|\mathsf{v}\|_1, \|\mathsf{u}\|_2\}$. There are many equivalent norms for $\mathsf{V}_1 \times \mathsf{V}_2$ including

$$\|(\mathsf{v},\mathsf{u})\|' := \sqrt{\|\mathsf{v}\|_1^2 + \|\mathsf{u}\|_2^2}$$
$$\|(\mathsf{v},\mathsf{u})\|'' := \|\mathsf{v}\|_1 + \|\mathsf{u}\|_2 .$$

Recall that two norms on V, say $\|.\|'$and $\|.\|''$ are equivalent if there exist positive constants c and C such that

$$c\,\|\mathsf{x}\|' \le \|\mathsf{x}\|'' \le C\,\|\mathsf{x}\|'$$

for all $\mathsf{x} \in \mathsf{V}$. Also, if V is a Banach space and W_1 and W_2 are closed subspaces such that $\mathsf{W}_1 \cap \mathsf{W}_2 = \{0\}$ and such that every $\mathsf{v} \in \mathsf{V}$ can be written uniquely in the form $\mathsf{v} = \mathsf{w}_1 + \mathsf{w}_2$ where $\mathsf{w}_1 \in \mathsf{W}_1$ and $\mathsf{w}_2 \in \mathsf{W}_2$ then we write $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2$. In this case there is the natural continuous linear isomorphism $\mathsf{W}_1 \times \mathsf{W}_2 \cong \mathsf{W}_1 \oplus \mathsf{W}_2$ given by

$$(\mathsf{w}_1, \mathsf{w}_2) \longleftrightarrow \mathsf{w}_1 + \mathsf{w}_2.$$

When it is convenient, we can identify $\mathsf{W}_1 \oplus \mathsf{W}_2$ with $\mathsf{W}_1 \times \mathsf{W}_2$. Under the representation $(\mathsf{w}_1, \mathsf{w}_2)$ we need to specify what norm we are using and there is more than one natural choice. We take $\|(\mathsf{w}_1, \mathsf{w}_2)\| := \max\{\|\mathsf{w}_1\|, \|\mathsf{w}_2\|\}$ but equivalent norms include, for example, $\|(\mathsf{w}_1, \mathsf{w}_2)\|_2 := \sqrt{\|\mathsf{w}_1\|^2 + \|\mathsf{w}_2\|^2}$ which is a good choice if the spaces happen to be Hilbert spaces.

Let E be a Banach space and $\mathsf{W} \subset \mathsf{E}$ a *closed* subspace. We say that W is **complemented** or a **split subspace** of E if there is a *closed* subspace W' such that $\mathsf{E} = \mathsf{W} \oplus \mathsf{W}'$.

Definition 26.31 (Notation) *We will denote the set of all continuous (bounded) linear maps from a Banach space* E *to a Banach space* F *by* $L(\mathsf{E}, \mathsf{F})$*. The set of all continuous linear isomorphisms from* E *onto* F *will be denoted by* $\mathrm{GL}(\mathsf{E}, \mathsf{F})$*. In case,* $\mathsf{E} = \mathsf{F}$ *the corresponding spaces will be denoted by* $\mathfrak{gl}(\mathsf{E})$ *and* $\mathrm{GL}(\mathsf{E})$*.*

Here $\mathrm{GL}(\mathsf{E})$ is a group under composition and is called the general linear group . In particular, $\mathrm{GL}(\mathsf{E}, \mathsf{F})$ is a subset of $L(\mathsf{E}, \mathsf{F})$ but not a linear subspace.

Definition 26.32 *Let* V_i*,* $i = 1, ..., k$ *and* W *be Banach spaces. A map* $\mu : V_1 \times \cdots \times V_k \to \mathsf{W}$ *is called* ***multilinear*** *(k-multilinear) if for each* i*,* $1 \leq i \leq k$ *and each fixed* $(\mathsf{w}_1, ..., \widehat{\mathsf{w}_i}, ..., \mathsf{w}_k) \in V_1 \times \cdots \times \widehat{V_1} \times \cdots \times V_k$ *we have that the map*

$$\mathsf{v} \mapsto \mu(\mathsf{w}_1, ..., \underset{i-th}{\mathsf{v}}, ..., \mathsf{w}_{k-1}),$$

obtained by fixing all but the i-th variable, is a bounded linear map. In other words, we require that μ *be* R*- linear in each slot separately.*

A multilinear map $\mu : \mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ is said to be **bounded** if and only if there is a constant C such that

$$\|\mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k)\|_{\mathsf{W}} \leq C \|\mathsf{v}_1\|_{\mathsf{E}_1} \|\mathsf{v}_2\|_{\mathsf{E}_2} \cdots \|\mathsf{v}_k\|_{\mathsf{E}_k}$$

for all $(\mathsf{v}_1, ..., \mathsf{v}_k) \in \mathsf{E}_1 \times \cdots \times \mathsf{E}_k$. The set of all bounded multilinear maps $\mathsf{E}_1 \times \cdots \times \mathsf{E}_k \to \mathsf{W}$ will be denoted by $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$. If $\mathsf{E}_1 = \cdots = \mathsf{E}_k = \mathsf{E}$ then we write $L^k(\mathsf{E}; \mathsf{W})$ instead of $L(\mathsf{E}, ..., \mathsf{E}; \mathsf{W})$

In case we are dealing with a be a Hilbert space $\mathsf{E}, \langle ., . \rangle$ then we have the group of linear isometries $O(\mathsf{E})$ from E onto itself. That is, $O(\mathsf{E})$ consists of the bijective linear maps $\Phi : \mathsf{E} \to \mathsf{E}$ such that $\langle \Phi \mathsf{v}, \Phi \mathsf{w} \rangle = \langle \mathsf{v}, \mathsf{w} \rangle$ for all $\mathsf{v}, \mathsf{w} \in \mathsf{E}$. The group $O(\mathsf{E})$ is called the **orthogonal group** (or sometimes the **Hilbert group** in the infinite dimensional case).

Definition 26.33 *A (bounded) multilinear map* $\mu : \mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ *is called* ***symmetric*** *(resp.* ***skew-symmetric*** *or* ***alternating****) iff for any* $\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k \in \mathsf{V}$ *we have that*

$$\mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k) = K(\mathsf{v}_{\sigma 1}, \mathsf{v}_{\sigma 2}, ..., \mathsf{v}_{\sigma k})$$
$$\textit{resp. } \mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k) = sgn(\sigma)\mu(\mathsf{v}_{\sigma 1}, \mathsf{v}_{\sigma 2}, ..., \mathsf{v}_{\sigma k})$$

for all permutations σ *on the letters* $\{1, 2, ..., k\}$*. The set of all bounded symmetric (resp. skew-symmetric) multilinear maps* $\mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ *is denoted* $L^k_{sym}(\mathsf{V}; \mathsf{W})$ *(resp.* $L^k_{skew}(\mathsf{V}; \mathsf{W})$ *or* $L^k_{alt}(\mathsf{V}; \mathsf{W})$*).*

Now the space $L(\mathsf{V}, \mathsf{W})$ is a Banach space in its own right with the norm

$$\|l\| = \sup_{\mathsf{v} \in \mathsf{V}} \frac{\|l(\mathsf{v})\|_{\mathsf{W}}}{\|\mathsf{v}\|_{\mathsf{V}}} = \sup\{\|l(\mathsf{v})\|_{\mathsf{W}} : \|\mathsf{v}\|_{\mathsf{V}} = 1\}.$$

The spaces $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ are also Banach spaces normed by

$$\|\mu\| := \sup\{\|\mu(\mathsf{v}_1, \mathsf{v}_2, ..., \mathsf{v}_k)\|_{\mathsf{W}} : \|\mathsf{v}_i\|_{\mathsf{E}_i} = 1 \text{ for } i = 1, .., k\}$$

Proposition 26.3 *A k-multilinear map $\mu \in L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ is continuous if and only if it is bounded.*

Proof. ($\Leftarrow$) We shall simplify by letting $k = 2$. Let $(\mathsf{a}_1, \mathsf{a}_2)$ and $(\mathsf{v}_1, \mathsf{v}_2)$ be elements of $\mathsf{E}_1 \times \mathsf{E}_2$ and write

$$\begin{aligned}&\mu(\mathsf{v}_1, \mathsf{v}_2) - \mu(\mathsf{a}_1, \mathsf{a}_2)\\ &= \mu(\mathsf{v}_1 - \mathsf{a}_1, \mathsf{v}_2) + \mu(\mathsf{a}_1, \mathsf{v}_2 - \mathsf{a}_2).\end{aligned}$$

We then have

$$\begin{aligned}&\|\mu(\mathsf{v}_1, \mathsf{v}_2) - \mu(\mathsf{a}_1, \mathsf{a}_2)\|\\ &\leq C\,\|\mathsf{v}_1 - \mathsf{a}_1\|\,\|\mathsf{v}_2\| + C\,\|\mathsf{a}_1\|\,\|\mathsf{v}_2 - \mathsf{a}_2\|\end{aligned}$$

and so if $\|(\mathsf{v}_1, \mathsf{v}_2) - (\mathsf{a}_1, \mathsf{a}_2)\| \to 0$ then $\|\mathsf{v}_i - \mathsf{a}_i\| \to 0$ and we see that

$$\|\mu(\mathsf{v}_1, \mathsf{v}_2) - \mu(\mathsf{a}_1, \mathsf{a}_2)\| \to 0.$$

($\Rightarrow$) Start out by assuming that μ is continuous at $(0, 0)$. Then for $r > 0$ sufficiently small, $(\mathsf{v}_1, \mathsf{v}_2) \in B((0, 0), r)$ implies that $\|\mu(\mathsf{v}_1, \mathsf{v}_2)\| \leq 1$ so if for $i = 1, 2$ we let

$$\mathsf{z}_i := \frac{r\mathsf{v}_i}{\|\mathsf{v}_1\|_i + \epsilon} \quad \text{for some } \epsilon > 0$$

then $(\mathsf{z}_1, \mathsf{z}_2) \in B((0, 0), r)$ and $\|\mu(\mathsf{z}_1, \mathsf{z}_2)\| \leq 1$. The case $(\mathsf{v}_1, \mathsf{v}_2) = (0, 0)$ is trivial so assume $(\mathsf{v}_1, \mathsf{v}_2) \neq (0, 0)$. Then we have

$$\begin{aligned}\mu(\mathsf{z}_1, \mathsf{z}_2) &= \mu\left(\frac{r\mathsf{v}_1}{\|\mathsf{v}_1\| + \epsilon}, \frac{r\mathsf{v}_2}{\|\mathsf{v}_2\| + \epsilon}\right)\\ &= \frac{r^2}{(\|\mathsf{v}_1\| + \epsilon)(\|\mathsf{v}_2\| + \epsilon)}\mu(\mathsf{v}_1, \mathsf{v}_2) \leq 1\end{aligned}$$

and so $\mu(\mathsf{v}_1, \mathsf{v}_2) \leq r^{-2}(\|\mathsf{v}_1\| + \epsilon)(\|\mathsf{v}_2\| + \epsilon)$. Now let $\epsilon \to 0$ to get the result. ■

We shall need to have several Banach spaces handy for examples. For the next example we need some standard notation.

Notation 26.3 *In the context of $\mathbb{R}^n$, we often use the so called "multiindex notation". Let $\alpha = (\alpha_1, ..., \alpha_n)$ where the α_i are integers and $0 \leq \alpha_i \leq n$. Such an n-tuple is called a multiindex. Let $|\alpha| := \alpha_1 + ... + \alpha_n$ and*

$$\frac{\partial^\alpha f}{\partial x^\alpha} := \frac{\partial^{|\alpha|} f}{\partial (x^1)^{\alpha_1} \partial (x^1)^{\alpha_2} \cdots \partial (x^1)^{\alpha_n}}.$$

Example 26.9 *Consider a bounded open subset Ω of $\mathbb{R}^n$. Let $L_k^p(\Omega)$ denote the Banach space obtained by taking the Banach space completion of the set $C^k(\Omega)$ of k-times continuously differentiable real valued functions on Ω with the norm given by*

$$\|f\|_{k,p} := \left(\int_\Omega \sum_{|\alpha| \leq k} \left|\frac{\partial^\alpha}{\partial x^\alpha} f(x)\right|^p\right)^{1/p}$$

Note that in particular $L_0^p(\Omega) = L^p(\Omega)$ is the usual L^p-space from real analysis.

Exercise 26.2 *Show that the map $C^k(\Omega) \to C^{k-1}(\Omega)$ given by $f \mapsto \frac{\partial f}{\partial x^i}$ is bounded if we use the norms $\|f\|_{2,p}$ and $\|f\|_{2-1,p}$. Show that we may extend this to a bounded map $L^p_2(\Omega) \to L^p_1(\Omega)$.*

Proposition 26.4 *There is a natural linear isomorphism $L(\mathsf{V}, L(\mathsf{V},\mathsf{W})) \cong L^2(\mathsf{V},\mathsf{W})$ given by*

$$l(v_1)(v_2) \longleftrightarrow l(v_1, v_2)$$

and we identify the two spaces. In fact, $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V},\mathsf{W})) \cong L^3(\mathsf{V};\mathsf{W})$ and in general $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, ..., L(\mathsf{V},\mathsf{W})) \cong L^k(\mathsf{V};\mathsf{W})$ etc.

Proof. It is easily checked that if we just define $(\iota\ T)(v_1)(v_2) = T(v_1, v_2)$ then $\iota\ T \leftrightarrow T$ does the job for the $k = 2$ case. The $k > 2$ case can be done by an inductive construction and is left as an exercise. It is also not hard to show that the isomorphism is continuous and in fact, norm preserving. ■

Definition 26.34 *A function $f : U \subset \mathsf{V} \to \mathsf{W}$ between Banach spaces and defined on an open set $U \subset \mathsf{V}$ is said to be* ***differentiable at*** *$\mathsf{p} \in U$ iff there is a bounded linear map $A_\mathsf{p} \in L(\mathsf{V},\mathsf{W})$ such that*

$$\lim_{\|\mathsf{h}\|\to 0} \frac{f(\mathsf{p}+\mathsf{h}) - f(\mathsf{p}) - A_\mathsf{p} \cdot \mathsf{h}}{\|\mathsf{h}\|} = 0$$

In anticipation of the following proposition we write $A_\mathsf{p} = Df(\mathsf{p})$. We will also use the notation $Df|_\mathsf{p}$ or sometimes $f'(p)$. The linear map $Df(\mathsf{p})$ is called the ***derivative*** *of f at p.*

We often write $Df|_\mathsf{p} \cdot \mathsf{h}$. The dot in the notation just indicate a linear dependence and is not a literal "dot product". We could also write $Df(\mathsf{p})(\mathsf{h})$.

Exercise 26.3 *Show that the map $F : L^2(\Omega) \to L^1(\Omega)$ given by $F(f) = f^2$ is differentiable at any $f_0 \in L^2(\Omega)$.*

Proposition 26.5 *If A_p exists for a given function f then it is unique.*

Proof. Suppose that A_p and B_p both satisfy the requirements of the definition. That is the limit in question equals zero. For $\mathsf{p}+\mathsf{h} \in U$ we have

$$\begin{aligned} A_\mathsf{p}\mathsf{h} - B_\mathsf{p}\mathsf{h} = &\left(f(\mathsf{p}+\mathsf{h}) - f(\mathsf{p}) - A_\mathsf{p} \cdot \mathsf{h}\right) \\ &- \left(f(\mathsf{p}+\mathsf{h}) - f(\mathsf{p}) - B_\mathsf{p} \cdot \mathsf{h}\right). \end{aligned}$$

Dividing by $\|\mathsf{h}\|$ and taking the limit as $\|\mathsf{h}\| \to 0$ we get

$$\|A_\mathsf{p}\mathsf{h} - B_\mathsf{p}\mathsf{h}\| / \|\mathsf{h}\| \to 0$$

Now let $\mathsf{h} \neq 0$ be arbitrary and choose $\epsilon > 0$ small enough that $\mathsf{p} + \epsilon\mathsf{h} \in U$. Then we have

$$\|A_\mathsf{p}(\epsilon\mathsf{h}) - B_\mathsf{p}(\epsilon\mathsf{h})\| / \|\epsilon\mathsf{h}\| \to 0.$$

But by linearity $\|A_{\mathsf{p}}(\epsilon\mathsf{h}) - B_{\mathsf{p}}(\epsilon\mathsf{h})\| / \|\epsilon\mathsf{h}\| = \|A_{\mathsf{p}}\mathsf{h} - B_{\mathsf{p}}\mathsf{h}\| / \|\mathsf{h}\|$ which doesn't even depend on ϵ so in fact $\|A_{\mathsf{p}}\mathsf{h} - B_{\mathsf{p}}\mathsf{h}\| = 0$. ∎

If we are interested in differentiating "in one direction at a time" then we may use the natural notion of directional derivative. A map has a directional derivative $D_{\mathsf{h}}f$ at p in the direction h if the following limit exists:

$$(D_{\mathsf{h}}f)(\mathsf{p}) := \lim_{\varepsilon \to 0} \frac{f(\mathsf{p} + \varepsilon\mathsf{h}) - f(\mathsf{p})}{\varepsilon}$$

In other words, $D_{\mathsf{h}}f(\mathsf{p}) = \frac{d}{dt}\big|_{t=0} f(\mathsf{p}+t\mathsf{h})$. But a function may have a directional derivative[2] in every direction (at some fixed p), that is, for every $\mathsf{h} \in \mathsf{E}$ and yet still not be differentiable at p in the sense of definition 26.34. Do you remember the usual example of this from multivariable calculus?

Definition 26.35 *If it happens that a function f is differentiable for all p throughout some open set U then we say that f is differentiable on U. We then have a map $Df : U \subset \mathsf{V} \to L(\mathsf{V}, \mathsf{W})$ given by $\mathsf{p} \mapsto Df(\mathsf{p})$. If this map is differentiable at some $\mathsf{p} \in \mathsf{V}$ then its derivative at p is denoted $DDf(\mathsf{p}) = D^2f(\mathsf{p})$ or $D^2f\big|_{\mathsf{p}}$ and is an element of $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}; \mathsf{W})$. Similarly, we may inductively define $D^kf \in L^k(\mathsf{V}; \mathsf{W})$ whenever f is sufficiently nice that the process can continue.*

Definition 26.36 *We say that a map $f : U \subset \mathsf{V} \to \mathsf{W}$ is C^r-differentiable on U if $D^rf\big|_{\mathsf{p}} \in L^r(\mathsf{V}, \mathsf{W})$ exists for all $\mathsf{p} \in U$ and if continuous D^rf as map $U \to L^r(\mathsf{V}, \mathsf{W})$. If f is C^r-differentiable on U for all $r > 0$ then we say that f is C^∞ or **smooth** (on U).*

Exercise 26.4 *Show directly that a bounded multilinear map is C^∞.*

Definition 26.37 *A bijection f between open sets $U_\alpha \subset \mathsf{V}$ and $U_\beta \subset \mathsf{W}$ is called a C^r-**diffeomorphism** iff f and f^{-1} are both C^r-differentiable (on U_α and U_β respectively). If $r = \infty$ then we simply call f a diffeomorphism. Often, we will have $\mathsf{W} = \mathsf{V}$ in this situation.*

*Let U be open in V. A map $f : U \to \mathsf{W}$ is called a **local** C^r **diffeomorphism** iff for every $\mathsf{p} \in U$ there is an open set $U_{\mathsf{p}} \subset U$ with $\mathsf{p} \in U_{\mathsf{p}}$ such that $f|_{U_{\mathsf{p}}} : U_{\mathsf{p}} \to f(U_{\mathsf{p}})$ is a C^r-diffeomorphism.*

In the context of undergraduate calculus courses we are used to thinking of the derivative of a function at some $a \in \mathbb{R}$ as a number $f'(a)$ which is the slope of the tangent line on the graph at $(a, f(a))$. From the current point of view $Df(a) = Df|_a$ just gives the linear transformation $h \mapsto f'(a) \cdot h$ and the equation of the tangent line is given by $y = f(a) + f'(a)(x - a)$. This generalizes to an arbitrary differentiable map as $\mathsf{y} = f(\mathsf{a}) + Df(\mathsf{a}) \cdot (\mathsf{x} - \mathsf{a})$ giving a map which is the linear approximation of f at a.

[2] The directional derivative is written as $(D_{\mathsf{h}}f)(\mathsf{p})$ and if f is differentiable at p this is equal to $Df|_{\mathsf{p}}\,\mathsf{h}$. The notation $D_{\mathsf{h}}f$ should not be confused with $Df|_{\mathsf{h}}$.

We will sometimes think of the derivative of a curve[3] $c : I \subset \mathbb{R} \to \mathsf{E}$ at $t_0 \in I$, written $\dot{c}(t_0)$, as a velocity vector and so we are identifying $\dot{c}(t_0) \in L(\mathbb{R}, \mathsf{E})$ with $Dc|_{t_0} \cdot 1 \in \mathsf{E}$. Here the number 1 is playing the role of the unit vector in $\mathbb{R}$.

Let $f : U \subset \mathsf{E} \to \mathsf{F}$ be a map and suppose that we have a splitting $\mathsf{E} = \mathsf{E}_1 \times \mathsf{E}_2$. We will write $f(\mathsf{x}, \mathsf{y})$ for $(\mathsf{x}, \mathsf{y}) \in \mathsf{E}_1 \times \mathsf{E}_2$. Now for every $(\mathsf{a}, \mathsf{b}) \in \mathsf{E}_1 \times \mathsf{E}_2$ the partial map $f_{\mathsf{a},} : \mathsf{y} \mapsto f(\mathsf{a}, \mathsf{y})$ (resp. $f_{,\mathsf{b}} : \mathsf{x} \mapsto f(\mathsf{x}, \mathsf{b})$) is defined in some neighborhood of b (resp. a). We define the partial derivatives when they exist by $D_2 f(\mathsf{a}, \mathsf{b}) = Df_{\mathsf{a},}(\mathsf{b})$ (resp. $D_1 f(\mathsf{a}, \mathsf{b}) = Df_{,\mathsf{b}}(\mathsf{a})$). These are, of course, linear maps.

$$
\begin{aligned}
D_1 f(\mathsf{a}, \mathsf{b}) &: \mathsf{E}_1 \to \mathsf{F} \\
D_2 f(\mathsf{a}, \mathsf{b}) &: \mathsf{E}_2 \to \mathsf{F}
\end{aligned}
$$

The partial derivative can exist even in cases where f might not be differentiable in the sense we have defined. The point is that f might be differentiable only in certain directions.

If f has continuous partial derivatives $D_i f(\mathsf{x}, \mathsf{y}) : \mathsf{E}_i \to \mathsf{F}$ near $(\mathsf{x}, \mathsf{y}) \in \mathsf{E}_1 \times \mathsf{E}_2$ then $Df(\mathsf{x}, \mathsf{y})$ exists and is continuous near p . In this case,

$$
\begin{aligned}
& Df(\mathsf{x}, \mathsf{y}) \cdot (\mathrm{v}_1, \mathrm{v}_2) \\
& = D_1 f(\mathsf{x}, \mathsf{y}) \cdot \mathrm{v}_1 + D_2 f(\mathsf{x}, \mathsf{y}) \cdot \mathrm{v}_2
\end{aligned}
$$

26.8 Chain Rule, Product rule and Taylor's Theorem

Theorem 26.3 (Chain Rule) *Let U_1 and U_2 be open subsets of Banach spaces E_1 and E_2 respectively. Suppose we have continuous maps composing as*

$$
U_1 \xrightarrow{f} U_2 \xrightarrow{g} \mathsf{E}_3
$$

where E_3 is a third Banach space. If f is differentiable at p and g is differentiable at $f(p)$ then the composition is differentiable at p and $D(g \circ f) = Dg(f(\mathsf{p})) \circ Dg(\mathsf{p})$. In other words, if $\mathsf{v} \in \mathsf{E}_1$ then

$$
D(g \circ f)|_p \cdot \mathsf{v} = Dg|_{f(p)} \cdot (Df|_p \cdot \mathsf{v}).
$$

Furthermore, if $f \in C^r(U_1)$ and $g \in C^r(U_2)$ then $g \circ f \in C^r(U_1)$.

Proof. Let us use the notation $O_1(\mathsf{v})$, $O_2(\mathsf{v})$ etc. to mean functions such that $O_i(\mathsf{v}) \to 0$ as $\|\mathsf{v}\| \to 0$. Let $\mathsf{y} = f(\mathsf{p})$. Since f is differentiable at p we have $f(\mathsf{p} + \mathsf{h}) = \mathsf{y} + Df|_p \cdot \mathsf{h} + \|\mathsf{h}\| O_1(\mathsf{h}) := \mathsf{y} + \Delta\mathsf{y}$ and since g is differentiable at y we have $g(\mathsf{y} + \Delta\mathsf{y}) = Dg|_{\mathsf{y}} \cdot (\Delta\mathsf{y}) + \|\Delta\mathsf{y}\| O_2(\Delta\mathsf{y})$. Now $\Delta\mathsf{y} \to 0$ as $\mathsf{h} \to 0$ and in

[3] We will often use the letter I to denote a generic (usually open) interval in the real line.

turn $O_2(\Delta \mathsf{y}) \to 0$ hence

$$\begin{aligned} g \circ f(\mathsf{p}+\mathsf{h}) &= g(\mathsf{y}+\Delta\mathsf{y}) \\ &= Dg|_{\mathsf{y}} \cdot (\Delta\mathsf{y}) + \|\Delta\mathsf{y}\|\, O_2(\Delta\mathsf{y}) \\ &= Dg|_{\mathsf{y}} \cdot (\, Df|_p \cdot \mathsf{h} + \|\mathsf{h}\|\, O_1(\mathsf{h})) + \|\mathsf{h}\|\, O_3(\mathsf{h}) \\ &= Dg|_{\mathsf{y}} \cdot Df|_p \cdot \mathsf{h} + \|\mathsf{h}\|\, Dg|_{\mathsf{y}} \cdot O_1(\mathsf{h}) + \|\mathsf{h}\|\, O_3(\mathsf{h}) \\ &= Dg|_{\mathsf{y}} \cdot Df|_p \cdot \mathsf{h} + \|\mathsf{h}\|\, O_4(\mathsf{h}) \end{aligned}$$

which implies that $g \circ f$ is differentiable at p with the derivative given by the promised formula.

Now we wish to show that $f, g \in C^r$ $r \geq 1$ implies that $g \circ f \in C^r$ also. The bilinear map defined by composition $comp : L(\mathsf{E}_1, \mathsf{E}_2) \times L(\mathsf{E}_2, \mathsf{E}_3) \to L(\mathsf{E}_1, \mathsf{E}_3)$ is bounded . Define a map

$$m_{f,g} : \mathsf{p} \mapsto (Dg(f(\mathsf{p}), Df(\mathsf{p}))$$

which is defined on U_1. Consider the composition $comp \circ m_{f,g}$. Since f and g are at least C^1 this composite map is clearly continuous. Now we may proceed inductively. Consider the $r-th$ statement:

$$\text{composition of } C^r \text{ maps are } C^r$$

Suppose f and g are C^{r+1} then Df is C^r and $Dg \circ f$ is C^r by the inductive hypothesis so that $m_{f,g}$ is C^r. A bounded bilinear functional is C^∞. Thus $comp$ is C^∞ and by examining $comp \circ m_{f,g}$ we see that the result follows. ■

We will often use the following lemma without explicit mention when calculating:

Lemma 26.2 *Let* $f : U \subset \mathsf{V} \to \mathsf{W}$ *be twice differentiable at* $\mathsf{x}_0 \in U \subset \mathsf{V}$ *then the map* $D_{\mathsf{v}} f : \mathsf{x} \mapsto Df(\mathsf{x}) \cdot \mathsf{v}$ *is differentiable at* x_0 *and its derivative at* x_0 *is given by*

$$D(D_{\mathsf{v}} f)|_{\mathsf{x}_0} \cdot \mathsf{h} = D^2 f(\mathsf{x}_0)(\mathsf{h}, \mathsf{v}).$$

Proof. The map $D_{\mathsf{v}} f : \mathsf{x} \mapsto Df(\mathsf{x}) \cdot \mathsf{v}$ is decomposed as the composition

$$\mathsf{x} \overset{Df}{\mapsto} Df|_{\mathsf{x}} \overset{R^{\mathsf{v}}}{\mapsto} Df|_{\mathsf{x}} \cdot \mathsf{v}$$

where $R^{\mathsf{v}} : L(\mathsf{V}, \mathsf{W}) \mapsto \mathsf{W}$ is the map $(A, b) \mapsto A \cdot b$. The chain rule gives

$$\begin{aligned} D(D_{\mathsf{v}} f)(\mathsf{x}_0) \cdot \mathsf{h} &= DR^{\mathsf{v}}(Df|_{\mathsf{x}_0}) \cdot D(Df)|_{\mathsf{x}_0} \cdot \mathsf{h}) \\ &= DR^{\mathsf{v}}(Df(\mathsf{x}_0)) \cdot (D^2 f(\mathsf{x}_0) \cdot \mathsf{h}). \end{aligned}$$

But R^{v} is linear and so $DR^{\mathsf{v}}(y) = R^{\mathsf{v}}$ for all y. Thus

$$\begin{aligned} D(D_{\mathsf{v}} f)|_{\mathsf{x}_0} \cdot \mathsf{h} &= R^{\mathsf{v}}(D^2 f(\mathsf{x}_0) \cdot \mathsf{h}) \\ &= (D^2 f(\mathsf{x}_0) \cdot \mathsf{h}) \cdot \mathsf{v} = D^2 f(\mathsf{x}_0)(\mathsf{h}, \mathsf{v}). \end{aligned}$$

■

Theorem 26.4 *If $f : U \subset \mathsf{V} \to \mathsf{W}$ is twice differentiable on U such that $D^2 f$ is continuous, i.e. if $f \in C^2(U)$ then $D^2 f$ is symmetric:*

$$D^2 f(p)(\mathsf{w}, \mathsf{v}) = D^2 f(p)(\mathsf{v}, \mathsf{w}).$$

More generally, if $D^k f$ exists and is continuous then $D^k f\ (\mathbf{p}) \in \mathbf{L}^k_{sym}(\mathsf{V}; \mathsf{W})$.

Proof. Let $p \in U$ and define an affine map $A : \mathbb{R}^2 \to \mathsf{V}$ by $A(s,t) := p + s\mathsf{v} + t\mathsf{w}$. By the chain rule we have

$$\frac{\partial^2 (f \circ A)}{\partial s \partial t}(0) = D^2 (f \circ A)(0) \cdot (\mathbf{e}_1, \mathbf{e}_2) = D^2 f(\mathsf{p}) \cdot (\mathsf{v}, \mathsf{w})$$

where $\mathbf{e}_1, \mathbf{e}_2$ is the standard basis of $\mathbb{R}^2$. Thus it suffices to prove that

$$\frac{\partial^2 (f \circ A)}{\partial s \partial t}(0) = \frac{\partial^2 (f \circ A)}{\partial t \partial s}(0).$$

In fact, for any $\ell \in \mathsf{V}^*$ we have

$$\frac{\partial^2 (\ell \circ f \circ A)}{\partial s \partial t}(0) = \ell \left(\frac{\partial^2 (f \circ A)}{\partial s \partial t} \right)(0)$$

and so by the Hahn-Banach theorem it suffices to prove that $\frac{\partial^2 (\ell \circ f \circ A)}{\partial s \partial t}(0) = \frac{\partial^2 (\ell \circ f \circ A)}{\partial t \partial s}(0)$ which is the standard 1-variable version of the theorem which we assume known. The result for $D^k f$ is proven by induction. ∎

Theorem 26.5 *Let $\varrho \in L(\mathsf{F}_1, \mathsf{F}_2; \mathsf{W})$ be a bilinear map and let $f_1 : U \subset \mathsf{E} \to \mathsf{F}_1$ and $f_2 : U \subset \mathsf{E} \to \mathsf{F}_2$ be differentiable (resp. $C^r, r \geq 1$) maps. Then the composition $\varrho(f_1, f_2)$ is differentiable (resp. $C^r, r \geq 1$) on U where $\varrho(f_1, f_2) : \mathsf{x} \mapsto \varrho(f_1(\mathsf{x}), f_2(\mathsf{x}))$. Furthermore,*

$$D\varrho|_{\mathsf{x}} (f_1, f_2) \cdot \mathsf{v} = \varrho(Df_1|_{\mathsf{x}} \cdot \mathsf{v}, f_2(\mathsf{x})) + \varrho(f_1(\mathsf{x}), Df_2|_{\mathsf{x}} \cdot \mathsf{v}).$$

In particular, if F is an algebra with differentiable product $\star$ and $f_1 : U \subset \mathsf{E} \to \mathsf{F}$ and $f_2 : U \subset \mathsf{E} \to \mathsf{F}$ then $f_1 \star f_2$ is defined as a function and

$$D(f_1 \star f_2) \cdot \mathsf{v} = (Df_1 \cdot \mathsf{v}) \star (f_2) + (Df_1 \cdot \mathsf{v}) \star (Df_2 \cdot \mathsf{v}).$$

Proof. This is completely similar to the usual proof of the product rule and is left as an exercise. ∎

The proof of this useful lemma is left as an easy exercise. It is actually quite often that this little lemma saves the day as it were.

It will be useful to define an integral for maps from an interval $[a, b]$ into a Banach space V. First we define the integral for step functions. A function f on an interval $[a, b]$ is a **step function** if there is a partition $a = t_0 < t_1 < \cdots < t_k = b$ such that f is constant, with value say f_i, on each subinterval $[t_i, t_{i+1})$.

The set of step functions so defined is a vector space. We define the integral of a step function f over $[a,b]$ by

$$\int_{[a,b]} f := \sum_{i=0}^{k-1}(t_{i+1}-t_i)f_i = \sum_{i=0}^{k-1}(t_{i+1}-t_i)f(t_i).$$

One checks that the definition is independent of the partition chosen. Now the set of all step functions from $[a,b]$ into V is a linear subspace of the Banach space $\mathcal{B}(a,b,\mathsf{V})$ of all bounded functions of $[a,b]$ into V and the integral is a linear map on this space. Recall that the norm on $\mathcal{B}(a,b,\mathsf{V})$ is $\sup_{a\leq t<b}\{f(t)\}$. If we denote the closure of the space of step functions in this Banach space by $\bar{\mathcal{S}}(a,b,\mathsf{V})$ then we can extend the definition of the integral to $\bar{\mathcal{S}}(a,b,\mathsf{V})$ by continuity since on step functions we have

$$\left|\int_{[a,b]} f\right| \leq (b-a)\left\|f\right\|_\infty .$$

In the limit, this bound persists. This integral is called the **Cauchy-Bochner** integral and is a bounded linear map $\bar{\mathcal{S}}(a,b,\mathsf{V}) \to \mathsf{V}$. It is important to notice that $\bar{\mathcal{S}}(a,b,\mathsf{V})$ contains the continuous functions $C([a,b],\mathsf{V})$ because such may be uniformly approximated by elements of $\mathcal{S}(a,b,\mathsf{V})$ and so we can integrate these functions using the Cauchy-Bochner integral.

Lemma 26.3 *If $\ell : \mathsf{V} \to \mathsf{W}$ is a bounded linear map of Banach spaces then for any $f \in \bar{\mathcal{S}}(a,b,\mathsf{V})$ we have*

$$\int_{[a,b]} \ell\circ f = \ell\circ\int_{[a,b]} f.$$

Proof. This is obvious for step functions. The general result follows by taking a limit for step functions converging in $\bar{\mathcal{S}}(a,b,\mathsf{V})$ to f. ■

Some facts about maps on finite dimensional spaces.

Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be a map which is differentiable at $a = (a^1,...,a^n) \in \mathbb{R}^n$. The map f is given by m functions $f^i : U \to \mathbb{R}^m$, $1 \leq i \leq m$. Now with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the derivative is given by an $n \times m$ matrix called the **Jacobian** matrix:

$$J_a(f) := \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(a) & \frac{\partial f^1}{\partial x^2}(a) & \cdots & \frac{\partial f^1}{\partial x^n}(a) \\ \frac{\partial f^2}{\partial x^1}(a) & & & \\ \vdots & & \ddots & \\ \frac{\partial f^m}{\partial x^1}(a) & & & \frac{\partial f^m}{\partial x^n}(a) \end{pmatrix}.$$

The rank of this matrix is called the rank of f at a. If $n = m$ so that $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ then the Jacobian is a square matrix and $\det(J_a(f))$ is called the **Jacobian determinant** at a. If f is differentiable near a then it follows from the inverse mapping theorem proved below that if $\det(J_a(f)) \neq 0$ then there is some open set containing a on which f has a differentiable inverse. The Jacobian of this inverse at $f(x)$ is the inverse of the Jacobian of f at x.

The following is the **mean value theorem**:

Theorem 26.6 *Let* V *and* W *be Banach spaces. Let* $c : [a,b] \to \mathsf{V}$ *be a* C^1-*map with image contained in an open set* $U \subset \mathsf{V}$. *Also, let* $f : U \to \mathsf{W}$ *be a* C^1 *map. Then*

$$f(c(b)) - f(c(a)) = \int_0^1 Df(c(t)) \cdot c'(t)dt.$$

If $c(t) = (1-t)\mathsf{x} + t\mathsf{y}$ *then*

$$f(\mathsf{y}) - f(\mathsf{x}) = \int_0^1 Df(c(t))dt \cdot (\mathsf{y} - \mathsf{x})$$

Notice that $\int_0^1 Df(c(t))dt \in L(\mathsf{V}, \mathsf{W})$.

Proof. Use the chain rule and the 1-variable fundamental theorem of calculus for the first part. For the second use lemma 26.3. ∎

Corollary 26.1 *Let* U *be a convex open set in a Banach space* V *and* $f : U \to \mathsf{W}$ *a* C^1 *map into another Banach space* W. *Then for any* $\mathsf{x}, \mathsf{y} \in U$ *we have*

$$\|f(\mathsf{y}) - f(\mathsf{x})\| \leq C_{\mathsf{x},\mathsf{y}} \|\mathsf{y} - \mathsf{x}\|$$

where $C_{\mathsf{x},\mathsf{y}}$ *is the supremum over all values taken by* f *along the line segment which is the image of the path* $t \mapsto (1-t)\mathsf{x} + t\mathsf{y}$.

Recall that for a fixed x higher derivatives $D^p f\big|_{\mathsf{x}}$ are symmetric multilinear maps. For the following let $(y)^k$ denote $\underbrace{(y, y, ..., y)}_{k\text{-times}}$. With this notation we have the following version of Taylor's theorem.

Theorem 26.7 (Taylor's theorem) *Given Banach spaces* V *and* W, *a* C^r *function* $f : U \to \mathsf{W}$ *and a line segment* $t \mapsto (1-t)\mathsf{x} + t\mathsf{y}$ *contained in* U, *we have that* $t \mapsto D^p f(x + ty) \cdot (y)^p$ *is defined and continuous for* $1 \leq p \leq k$ *and*

$$f(x+y) = f(x) + \frac{1}{1!} Df\big|_x \cdot y + \frac{1}{2!} D^2 f\big|_x \cdot (y)^2 + \cdots + \frac{1}{(k-1)!} D^{k-1} f\big|_x \cdot (y)^{k-1}$$
$$+ \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} D^k f(x+ty) \cdot (y)^k dt$$

Proof. The proof is by induction and follows the usual proof closely. The point is that we still have an integration by parts formula coming from the product rule and we still have the fundamental theorem of calculus. ∎

26.9 Local theory of maps

Inverse Mapping Theorem

The main reason for restricting our calculus to Banach spaces is that the inverse mapping theorem holds for Banach spaces and there is no simple and general inverse mapping theory on more general topological vector spaces.

Definition 26.38 *Let* E *and* F *be Banach spaces. A map will be called a* C^r ***diffeomorphism near*** p *if there is some open set* $U \subset \operatorname{dom}(f)$ *containing* p *such that* $f|_U : U \to f(U)$ *is a* C^r *diffeomorphism onto an open set* $f(U)$*. The set of all maps which are diffeomorphisms near* p *will be denoted* $\operatorname{Diff}^r_p(\mathsf{E},\mathsf{F})$*. If* f *is a* C^r *diffeomorphism near* p *for all* $\mathsf{p} \in U = \operatorname{dom}(f)$ *then we say that* f *is a* ***local*** C^r ***diffeomorphism****.*

Definition 26.39 *Let* X, d_1 *and* Y, d_2 *be metric spaces. A map* $f : X \to Y$ *is said to be* ***Lipschitz continuous*** *(with constant* k*) if there is a* $k > 0$ *such that* $d(f(x_1), f(x_2)) \le kd(x_1, x_2)$ *for all* $x_1, x_2 \in X$*. If* $0 < k < 1$ *the map is called a* ***contraction mapping*** *(with constant* k*) or is said to be* k***-contractive****.*

The following technical result has numerous applications and uses the idea of iterating a map. **Warning**: For this theorem f^n will denote the $n-$fold composition $f \circ f \circ \cdots \circ f$ rather than a product.

Proposition 26.6 (Contraction Mapping Principle) *Let* F *be a closed subset of a complete metric space* (M, d)*. Let* $f : F \to F$ *be a* k*-contractive map such that*

$$d(f(x), f(y)) \le kd(x, y)$$

for some fixed $0 \le k < 1$*. Then*

1) there is exactly one $x_0 \in F$ *such that* $f(x_0) = x_0$*. Thus* x_0 *is a fixed point for* f*. Furthermore,*

2) for any $y \in F$ *the sequence* $y_n := f^n(y)$ *converges to the fixed point* x_0 *with the error estimate* $d(y_n, x_0) \le \frac{k^n}{1-k} d(y_1, x_0)$*.*

Proof. Let $y \in F$. By iteration

$$d(f^n(y), f^{n-1}(y)) \le kd(f^{n-1}(y), f^{n-2}(y)) \le \cdots \le k^{n-1} d(f(y), y)$$

as follows:

$$\begin{aligned} d(f^{n+j+1}(y), f^n(y)) &\le d(f^{n+j+1}(y), f^{n+j}(y)) + \cdots + d(f^{n+1}(y), f^n(y)) \\ &\le (k^{j+1} + \cdots + k) d(f^n(y), f^{n-1}(y)) \\ &\le \frac{k}{1-k} d(f^n(y), f^{n-1}(y)) \\ &\frac{k^n}{1-k} d(f^1(y), y)) \end{aligned}$$

From this, and the fact that $0 \leq k < 1$, one can conclude that the sequence $f^n(y) = x_n$ is Cauchy. Thus $f^n(y) \to x_0$ for some x_0 which is in F since F is closed. On the other hand,

$$x_0 = \lim_{n\to 0} f^n(y) = \lim_{n\to 0} f(f^{n-1}(y)) = f(x_0)$$

by continuity of f. Thus x_0 is a fixed point. If u_0 where also a fixed point then

$$d(x_0, u_0) = d(f(x_0), f(u_0)) \leq kd(x_0, u_0)$$

which forces $x_0 = u_0$. The error estimate in (2) of the statement of the theorem is left as an easy exercise. ■

Remark 26.9 *Note that a Lipschitz map f may not satisfy the hypotheses of the last theorem even if $k < 1$ since U is not a complete metric space unless $U = \mathsf{E}$.*

Definition 26.40 *A continuous map $f : U \to \mathsf{E}$ such that $L_f := \mathrm{id}_U - f$ is injective has a not necessarily continuous inverse G_f and the invertible map $R_f := \mathrm{id}_{\mathsf{E}} - G_f$ will be called the* ***resolvent operator*** *for f.*

The resolvent is a term that is usually used in the context of linear maps and the definition in that context may vary slightly. Namely, what we have defined here would be the resolvent of $\pm L_f$. Be that as it may, we have the following useful result.

Theorem 26.8 *Let E be a Banach space. If $f : \mathsf{E} \to \mathsf{E}$ is continuous map that is Lipschitz continuous with constant k where $0 \leq k < 1$, then the resolvent R_f exists and is Lipschitz continuous with constant $\frac{k}{1-k}$.*

Proof. Consider the equation $x - f(x) = y$. We claim that for any $y \in \mathsf{E}$ this equation has a unique solution. This follows because the map $F : \mathsf{E} \to \mathsf{E}$ defined by $F(x) = f(x) + y$ is k-contractive on the complete normed space E as a result of the hypotheses. Thus by the contraction mapping principle there is a unique x fixed by F which means a unique x such that $f(x) + y = x$. Thus the inverse G_f exists and is defined on all of E. Let $R_f := \mathrm{id}_{\mathsf{E}} - G_f$ and choose $y_1, y_2 \in \mathsf{E}$ and corresponding unique x_i , $i = 1, 2$ with $x_i - f(x_i) = y_i$. We have

$$\begin{aligned}
\|R_f(y_1) - R_f(y_2)\| &= \|f(x_1) - f(x_2)\| \\
&\leq k\,\|x_1 - x_2\| \ \ \leq \\
&\leq k\,\|y_1 - R_f(y_1) - (y_2 - R_f(y_2))\| \leq \\
&\leq k\,\|y_1 - y_2\| + \|R_f(y_1) - R_f(y_2)\|.
\end{aligned}$$

Solving this inequality we get

$$\|R_f(y_1) - R_f(y_2)\| \leq \frac{k}{1-k}\,\|y_1 - y_2\|.$$

■

Lemma 26.4 *The space* $\mathrm{GL}(\mathsf{E},\mathsf{F})$ *of continuous linear isomorphisms is an open subset of the Banach space* $L(\mathsf{E},\mathsf{F})$*. In particular, if* $\|\mathrm{id}-A\|<1$ *for some* $A\in\mathrm{GL}(\mathsf{E})$ *then* $A^{-1}=\lim_{N\to\infty}\sum_{n=0}^{N}(\mathrm{id}-A)^n$*.*

Proof. Let $A_0\in\mathrm{GL}(\mathsf{E},\mathsf{F})$. The map $A\mapsto A_0^{-1}\circ A$ is continuous and maps $\mathrm{GL}(\mathsf{E},\mathsf{F})$ onto $\mathrm{GL}(\mathsf{E},\mathsf{F})$. If follows that we may assume that $\mathsf{E}=\mathsf{F}$ and that $A_0=\mathrm{id}_{\mathsf{E}}$. Our task is to show that elements of $\mathrm{L}(\mathsf{E},\mathsf{E})$ close enough to id_{E} are in fact elements of $\mathrm{GL}(\mathsf{E})$. For this we show that

$$\|\mathrm{id}-A\|<1$$

implies that $A\in\mathrm{GL}(\mathsf{E})$. We use the fact that the norm on $\mathrm{L}(\mathsf{E},\mathsf{E})$ is an algebra norm. Thus $\|A_1\circ A_2\|\le\|A_1\|\,\|A_2\|$ for all $A_1,A_2\in\mathrm{L}(\mathsf{E},\mathsf{E})$. We abbreviate id by "1" and denote $\mathrm{id}-A$ by Λ. Let $\Lambda^2:=\Lambda\circ\Lambda$, $\Lambda^3:=\Lambda\circ\Lambda\circ\Lambda$ and so forth. We now form a Neumann series :

$$\begin{aligned}
\pi_0 &= 1\\
\pi_1 &= 1+\Lambda\\
\pi_2 &= 1+\Lambda+\Lambda^2\\
&\vdots\\
\pi_n &= 1+\Lambda+\Lambda^2+\cdots+\Lambda^n.
\end{aligned}$$

By comparison with the Neumann series of real numbers formed in the same way using $\|A\|$ instead of A we see that $\{\pi_n\}$ is a Cauchy sequence since $\|\Lambda\|=\|\mathrm{id}-A\|<1$. Thus $\{\pi_n\}$ is convergent to some element ρ. Now we have $(1-\Lambda)\pi_n=1-\Lambda^{n+1}$ and letting $n\to\infty$ we see that $(1-\Lambda)\rho=1$ or in other words, $A\rho=1$. ■

Lemma 26.5 *The map* $\mathrm{inv}:\mathrm{GL}(\mathsf{E},\mathsf{F})\to\mathrm{GL}(\mathsf{E},\mathsf{F})$ *given by taking inverses is a* C^∞*map and the derivative of* $\mathrm{inv}: g\mapsto g^{-1}$ *at some* $g_0\in\mathrm{GL}(\mathsf{E},\mathsf{F})$ *is the linear map given by the formula:* $D\,\mathrm{inv}|_{g_0}: A\mapsto -g_0^{-1}Ag_0^{-1}$*.*

Proof. Suppose that we can show that the result is true for $g_0=\mathrm{id}$. Then pick any $h_0\in\mathrm{GL}(\mathsf{E},\mathsf{F})$ and consider the isomorphisms $L_{h_0}:\mathrm{GL}(\mathsf{E})\to\mathrm{GL}(\mathsf{E},\mathsf{F})$ and $R_{h_0^{-1}}:\mathrm{GL}(\mathsf{E})\to\mathrm{GL}(\mathsf{E},\mathsf{F})$ given by $\phi\mapsto h_0\phi$ and $\phi\mapsto\phi h_0^{-1}$ respectively. The map $g\mapsto g^{-1}$ can be decomposed as

$$g\overset{L_{h_0^{-1}}}{\mapsto}h_0^{-1}\circ g\overset{\mathrm{inv}_{\mathsf{E}}}{\mapsto}(h_0^{-1}\circ g)^{-1}\overset{R_{h_0^{-1}}}{\mapsto}g^{-1}h_0h_0^{-1}=g^{-1}.$$

Now suppose that we have the result at $g_0=\mathrm{id}$ in $\mathrm{GL}(\mathsf{E})$. This means that $D\mathrm{inv}_{\mathsf{E}}|_{h_0}: A\mapsto -A$. Now by the chain rule we have

$$\begin{aligned}
(D\,\mathrm{inv}|_{h_0})\cdot \mathrm{A} &= D(R_{h_0^{-1}}\circ\mathrm{inv}_{\mathsf{E}}\circ L_{h_0^{-1}})\cdot\mathrm{A}\\
&= \left(R_{h_0^{-1}}\circ D\,\mathrm{inv}_{\mathsf{E}}|_{id}\circ L_{h_0^{-1}}\right)\cdot\mathrm{A}\\
&= R_{h_0^{-1}}\circ(-\mathrm{A})\circ L_{h_0^{-1}} = -h_0^{-1}\mathrm{A}h_0^{-1}
\end{aligned}$$

so the result is true for an arbitrary $h_0 \in \mathrm{GL}(\mathsf{E},\mathsf{F})$. Thus we are reduced to showing that $D\mathrm{inv}_{\mathsf{E}}\big|_{\mathrm{id}} : A \mapsto -A$. The definition of derivative leads us to check that the following limit is zero.

$$\lim_{\|\mathrm{A}\|\to 0} \frac{\left\|(\mathrm{id}+\mathrm{A})^{-1} - (\mathrm{id})^{-1} - (-\mathrm{A})\right\|}{\|\mathrm{A}\|}.$$

Note that for small enough $\|\mathrm{A}\|$, the inverse $(\mathrm{id}+A)^{-1}$ exists and so the above limit makes sense. By our previous result (26.4) the above difference quotient becomes

$$\begin{aligned}
&\lim_{\|\mathrm{A}\|\to 0} \frac{\left\|(\mathrm{id}+\mathrm{A})^{-1} - \mathrm{id}+\mathrm{A}\right\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \frac{\|\sum_{n=0}^{\infty}(\mathrm{id}-(\mathrm{id}+\mathrm{A}))^n - \mathrm{id}+\mathrm{A}\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \frac{\|\sum_{n=0}^{\infty}(-\mathrm{A})^n - \mathrm{id}+\mathrm{A}\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \frac{\|\sum_{n=2}^{\infty}(-\mathrm{A})^n\|}{\|\mathrm{A}\|} \le \lim_{\|\mathrm{A}\|\to 0} \frac{\sum_{n=2}^{\infty}\|\mathrm{A}\|^n}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \sum_{n=1}^{\infty}\|\mathrm{A}\|^n = \lim_{\|\mathrm{A}\|\to 0} \frac{\|\mathrm{A}\|}{1-\|\mathrm{A}\|} = 0.
\end{aligned}$$

■

Theorem 26.9 (Inverse Mapping Theorem) *Let E and F be Banach spaces and $f : U \to \mathsf{F}$ be a C^r mapping defined an open set $U \subset \mathsf{E}$. Suppose that $\mathsf{x}_0 \in U$ and that $f'(\mathsf{x}_0) = Df|_x : \mathsf{E} \to \mathsf{F}$ is a continuous linear isomorphism. Then there exists an open set $V \subset U$ with $\mathsf{x}_0 \in V$ such that $f : V \to f(V) \subset \mathsf{F}$ is a C^r-diffeomorphism. Furthermore the derivative of f^{-1} at y is given by $Df^{-1}\big|_{\mathsf{y}} = (Df|_{f^{-1}(\mathsf{y})})^{-1}$.*

Proof. By considering $(Df|_x)^{-1} \circ f$ and by composing with translations we may as well just assume from the start that $f : \mathsf{E} \to \mathsf{E}$ with $\mathsf{x}_0 = 0$, $f(0) = 0$ and $Df|_0 = \mathrm{id}_E$. Now if we let $g = x - f(x)$, then $Dg|_0 = 0$ and so if $r > 0$ is small enough then

$$\|Dg|_0\| < \frac{1}{2}$$

for $x \in B(0,2r)$. The mean value theorem now tells us that $\|g(x_2) - g(x_1)\| \le \frac{1}{2}\|x_2 - x_1\|$ for $x_2, x_1 \in \overline{B}(0,r)$ and that $g(\overline{B}(0,r)) \subset \overline{B}(0,r/2)$. Let $y_0 \in \overline{B}(0,r/2)$. It is not hard to show that the map $c : x \mapsto y_0 + x - f(x)$ is a contraction mapping $c : \overline{B}(0,r) \to \overline{B}(0,r)$ with constant $\frac{1}{2}$. The contraction mapping principle 26.6 says that c has a unique fixed point $x_0 \in \overline{B}(0,r)$. But $c(x_0) = x_0$ just translates to $y_0 + x_0 - f(x_0) = x_0$ and then $f(x_0) = y_0$. So x_0 is the unique element of $\overline{B}(0,r)$ satisfying this equation. But then since $y_0 \in$

$\overline{B}(0,r/2)$ was an arbitrary element of $\overline{B}(0,r/2)$ it follows that the restriction $f:\overline{B}(0,r/2)\to f(\overline{B}(0,r/2))$ is invertible. But f^{-1} is also continuous since

$$\begin{aligned}\left\|f^{-1}(y_2)-f^{-1}(y_1)\right\| &= \|x_2-x_1\| \\ &\le \|f(x_2)-f(x_1)\|+\|g(x_2)-g(x_1)\| \\ &\le \|f(x_2)-f(x_1)\|+\frac{1}{2}\|x_2-x_1\| \\ &= \|y_2-y_1\|+\frac{1}{2}\left\|f^{-1}(y_2)-f^{-1}(y_1)\right\|\end{aligned}$$

Thus $\left\|f^{-1}(y_2)-f^{-1}(y_1)\right\|\le 2\|y_2-y_1\|$ and so f^{-1} is continuous. In fact, f^{-1} is also differentiable on $B(0,r/2)$. To see this let $f(x_2)=y_2$ and $f(x_1)=y_1$ with $x_2,x_1\in\overline{B}(0,r)$ and $y_2,y_1\in\overline{B}(0,r/2)$. The norm of $Df(x_1))^{-1}$ is bounded (by continuity) on $\overline{B}(0,r)$ by some number B. Setting $x_2-x_1=\Delta x$ and $y_2-y_1=\Delta y$ and using $(Df(x_1))^{-1}Df(x_1)=\mathrm{id}$ we have

$$\begin{aligned}&\left\|f^{-1}(y_2)-f^{-1}(y_1)-(Df(x_1))^{-1}\cdot\Delta y\right\| \\ &=\left\|\Delta x-(Df(x_1))^{-1}(f(x_2)-f(x_1))\right\| \\ &=\left\|\{(Df(x_1))^{-1}Df(x_1)\}\Delta x-\{(Df(x_1))^{-1}Df(x_1)\}(Df(x_1))^{-1}(f(x_2)-f(x_1))\right\| \\ &\le B\|Df(x_1)\Delta x-(f(x_2)-f(x_1))\|\le o(\Delta x)=o(\Delta y)\ \text{(by continuity)}.\end{aligned}$$

Thus $Df^{-1}(y_1)$ exists and is equal to $(Df(x_1))^{-1}=(Df(f^{-1}(y_1)))^{-1}$. A simple argument using this last equation shows that $Df^{-1}(y_1)$ depends continuously on y_1 and so f^{-1} is C^1. The fact that f^{-1} is actually C^r follows from a simple induction argument that uses the fact that Df is C^{r-1} together with lemma 26.5. This last step is left to the reader. ∎

Exercise 26.5 *Complete the last part of the proof of theorem*

Corollary 26.2 *Let $U\subset\mathsf{E}$ be an open set and $0\in U$. Suppose that $f:U\to\mathsf{F}$ is differentiable with $Df(p):\mathsf{E}\to\mathsf{F}$ a (bounded) linear isomorphism for each $p\in U$. Then f is a local diffeomorphism.*

Theorem 26.10 (Implicit Function Theorem I) *Let $\mathsf{E}_1,\mathsf{E}_2$ and F be Banach spaces and $U\times V\subset\mathsf{E}_1\times\mathsf{E}_2$ open. Let $f:U\times V\to\mathsf{F}$ be a C^r mapping such that $f(\mathsf{x}_0,\mathsf{y}_0)=0$. If $D_2f_{(\mathsf{x}_0,\mathsf{y}_0)}:\mathsf{E}_2\to\mathsf{F}$ is a continuous linear isomorphism then there exists a (possibly smaller) open set $U_0\subset U$ with $\mathsf{x}_0\in U_0$ and unique mapping $g:U_0\to V$ with $g(x_0)=y_0$ and such that*

$$f(\mathsf{x},g(\mathsf{x}))=0$$

for all $x\in U_0$.

Proof. Follows from the following theorem. ∎

Theorem 26.11 (Implicit Function Theorem II) *Let $\mathsf{E}_1,\mathsf{E}_2$ and F be Banach spaces and $U\times V\subset\mathsf{E}_1\times\mathsf{E}_2$ open. Let $f:U\times V\to\mathsf{F}$ be a C^r mapping such* *that* $f(\mathsf{x}_0, \mathsf{y}_0) = \mathsf{w}_0$. *If* $D_2 f(\mathsf{x}_0, \mathsf{y}_0) : \mathsf{E}_2 \to \mathsf{F}$ *is a continuous linear isomorphism then there exists (possibly smaller) open sets* $U_0 \subset U$ *and* $W_0 \subset \mathsf{F}$ *with* $\mathsf{x}_0 \in U_0$ *and* $\mathsf{w}_0 \in W_0$ *together with a unique mapping* $g : U_0 \times W_0 \to V$ *such that*

$$f(\mathsf{x}, g(\mathsf{x}, \mathsf{w})) = \mathsf{w}$$

for all $\mathsf{x} \in U_0$. *Here unique means that any other such function* h *defined on a neighborhood* $U_0' \times W_0'$ *will equal* g *on some neighborhood of* $(\mathsf{x}_0, \mathsf{w}_0)$.

Proof. Sketch: Let $\Psi : U \times V \to \mathsf{E}_1 \times \mathsf{F}$ be defined by $\Psi(\mathsf{x}, \mathsf{y}) = (\mathsf{x}, f(\mathsf{x}, \mathsf{y}))$. Then $D\Psi(\mathsf{x}_0, \mathsf{y}_0)$ has the operator matrix

$$\begin{bmatrix} \mathrm{id}_{E_1} & 0 \\ D_1 f(\mathsf{x}_0, \mathsf{y}_0) & D_2 f(\mathsf{x}_0, \mathsf{y}_0) \end{bmatrix}$$

which shows that $D\Psi(\mathsf{x}_0, \mathsf{y}_0)$ is an isomorphism. Thus Ψ has a unique local inverse Ψ^{-1} which we may take to be defined on a product set $U_0 \times W_0$. Now Ψ^{-1} must have the form $(\mathsf{x}, \mathsf{y}) \mapsto (\mathsf{x}, g(\mathsf{x}, \mathsf{y}))$ which means that $(\mathsf{x}, f(\mathsf{x}, g(\mathsf{x}, \mathsf{w}))) = \Psi(\mathsf{x}, g(\mathsf{x}, \mathsf{w})) = (\mathsf{x}, \mathsf{w})$. Thus $f(\mathsf{x}, g(\mathsf{x}, \mathsf{w})) = \mathsf{w}$. The fact that g is unique follows from the local uniqueness of the inverse Ψ^{-1} and is left as an exercise. ∎

Let U be an open subset of a Banach space V and let $I \subset \mathbb{R}$ be an open interval containing 0. A (local) time dependent vector field on U is a C^r-map $F : I \times U \to \mathsf{V}$ (where $r \geq 0$). An integral curve of F with initial value x_0 is a map c defined on an open subinterval $J \subset I$ also containing 0 such that

$$\begin{aligned} c'(t) &= F(t, c(t)) \\ c(0) &= x_0 \end{aligned}$$

A local flow for F is a map $\alpha : I_0 \times U_0 \to \mathsf{V}$ such that $U_0 \subset U$ and such that the curve $\alpha_x(t) = \alpha(t, x)$ is an integral curve of F with $\alpha_x(0) = x$

In the case of a map $f : U \to V$ between open subsets of Euclidean spaces (say $\mathbb{R}^n$ and $\mathbb{R}^m$) we have the notion of **rank at** $p \in U$ which is just the rank of the linear map $D_p f : \mathbb{R}^n \to \mathbb{R}^m$.

Definition 26.41 *Let* X, Y *be topological spaces. When we write* $f :: \mathsf{X} \to \mathsf{Y}$ *we imply only that* f *is defined on some open set in* X. *If we wish to indicate that* f *is defined near* $p \in \mathsf{X}$ *and that* $f(\mathsf{p}) = \mathsf{q}$ *we will used the pointed category notation together with the symbol "::":*

$$f :: (\mathsf{X}, \mathsf{p}) \to (\mathsf{Y}, \mathsf{q})$$

We will refer to such maps as ***local maps at*** p. *Local maps may be composed with the understanding that the domain of the composite map may become smaller: If* $f :: (\mathsf{X}, \mathsf{p}) \to (\mathsf{Y}, \mathsf{q})$ *and* $g :: (\mathsf{Y}, \mathsf{q}) \to (\mathsf{G}, \mathsf{z})$ *then* $g \circ f :: (\mathsf{X}, \mathsf{p}) \to (\mathsf{G}, \mathsf{z})$ *and the domain of* $g \circ f$ *will be a non-empty open set.*

Recall that for a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ which is injective with rank r there exist linear isomorphisms $C_1 : \mathbb{R}^m \to \mathbb{R}^m$ and $C_2 : \mathbb{R}^n \to \mathbb{R}^n$ such that $C_1 \circ A \circ C_2^{-1}$ is just a projection followed by an injection:

$$\mathbb{R}^n = \mathbb{R}^r \times \mathbb{R}^{n-r} \to \mathbb{R}^r \times 0 \to \mathbb{R}^r \times \mathbb{R}^{m-r} = \mathbb{R}^m$$

We have obvious special cases when $r = n$ or $r = m$. This fact has a local version that applies to C^∞nonlinear maps.

26.9.1 Linear case.

Definition 26.42 *We say that a continuous linear map $A_1 : \mathsf{E}_1 \to \mathsf{F}_1$ is* ***equivalent*** *to a map $A_2 : \mathsf{E}_2 \to \mathsf{F}_2$ if there are continuous linear isomorphisms $\alpha : \mathsf{E}_1 \to \mathsf{E}_2$ and $\beta : \mathsf{F}_1 \to \mathsf{F}_2$ such that $A_2 = \beta \circ A_1 \circ \alpha^{-1}$.*

Definition 26.43 *Let $A : \mathsf{E} \to \mathsf{F}$ be an injective continuous linear map. We say that A is a* ***splitting injection*** *if there are Banach spaces F_1 and F_2 with $\mathsf{F} \cong \mathsf{F}_1 \times \mathsf{F}_2$ and if A is equivalent to the injection $\mathrm{inj}_1 : \mathsf{F}_1 \to \mathsf{F}_1 \times \mathsf{F}_2$.*

Lemma 26.6 *If $A : \mathsf{E} \to \mathsf{F}$ is a splitting injection as above then there exists a linear isomorphism $\delta : \mathsf{F} \to \mathsf{E} \times \mathsf{F}_2$ such that $\delta \circ A : \mathsf{E} \to \mathsf{E} \times \mathsf{F}_2$ is the injection $\mathsf{x} \mapsto (\mathsf{x}, 0)$.*

Proof. By definition there are isomorphisms $\alpha : \mathsf{E} \to \mathsf{F}_1$ and $\beta : \mathsf{F} \to \mathsf{F}_1 \times \mathsf{F}_2$ such that $\beta \circ A \circ \alpha^{-1}$ is the injection $\mathsf{F}_1 \to \mathsf{F}_1 \times \mathsf{F}_2$. Since α is an isomorphism we may compose as follows

$$\begin{aligned}
&(\alpha^{-1} \times \mathrm{id}_{\mathsf{E}}) \circ \beta \circ A \circ \alpha^{-1} \circ \alpha \\
&= (\alpha^{-1} \times \mathrm{id}_{\mathsf{E}}) \circ \beta \circ A \\
&= \delta \circ A
\end{aligned}$$

to get a map which is can seen to have the correct form.

$$\begin{array}{lcc}
(\mathsf{E}, \mathsf{p}) & \to & (\mathsf{F}, \mathsf{q}) \\
\alpha \downarrow\uparrow \alpha^{-1} & & \updownarrow \beta \\
(\mathsf{F}_1, 0) & \to & (\mathsf{F}_1 \times \mathsf{F}_2, \quad (0,0)) \\
\alpha^{-1} \downarrow\uparrow \alpha & & \updownarrow \alpha^{-1} \times \mathrm{id} \\
(\mathsf{E}, \mathsf{p}) & \overset{\mathrm{inj}_1}{\to} & (\mathsf{E} \times \mathsf{F}_2, \quad (\mathsf{p}, 0))
\end{array}$$

■

If A is a splitting injection as above it easy to see that there are closed subspaces F_1 and F_2 of F such that $\mathsf{F} = \mathsf{F}_1 \oplus \mathsf{F}_2$ and such that A maps E isomorphically onto F_1 .

Definition 26.44 *Let $A : \mathsf{E} \to \mathsf{F}$ be an surjective continuous linear map. We say that A is a* ***splitting surjection*** *if there are Banach spaces E_1 and E_2 with $\mathsf{E} \cong \mathsf{E}_1 \times \mathsf{E}_2$ and if A is equivalent to the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$.*

Lemma 26.7 *If Let $A : \mathsf{E} \to \mathsf{F}$ is a splitting surjection then there is a linear isomorphism $\delta : \mathsf{F} \times \mathsf{E}_2 \to \mathsf{E}$ such that $A \circ \delta : \mathsf{F} \times \mathsf{E}_2 \to \mathsf{F}$ is the projection $(\mathsf{x}, \mathsf{y}) \mapsto \mathsf{x}$.*

Proof. By definition there exist isomorphisms $\alpha : \mathsf{E} \to \mathsf{E}_1 \times \mathsf{E}_2$ and $\beta : \mathsf{F} \to \mathsf{E}_1$ such that $\beta \circ A \circ \alpha^{-1}$ is the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$. We form another map by composition by isomorphisms;

$$\begin{aligned}&\beta^{-1} \circ \beta \circ A \circ \alpha^{-1} \circ (\beta, \mathrm{id}_{\mathsf{E}_2}) \\ &= A \circ \alpha^{-1} \circ (\beta, \mathrm{id}_{\mathsf{E}_2}) := A \circ \delta\end{aligned}$$

and check that this does the job. Examine the diagram for guidance if you get lost:

$$\begin{array}{ccc} (\mathsf{E}, \mathsf{p}) & \overset{A}{\to} & (\mathsf{F}, \mathsf{q}) \\ \alpha \downarrow & & \beta^{-1} \uparrow\downarrow \beta \\ (\mathsf{E}_1 \times \mathsf{E}_2, (0,0)) & \to & (\mathsf{E}_1, 0) \\ (\beta, \mathrm{id}_{\mathsf{E}_2}) \updownarrow & & \beta \uparrow\downarrow \beta^{-1} \\ (\mathsf{F} \times \mathsf{E}_2, (\mathsf{q}, 0)) & \overset{pr_1}{\to} & (\mathsf{F}, \mathsf{q}) \end{array}$$

■

If A is a splitting surjection as above it easy to see that there are closed subspaces E_1 and E_2 of E such that $\mathsf{E} = \mathsf{E}_1 \oplus \mathsf{E}_2$ and such that A maps E onto E_1 as a projection $\mathsf{x} + \mathsf{y} \mapsto \mathsf{x}$.

26.9.2 Local (nonlinear) case.

Definition 26.45 *Let $f_1 : (\mathsf{E}_1, \mathsf{p}_1) \to (\mathsf{F}_1, \mathsf{q}_1)$ be a local map. We say that f_1 is **locally equivalent** near p_1 to $f_2 : (\mathsf{E}_2, \mathsf{p}_2) \to (\mathsf{F}_2, \mathsf{q}_2)$ if there exist local diffeomorphisms $\alpha : \mathsf{E}_1 \to \mathsf{E}_2$ and $\beta : \mathsf{F}_1 \to \mathsf{F}_2$ such that $f_1 = \alpha \circ f_2 \circ \beta^{-1}$ (near p) or equivalently if $f_2 = \beta \circ f_1 \circ \alpha^{-1}$ (near p_2).*

Definition 26.46 *Let $f :: \mathsf{E}, \mathsf{p} \to \mathsf{F}, \mathsf{q}$ be a local map. We say that f is a **locally splitting injection** or **local immersion** if there are Banach spaces F_1 and F_2 with $\mathsf{F} \cong \mathsf{F}_1 \times \mathsf{F}_2$ and if f is locally equivalent near p to the injection $\mathrm{inj}_1 :: (\mathsf{F}_1, 0) \to (\mathsf{F}_1 \times \mathsf{F}_2, 0)$.*

By restricting the maps to possibly smaller open sets we can arrange that $\beta \circ f \circ \alpha^{-1}$ is given by $U' \overset{\alpha^{-1}}{\to} U \overset{f}{\to} V \overset{\beta}{\to} U' \times V'$ which we will call a **nice** local injection.

Lemma 26.8 *If f is a locally splitting injection as above there is an open set U_1 containing p and local diffeomorphism $\varphi : U_1 \subset \mathsf{F} \to U_2 \subset \mathsf{E} \times \mathsf{F}_2$ and such that $\varphi \circ f(\mathsf{x}) = (\mathsf{x}, 0)$ for all $\mathsf{x} \in U_1$.*

Proof. This is done using the same idea as in the proof of lemma 26.6. ■

Definition 26.47 *Let $f :: (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ be a local map. We say that f is a **locally splitting surjection** or **local submersion** if there are Banach spaces E_1 and E_2 with $\mathsf{E} \cong \mathsf{E}_1 \times \mathsf{E}_2$ and if f is locally equivalent (at p) to the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$.*

Again, by restriction of the domains to smaller open sets we can arrange that projection $\beta \circ f \circ \alpha^{-1}$ is given by $U' \times V' \overset{\alpha^{-1}}{\to} U \overset{f}{\to} V \overset{\beta}{\to} U'$ which we will call a **nice** local projection.

Lemma 26.9 *If f is a locally splitting surjection as above there are open sets $U_1 \times U_2 \subset \mathsf{F} \times \mathsf{E}_2$ and $V \subset \mathsf{F}$ together with a local diffeomorphism $\varphi : U_1 \times U_2 \subset \mathsf{F} \times \mathsf{E}_2 \to V \subset \mathsf{E}$ such that $f \circ \varphi(\mathsf{u}, \mathsf{v}) = \mathsf{u}$ for all $(\mathsf{u}, \mathsf{v}) \in U_1 \times U_2$.*

Proof. *This is the local (nonlinear) version of lemma 26.7 and is proved just as easily.* ■

Theorem 26.12 (local immersion) *Let $f :: (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ be a local map. If $Df|_p : (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ is a splitting injection then $f :: (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ is a local immersion.*

Theorem 26.13 (local immersion- finite dimensional case) *Let $f :: \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^k$ be a map of constant rank n in some neighborhood of $0 \in \mathbb{R}^n$. Then there is $g_1 :: \mathbb{R}^n \to \mathbb{R}^n$ with $g_1(0) = 0$, and a $g_2 :: \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^k$ such that $g_2 \circ f \circ g_1^{-1}$ is just given by $x \mapsto (x, 0) \in \mathbb{R}^n \times \mathbb{R}^k$.*

We have a similar but complementary theorem which we state in a slightly more informal manner.

Theorem 26.14 (local submersion) *Let $f :: (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ be a local map. If $Df|_p : (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ is a splitting surjection then $f :: (\mathsf{E}, \mathsf{p}) \to (\mathsf{F}, \mathsf{q})$ is a local submersion.*

Theorem 26.15 (local submersion -finite dimensional case) *Let $f :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n, 0)$ be a local map with constant rank n near 0. Then there are diffeomorphisms $g_1 :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n \times \mathbb{R}^k, 0)$ and $g_2 :: (\mathbb{R}^n, 0) \to (\mathbb{R}^n, 0)$ such that near 0 the map*

$$g_2 \circ f \circ g_1^{-1} :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n, 0)$$

is just the projection $(x, y) \mapsto x$.

If the reader thinks about what is meant by local immersion and local submersion he/she will realize that in each case the derivative map Df_p has full rank. That is, the rank of the Jacobian matrix in either case is a big as the dimensions of the spaces involved allow. Now rank is only a semicontinuous and this is what makes full rank extend from points out onto neighborhoods so to speak. On the other hand, we can get more general maps into the picture if we explicitly assume that the rank is locally constant.

26.10 The Tangent Bundle of an Open Subset of a Banach Space

Later on we will define the notion of a tangent space and tangent bundle for a differentiable manifold which locally looks like a Banach space. Here we give a definition that applies to the case of an open set U in a Banach space.

Definition 26.48 *Let E be a Banach space and $U \subset \mathsf{E}$ an open subset. A* ***tangent vector*** *at $\mathsf{x} \in U$ is a pair (x, v) where $\mathsf{v} \in \mathsf{E}$. The* ***tangent space*** *at $\mathsf{x} \in U$ is defined to be $T_{\mathsf{x}}U := T_{\mathsf{x}}\mathsf{E} := \{\mathsf{x}\} \times \mathsf{E}$ and the tangent bundle TU over U is the union of the tangent spaces and so is just $TU = U \times \mathsf{E}$. Similarly the cotangent bundle over U is defined to be $T^*U = U \times \mathsf{E}^*$. A tangent space $T_{\mathsf{x}}\mathsf{E}$ is also sometimes called the fiber at x.*

We give this definition in anticipation of our study of the tangent space at a point of a differentiable manifold. In this case however, it is often not necessary to distinguish between $T_{\mathsf{x}}U$ and E since we can often tell from context that an element $\mathsf{v} \in \mathsf{E}$ is to be interpreted as based at some point $\mathsf{x} \in U$. For instance a vector field in this setting is just a map $X : U \to \mathsf{E}$ but $X(\mathsf{x})$ should be thought of as based at x.

Definition 26.49 *If $f : U \to \mathsf{F}$ is a C^r map into a Banach space F then the* ***tangent map*** *$Tf : TU \to T\mathsf{F}$ is defined by*

$$Tf \cdot (\mathsf{x}, \mathsf{v}) = (f(\mathsf{x}), Df(\mathsf{x}) \cdot \mathsf{v}).$$

The map takes the tangent space $T_{\mathsf{x}}U = T_{\mathsf{x}}\mathsf{E}$ linearly into the tangent space $T_{f(\mathsf{x})}\mathsf{F}$ for each $\mathsf{x} \in U$. The projection onto the first factor is written $\tau_U : TU = U \times \mathsf{E} \to U$ and given by $\tau_U(\mathsf{x}, \mathsf{v}) = \mathsf{x}$. We also have a projection $\pi_U : T^*U = U \times \mathsf{E}^* \to U$ defined similarly.

If $f : U \to V$ is a diffeomorphism of open sets U and V in E and F respectively then Tf is a diffeomorphism that is linear on the fibers and such that we have a commutative diagram:

$$\begin{array}{llcl} TU = & U \times \mathsf{E} & \overset{Tf}{\to} & V \times \mathsf{F} = TV \\ pr_1 & \downarrow & & \downarrow pr_1 \\ & U & \to & V \\ & & f & \end{array}$$

The pair is an example of what is called a **local bundle map**. In this context we will denote the projection map $TU = U \times \mathsf{E} \to U$ by τ_U.

The chain rule looks much better if we use the tangent map:

Theorem 26.16 *Let U_1 and U_2 be open subsets of Banach spaces E_1 and E_2 respectively. Suppose we have differentiable (resp. $C^r, r \geq 1$) maps composing as*

$$U_1 \overset{f}{\to} U_2 \overset{g}{\to} \mathsf{E}_3$$

where E_3 is a third Banach space. Then the composition is $g \circ f$ differentiable (resp. $C^r, r \geq 1$) and $T(g \circ f) = Tg \circ Tf$

$$\begin{array}{lclcl} TU_1 & \overset{Tf}{\to} & TU_2 & \overset{Tg}{\to} & T\mathsf{E}_3 \\ \tau_{U_1} \downarrow & & \tau_{U_2} \downarrow & & \downarrow \tau_{E_3} \\ U_1 & \overset{f}{\to} & U_2 & \overset{g}{\to} & \mathsf{E}_3 \end{array}$$

Notation 26.4 *(and convention) There are three ways to express the "differential/derivative" of a differentiable map $f : U \subset \mathsf{E} \to \mathsf{F}$.*

1. *The first is just $Df : \mathsf{E} \to \mathsf{F}$ or more precisely $Df|_{\mathsf{x}} : \mathsf{E} \to \mathsf{F}$ for any point $\mathsf{x} \in U$.*
2. *This one is new for us. It is common but not completely standard :*

   $$dF : TU \to \mathsf{F}$$

   This is just the map $(\mathsf{x}, \mathsf{v}) \to Df|_{\mathsf{x}} \mathsf{v}$. We will use this notation also in the setting of maps from manifolds into vector spaces where there is a canonical trivialization of the tangent bundle of the target manifold (all of these terms will be defined). The most overused symbol for various "differentials" is d. We will use this in connection with Lie group also.
3. *Lastly the tangent map $Tf : TU \to T\mathsf{F}$ which we defined above. This is the one that generalizes to manifolds without problems.*

   In the local setting that we are studying now these three all contain essentially the same information so the choice to use one over the other is merely aesthetic.

It should be noted that some authors use *df* to mean any of the above maps and their counterparts in the general manifold setting. This leads to less confusion than one might think since one always has context on one's side.

26.11 Problem Set

Solution set 1

1. Find the matrix that represents the derivative the map $f : \mathbb{R}^n \to \mathbb{R}^m$ given by

   a) $f(x) = Ax$ for an $m \times n$ matrix A.

   b) $f(x) = x^t Ax$ for an $n \times n$ matrix A (here $m = 1$)

   c) $f(x) = x^1 x^2 \cdots x^n$ (here $m = 1$)
2. Find the derivative of the map $F : L^2([0,1]) \to L^2([0,1])$ given by

   $$F[f](x) = \int_0^1 k(x,y)f(y)dy$$

   where $k(x,y)$ is a bounded continuous function on $[0,1] \times [0,1]$. **Hint:** F is linear!
3. Let $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be C^∞ and define

   $$S[c] = \int_0^1 L(c(t), c'(t), t)dt$$

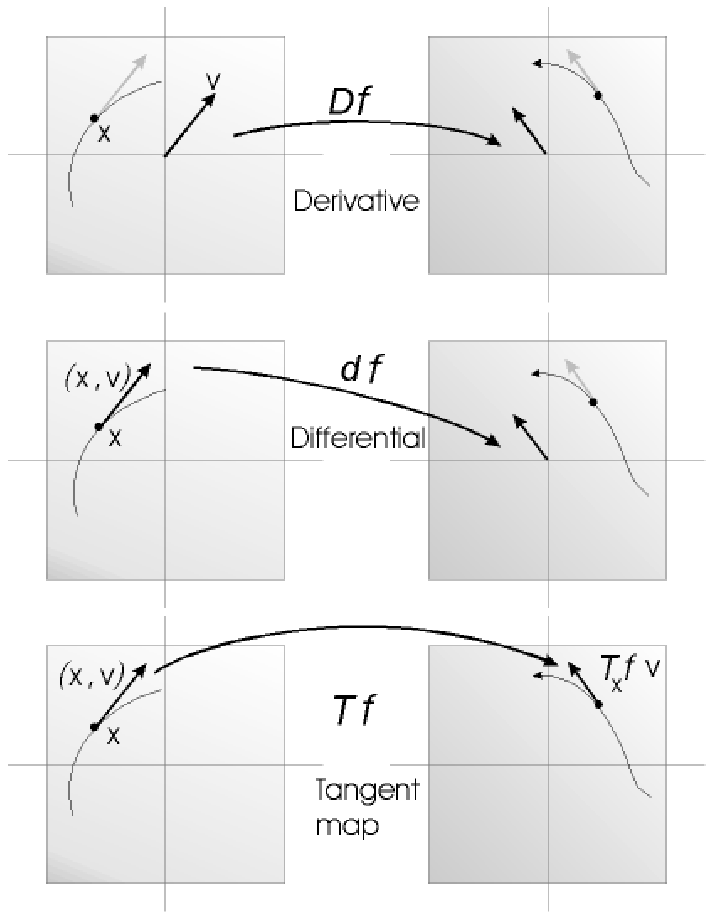


Figure 26.1: Versions of the “derivative” map.

which is defined on the Banach space B of all C^1 curves $c : [0,1] \to R^n$ with $c(0) = 0$ and $c(1) = 0$ and with the norm $\|c\| = \sup_{t\in[0,1]}\{|c(t)| + |c'(t)|\}$. Find a function $g_c : [0,1] \to R^n$ such that

$$DS|_c \cdot b = \int_0^1 \langle g_c(t), b(t)\rangle dt$$

4. In the last problem, if we hadn't insisted that $c(0) = 0$ and $c(1) = 0$, but rather that $c(0) = x_0$ and $c(1) = x_1$, then the space wouldn't even have been a vector space let alone a Banach space. But this fixed endpoint family of curves is exactly what is usually considered for functionals of this type. Anyway, convince yourself that this is not a serious problem by using the notion of an affine space (like a vector space but no origin and only differences are defined. (look it up)). Is the tangent space of the this space of fixed endpoint curves a Banach space? (hint: yep!)

   Hint: If we choose a fixed curve c_0 which is the point in the Banach space at which we wish to take the derivative then we can write $\mathcal{B}_{\vec{x}_0\vec{x}_1} = \mathcal{B} + c_0$ where

$$\begin{aligned} \mathcal{B}_{\vec{x}_0\vec{x}_1} &= \{c : c(0) = \vec{x}_0 \text{ and } c(1) = \vec{x}_1\} \\ \mathcal{B} &= \{c : c(0) = 0 \text{ and } c(1) = 0\} \end{aligned}$$

   Then we have $T_{c_0}\mathcal{B}_{\vec{x}_0\vec{x}_1} \cong \mathcal{B}$. Thus we should consider $DS|_{c_0} : \mathcal{B} \to \mathcal{B}$.

5. Let $Gl(n,\mathbb{R})$ be the nonsingular $n \times n$ matrices and show that $Gl(n,\mathbb{R})$ is an open subset of the vector space of all matrices $M_{n\times n}(\mathbb{R})$ and then find the derivative of the determinant map: $\det : Gl(n,\mathbb{R}) \to \mathbb{R}$ (for each A this should end up being a linear map $D\det|_A : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$).

   What is $\frac{\partial}{\partial x_{ij}} \det X$ where $X = (x_{ij})$?

6. Let $A : U \subset \mathsf{E} \to L(\mathsf{F},\mathsf{F})$ be a C^r map and define $F : U \times \mathsf{F} \to \mathsf{F}$ by $F(u,f) := A(u)f$. Show that F is also C^r.

   Hint: Leibnitz rule theorem.

26.11.1 Existence and uniqueness for differential equations

Theorem 26.17 *Let E be a Banach space and let $X : U \subset E \to E$ be a smooth map. Given any $x_0 \in U$ there is a smooth curve $c : (-\epsilon,\epsilon) \to U$ with $c(0) = x_0$ such that $c'(t) = X(c(t))$ for all $t \in (-\epsilon,\epsilon)$. If $c_1 : (-\epsilon_1,\epsilon_1) \to U$ is another such curve with $c_1(0) = x_0$ and $c_1'(t) = X(c(t))$ for all $t \in (-\epsilon_1,\epsilon_1)$ then $c = c_1$ on the intersection $(-\epsilon_1,\epsilon_1) \cap (-\epsilon,\epsilon)$. Furthermore, there is an open set V with $x_0 \in V \subset U$ and a smooth map $\Phi : V \times (-a,a) \to U$ such that $t \mapsto c_x(t) := \Phi(x,t)$ is a curve satisfying $c'(t) = X(c(t))$ for all $t \in (-a,a)$.*

26.11.2 Differential equations depending on a parameter.

Theorem 26.18 *Let J be an open interval on the real line containing 0 and suppose that for some Banach spaces E and F we have a smooth map $F : J \times U \times V \to \mathsf{F}$ where $U \subset \mathsf{E}$ and $V \subset \mathsf{F}$. Given any fixed point $(x_0, y_0) \in U \times V$ there exist a subinterval $J_0 \subset J$ containing 0 and open balls $B_1 \subset U$ and $B_2 \subset V$ with $(x_0, y_0) \in B_1 \times B_2$ and a unique smooth map*

$$\beta : J_0 \times B_1 \times B_2 \to V$$

such that
1) $\frac{d}{dt}\beta(t, x, y) = F(t, x, \beta(t, x, y))$ for all $(t, x, y) \in J_0 \times B_1 \times B_2$ and
2) $\beta(0, x, y) = y$.
Furthermore,
3) if we let $\beta(t, x) := \beta(t, x, y)$ for fixed y then

$$\begin{aligned}\frac{d}{dt} D_2\beta(t, x) \cdot v &= D_2F(t, x, \beta(t, x)) \cdot v \\ &\quad + D_3F(t, x, \beta(t, x)) \cdot D_2\beta(t, x) \cdot v\end{aligned}$$

for all $v \in \mathsf{E}$.

26.12 Multilinear Algebra

There a just a few things from multilinear algebra that are most important for differential geometry. Multilinear spaces and operations may be defined starting with in the category vector spaces and linear maps but we are also interested in vector bundles and their sections. At this point we may consider the algebraic structure possessed by the sections of the bundle. What we have then is not (only) a vector space but (also) a *module* over the *ring* of smooth functions. The algebraic operation we perform on this level are very similar to vector space calculations but instead of the scalars being the real (or complex) numbers the scalars are functions. So even though we could just start defining things for vector spaces it will be more efficient to consider modules over a commutative ring R since vector spaces are also modules whose scalar ring just happens to be a field. We remind the reader that unlike algebraic fields like $\mathbb{R}$ or $\mathbb{C}$ there may be elements a and b in a ring such that $ab = 0$ and yet neither a nor b is zero. The ring we mostly interested in is the ring of smooth functions on a manifold.

The set of all vector fields $\mathfrak{X}(M)$ on a manifold is a vector space over the real numbers. But $\mathfrak{X}(M)$ has more structure since not only can we add vector fields and scale by numbers but we may also scale by smooth functions. We say that the vector fields form a module over the ring of smooth functions. A module is similar to a vector space with the differences stemming from the use of elements of a ring R of the scalars rather than a field (such as the complex $\mathbb{C}$ or real numbers $\mathbb{R}$). For a module, one still has $1w = w$, $0w = 0$ and $-1w = -w$. Of course, every vector space is also a module since the latter is a generalization of the notion of vector space.

Definition 26.50 *Let R be a ring. A left R-**module** (or a left **module** over R) is an abelian group $W, +$ together with an operation $\mathsf{R} \times W \to W$ written $(a, w) \mapsto aw$ such that*

1) $(a + b)w = aw + bw$ for all $a, b \in \mathsf{R}$ and all $w \in W$,

2) $a(w_1 + w_2) = aw_1 + aw_2$ for all $a \in \mathsf{R}$ and all $w_2, w_1 \in W$.

*A right R-**module** is defined similarly with the multiplication of the right so that*

1) $w(a + b) = wa + wb$ for all $a, b \in \mathsf{R}$ and all $w \in W$,

2) $(w_1 + w_2)a = w_1a + w_2a$ for all $a \in \mathsf{R}$ and all $w_2, w_1 \in W$. .

If the ring is commutative (the usual case for us) then we may write $aw = wa$ and consider any right module as a left module and visa versa. Even if the ring is not commutative we will usually stick to left modules and so we drop the reference to "left" and refer to such as R-modules.

Remark 26.10 *We shall often refer to the elements of R as **scalars**.*

Example 26.10 *An abelian group $A, +$ is a $\mathbb{Z}$ module and a $\mathbb{Z}$-module is none other than an abelian group. Here we take the product of $n \in \mathbb{Z}$ with $x \in A$ to be $nx := x + \cdots + x$ if $n \geq 0$ and $nx := -(x + \cdots + x)$ if $n < 0$ (in either case we are adding $|n|$ terms).*

Example 26.11 *The set of all $m \times n$ matrices with entries being elements of a commutative ring R (for example real polynomials) is an R-module.*

Example 26.12 *The set of all module homomorphisms of a module W onto another module M is a module and is denoted $\mathrm{Hom}_{C^\infty(M)}(\mathrm{W}, \mathrm{M})$ or $L_{C^\infty(M)}(\mathrm{W}, \mathrm{M})$. In particular, we have $\mathrm{Hom}_{\mathsf{R}}(\Gamma(\xi_1),\Gamma(\xi_2))$ where ξ_1 and ξ_1 are vector bundles over a smooth manifold M.*

Example 26.13 *Let V be a vector space and $\ell : \mathrm{V} \to \mathrm{V}$ a linear operator. Using this one operator we may consider V as a module over the ring of polynomials $\mathbb{R}[t]$ by defining the "scalar" multiplication by*

$$p(t)v := p(\ell)v$$

for $p \in \mathbb{R}[t]$. For example, $(1 + t^2)v = v + \ell^2 v$.

Since the ring is usually fixed we often omit mention of the ring. In particular, we often abbreviate $L_{\mathsf{R}}(\mathrm{W}, \mathrm{M})$ to $L(\mathrm{W}, \mathrm{M})$. Similar omissions will be made without further mention. Also, since every real (resp. complex) Banach space E is a vector space and hence a module over $\mathbb{R}$ (resp. $\mathbb{C}$) we must distinguish between the bounded linear maps which we have denoted up until now as $L(\mathsf{E}; \mathsf{F})$ and the linear maps which would be denoted the same way in the context of modules. Our convention will be the following:

In case the modules in question are presented as infinite dimensional topological vector spaces, say E and F, we will let $L(\mathsf{E}; \mathsf{F})$ continue to mean the space of bounded linear operator unless otherwise stated.

A **submodule** is defined in the obvious way as a subset $S \subset \mathrm{W}$ which is closed under the operations inherited from W so that S itself is a module. The intersection of all submodules containing a subset $A \subset \mathrm{W}$ is called the **submodule generated by** A and is denoted $\langle A \rangle$ and A is called a **generating set**. If $\langle A \rangle = \mathrm{W}$ for a finite set A, then we say that W is finitely generated.

Let S be a submodule of W and consider the quotient abelian group W/S consisting of cosets, that is sets of the form $[v] := v + S = \{v + x : x \in S\}$ with addition given by $[v] + [w] = [v + w]$. We define a scalar multiplication by elements of the ring R by $a[v] := [av]$ respectively. In this way, W/S is a module called a quotient module.

Definition 26.51 *Let W_1 and W_2 be modules over a ring R. A map $L : W_1 \to W_2$ is called **module homomorphism** if*

$$L(aw_1 + bw_2) = aL(w_1) + bL(w_2).$$

*By analogy with the case of vector spaces, which module theory includes, we often characterize a module homomorphism L by saying that L is **linear over** R.*

A real (resp. complex) vector space is none other than a module over the field of real numbers $\mathbb{R}$ (resp. complex numbers $\mathbb{C}$). In fact, most of the modules we encounter will be either vector spaces or spaces of sections of some vector bundle.

Many of the operations that exist for vector spaces have analogues in the module category. For example, the direct sum of modules is defined in the obvious way. Also, for any module homomorphism $L : \mathrm{W}_1 \to \mathrm{W}_2$ we have the usual notions of kernel and image:

$$\begin{aligned} \ker L &= \{v \in \mathrm{W}_1 : L(v) = 0\} \\ \operatorname{img}(L) = L(\mathrm{W}_1) &= \{w \in \mathrm{W}_2 : w = Lv \text{ for some } v \in \mathrm{W}_1\}. \end{aligned}$$

These are submodules of W_1 and W_2 respectively.

On the other hand, modules are generally not as simple to study as vector spaces. For example, there are several notions of dimension. The following notions for a vector space all lead to the same notion of dimension. For a completely general module these are all potentially different notions:

1. Length of the longest chain of submodules

$$0 = \mathrm{W}_n \subsetneq \cdots \subsetneq \mathrm{W}_1 \subsetneq \mathrm{W}$$

2. The cardinality of the largest linearly independent set (see below).

3. The cardinality of a basis (see below).

For simplicity in our study of dimension, let us now assume that R is commutative.

Definition 26.52 *A set of elements $w_1, ..., w_k$ of a module are said to be* ***linearly dependent*** *if there exist ring elements $r_1, ..., r_k \in \mathsf{R}$ not all zero, such that $r_1 w_1 + \cdots + r_k w_k = 0$. Otherwise, they are said to be* ***linearly independent****. We also speak of the set $\{w_1, ..., w_k\}$ as being a* ***linearly independent set****.*

So far so good but it is important to realize that just because $w_1, ..., w_k$ are linearly independent doesn't mean that we may write each of these w_i as a linear combination of the others. It may even be that some element w forms a linearly dependent set since there may be a nonzero r such that $rw = 0$ (such a w is said to have **torsion**).

If a linearly independent set $\{w_1, ..., w_k\}$ is maximal in size then we say that the module has **rank** k. Another strange possibility is that a maximal linearly independent set may not be a generating set for the module and hence may not be a basis in the sense to be defined below. The point is that although for an arbitrary $w \in \mathrm{W}$ we must have that $\{w_1, ..., w_k\} \cup \{w\}$ is linearly dependent and hence there must be a nontrivial expression $rw + r_1 w_1 + \cdots + r_k w_k = 0$, it does **not** follow that we may solve for w since r may not be an invertible element of the ring (i.e. it may not be a **unit**).

Definition 26.53 *If B is a generating set for a module* W *such that every element of* W *has a unique expression as a* finite *R-linear combination of elements of B then we say that B is a* ***basis*** *for* W*.*

Definition 26.54 *If an R-module has a basis then it is referred to as a* ***free module****.*

If a module over a (commutative) ring R has a basis then the number of elements in the basis is called the **dimension** and must in this case be the same as the rank (the size of a maximal linearly independent set). If a module W is free with basis $w_1, ..., w_n$ then we have an isomorphism $\mathsf{R}^n \cong \mathrm{W}$ given by

$$(r_1, ..., r_n) \mapsto r_1 w_1 + \cdots + r_n w_n.$$

Exercise 26.6 *Show that every finitely generated R-module is the homomorphic image of a free module.*

It turns out that just as for vector spaces the cardinality of a basis for a free module W over a commutative ring is the same as that of every other basis for W. The cardinality of any basis for a free module W is called the dimension of W. If R is a field then every module is free and is a vector space by definition. In this case, the current definitions of dimension and basis coincide with the usual ones.

The ring R is itself a free R-module with standard basis given by $\{1\}$. Also, $\mathsf{R}^n := \mathsf{R} \times \cdots \times \mathsf{R}$ is a free module with standard basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ where, as usual $\mathbf{e}_i := (0, ..., 1, ..., 0)$; the only nonzero entry being in the i-th position.

Definition 26.55 *Let* k *be a commutative ring, for example a field such as* $\mathbb{R}$ *or* $\mathbb{C}$*. A ring* A *is called a* k***-algebra*** *if there is a ring homomorphism* $\mu : \mathsf{k} \rightarrow R$ *such that the image* $\mu(\mathsf{k})$ *consists of elements which commute with everything in* A*. In particular,* A *is a module over* k*.*

Example 26.14 *The ring* $\mathcal{C}^\infty_M(U)$ *is an* $\mathbb{R}$*-algebra.*

We shall have occasion to consider A-modules where A is an algebra over some k. In this context the elements of A are still called scalars but the elements of $\mathsf{k} \subset \mathsf{A}$ will be referred to as **constants**.

Example 26.15 *For an open set* $U \subset M$ *the set vector fields* $\mathfrak{X}_M(U)$ *is a vector space over* $\mathbb{R}$ *but it is also a module over the* $\mathbb{R}$*-algebra* $\mathcal{C}^\infty_M(U)$*. So for all* $X, Y \in \mathfrak{X}_M(U)$ *and all* $f, g \in \mathcal{C}^\infty_M(U)$ *we have*

1. $f(X + Y) = fX + fY$
2. $(f + g)X = fX + gX$
3. $f(gX) = (fg)X$

Similarly, $\mathfrak{X}^*_M(U) = \Gamma(U, T^*M)$ is also a module over $\mathcal{C}^\infty_M(U)$ which is naturally identified with the module dual $\mathfrak{X}_M(U)^*$ by the pairing $(\theta, X) \mapsto \theta(X)$. Here $\theta(X) \in \mathcal{C}^\infty_M(U)$ and is defined by $p \mapsto \theta_p(X_p)$. The set of all vector fields $\mathcal{Z} \subset \mathfrak{X}(U)$ which are zero at a fixed point $p \in U$ is a submodule in $\mathfrak{X}(U)$. If $U, (x^1, ..., x^n)$ is a coordinate chart then the set of vector fields

$$\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$$

is a basis (over $\mathcal{C}^\infty_M(U)$) for the module $\mathfrak{X}(U)$. Similarly,

$$dx^1, ..., dx^n$$

is a basis for $\mathfrak{X}^*_M(U)$. It is important to realize that if U is not a coordinate chart domain then it may be that $\mathfrak{X}_M(U)$ and $\mathfrak{X}_M(U)^*$ have no basis. In particular, we should not expect $\mathfrak{X}(M)$ to have a basis in the general case.

Example 26.16 *The sections of any vector bundle over a manifold* M *form a* $C^\infty(M)$*-module denoted* $\Gamma(E)$*. Let* $E \to M$ *be a trivial vector bundle of finite rank* n*. Then there exists a basis of vector fields* $E_1, ..., E_n$ *for the module* $\Gamma(E)$*. Thus for any section* X *there exist unique functions* f^i *such that*

$$X = \sum f^i E_i.$$

In fact, since E *is trivial we may as well assume that* $E = M \times \mathbb{R}^n \overset{pr_1}{\to} M$*. Then for any basis* $e_1, ..., e_n$ *for* $\mathbb{R}^n$ *we may take*

$$E_i(x) := (x, e_i).$$

In this section we intend all modules to be treated strictly as modules. Thus we do not require multilinear maps to be bounded. In particular, $L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$ does not refer to bounded multilinear maps even if the modules are coincidentally Banach spaces. We shall comment on how thing look in the Banach space category in a later section.

Definition 26.56 *The dual of an* $\mathsf{R}-$*module* W *is the module* $\mathrm{W}^* := \mathrm{Hom}_{\mathsf{R}}(\mathrm{W},\mathsf{R})$ *of all* R*-linear functionals on* W.

Definition 26.57 *If the map* $\widehat{()} : \mathrm{W} \to \mathrm{W}^{**} := \mathrm{Hom}_{\mathsf{R}}(\mathrm{W}^*,\mathsf{R})$ *given by*

$$\widehat{w}(\alpha) := \alpha(w)$$

is an isomorphism then we say that W *is* ***reflexive*** . *If* W *is reflexive then we are free to identify* W *with* W^{**} *and consider any* $w \in \mathrm{W}$ *as map* $w = \widehat{w} : \mathrm{W}^* \to \mathsf{R}$.

Exercise 26.7 *Show that if* W *is a free with finite dimension then* W *is reflexive. We sometimes write* $\widehat{w}(\alpha) = \langle w, \alpha \rangle = \langle \alpha, w \rangle$.

There is a bit of uncertainty about how to use the word "tensor". On the one hand, a tensor is a certain kind of multilinear mapping. On the other hand, a tensor is an element of a tensor product (defined below) of several copies of a module and its dual. For finite dimensional vector spaces these two viewpoints turn out to be equivalent as we shall see but since we are also interested in infinite dimensional spaces we must make a terminological distinction. We make the following slightly nonstandard definition:

Definition 26.58 *Let* V *and* W *be* R*-modules. A* W*-valued* $({}^{r}{}_{s})$*-**tensor map** on* V *is a multilinear mapping of the form*

$$\Lambda : \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \times \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \to \mathrm{W}.$$

The set of all tensor maps into W will be denoted $T^{r}{}_{s}(\mathrm{V}; \mathrm{W})$. Similarly, a W-valued $({}_{s}{}^{r})$-**tensor map** on V is a multilinear mapping of the form

$$\Lambda : \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \times \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \to \mathrm{W}$$

and the corresponding space of all such is denoted $T_{s}{}^{r}(\mathrm{V}; \mathrm{W})$.

There is, of course, a natural isomorphism $T_{s}{}^{r}(\mathrm{V}; \mathrm{W}) \cong T^{r}{}_{s}(\mathrm{V}; \mathrm{W})$ induced by the map $\mathrm{V}^s \times \mathrm{V}^{*r} \cong \mathrm{V}^{*r} \times \mathrm{V}^s$ given on homogeneous elements by $v \otimes \omega \mapsto \omega \otimes v$. (Warning) In the presence of an inner product there is another possible isomorphism here given by $v \otimes \omega \mapsto \flat v \otimes \sharp\omega$ **This map is a "transpose" map and just as we do not identify a matrix with its transpose we do not generally identify individual elements under this isomorphism.**

Notation 26.5 *For the most part we shall be needing only $T^{r}{}_{s}(\mathrm{V};\mathrm{W})$ and so we agree abbreviate this to $T^r_s(\mathrm{V};\mathrm{W})$ and call the elements (r,s)-**tensor maps**. So by convention*

$$T^r_s(\mathrm{V};\mathrm{W}) := T^{r}{}_{s}(\mathrm{V};\mathrm{W})$$
$$\text{but}$$
$$T^r_s(\mathrm{V};\mathrm{W}) \neq T_{r}{}^{s}(\mathrm{V};\mathrm{W}).$$

Elements of $T^r_0(\mathrm{V})$ are said to be of **contravariant type** and of degree r and those in $T^0_s(\mathrm{V})$ are of **covariant type** (and degree s). If $r,s>0$ then the elements of $T^r_s(\mathrm{V})$ are called **mixed tensors (of tensors of mixed type)** with contravariant degree r and covariant degree s.

Remark 26.11 *An $\mathbb{R}$ -valued (r,s)-**tensor map** is usually just called an (r,s)-**tensor** but as we mentioned above, there is another common meaning for this term which is equivalent in the case of finite dimensional vector spaces. The word* tensor *is also used to mean "tensor field" (defined below). The context will determine the proper meaning.*

Remark 26.12 *The space $T^r_s(\mathrm{V};\mathsf{R})$ is sometimes denoted by $T^r_s(\mathrm{V};\mathsf{R})$ (or even $T^r_s(\mathrm{V})$ in case $\mathsf{R}=\mathbb{R}$) but we reserve this notation for another space defined below which is canonically isomorphic to $T^r_s(\mathrm{V};\mathsf{R})$ in case V is free with finite dimension.*

Definition 26.59 *Given $\Upsilon_1 \in T^r_s(\mathrm{V};\mathsf{R})$ and $\Upsilon_2 \in T^p_q(\mathrm{V};\mathsf{R})$ we may define the (consolidated) tensor product $\Upsilon_1 \otimes \Upsilon_2 \in T^{r+p}_{s+q}(\mathrm{V};\mathsf{R})$ by*

$$\begin{aligned}&(\Upsilon_1 \otimes \Upsilon_2)(\alpha_1,...,\alpha_r,v_1,...,v_s)\\ &:= \Upsilon_1(\alpha_1,...,\alpha_r)\Upsilon_2(v_1,...,v_s)\end{aligned}$$

Remark 26.13 *We call this type of tensor product the **"map" tensor product** in case we need to distinguish it from the tensor product defined below.*

Now suppose that V is free with finite dimension n. Then there is a basis $f_1,...,f_n$ for V with dual basis $f^1,...,f^n$. Now we have $\mathrm{V}^* = T^0_1\ (\mathrm{V};\mathsf{R})$. Also, we may consider $f_i \in \mathrm{V}^{**} = T^0_1\ (\mathrm{V}^*;\mathsf{R})$ and then, as above, take tensor products to get elements of the form

$$f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f^{j_1} \otimes \cdots \otimes f^{j_s}.$$

These are multilinear maps in $T^r_s(\mathrm{V};\mathsf{R})$ by definition:

$$\begin{aligned}&(f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f^{j_1} \otimes \cdots \otimes f^{j_s})(\alpha_1,...,\alpha_r,v_1,...,v_s)\\ &= \alpha_1(f_{i_1})\cdots\alpha_r(f_{i_r})f^{j_1}(v_1)\cdots f^{j_s}(v_s).\end{aligned}$$

There are $n^s n^r$ such maps which form a basis for $T^r_s(\mathrm{V};\mathsf{R})$ which is therefore also free. Thus we may write any tensor map $\Upsilon \in T^r_s(\mathrm{V};\mathsf{R})$ as a sum

$$\Upsilon = \sum \Upsilon^{i_1...i_r}{}_{j_1,...j_s}\ f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f^{j_1} \otimes \cdots \otimes f^{j_s}$$

and the scalars $\Upsilon^{i_1 \ldots i_r}{}_{j_1,\ldots j_s} \in \mathsf{R}$

We shall be particularly interested in the case where all the modules are real (or complex) vector spaces such as the tangent space at a point on a smooth manifold. As we mention above, we will define a space $T^r_s(\mathrm{V};\mathsf{R})$ for each (r,s) which is canonically isomorphic to $T^r_s(\mathrm{V};\mathsf{R})$. There will be a product $\otimes T^r_s(\mathrm{V};\mathsf{R}) \times T^p_q(\mathrm{V};\mathsf{R}) \to T^{r+p}_{s+q}(\mathrm{V};\mathsf{R})$ for these spaces also and this will match up with the current definition under the canonical isomorphism.

Example 26.17 *The inner product (or "dot product") on the Euclidean vector space $\mathbb{R}^n$ given for vectors $\tilde{\mathrm{v}} = (v_1, ..., v_n)$ and $\tilde{\mathrm{w}} = (w_1, ..., w_n)$ by*

$$(\vec{v}, \vec{w}) \mapsto \langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i w_i$$

*is 2-multilinear (more commonly called **bilinear**).*

Example 26.18 *For any n vectors $\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n \in \mathbb{R}^n$ the determinant of $(\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n)$ is defined by considering $\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n$ as columns and taking the determinant of the resulting $n \times n$ matrix. The resulting function is denoted $\det(\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n)$ and is n-multilinear.*

Example 26.19 *Let $\mathfrak{X}(M)$ be the $C^\infty(M)$-module of vector fields on a manifold M and let $\mathfrak{X}^*(M)$ be the $C^\infty(M)$-module of 1-forms on M. The map $\mathfrak{X}^*(M) \times \mathfrak{X}(M) \to C^\infty(M)$ given by $(\alpha, X) \mapsto \alpha(X) \in C^\infty(M)$ is clearly multilinear (bilinear) over $C^\infty(M)$.*

Suppose now that we have two R-modules V_1 and V_2. Let us construct a category $\mathcal{C}_{\mathrm{V}_1 \times \mathrm{V}_2}$ whose objects are bilinear maps $\mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}$ where W varies over all R-modules but V_1 and V_2 are fixed. A morphism from, say $\mu_1 : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}_1$ to $\mu_2 : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}_2$ is defined to be a map $\ell : \mathrm{W}_1 \to \mathrm{W}_2$ such that the diagram

$$\begin{tikzcd} & \mathrm{W}_1 \arrow[dd, "\ell"] \\ \mathrm{V}_1 \times \mathrm{V}_2 \arrow[ur, "\mu_1"] \arrow[dr, "\mu_2"'] & \\ & \mathrm{W}_2 \end{tikzcd}$$

commutes. Suppose that there is an R-module $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}$ together with a bilinear map $\mathrm{t} : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}$ that has the following universal property for this category: For every bilinear map $\mu : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}$ there is a unique linear map $\widetilde{\mu} : \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \to \mathrm{W}$ such that the following diagram commutes:

$$\begin{tikzcd} \mathrm{V}_1 \times \mathrm{V}_2 \arrow[r, "\mu"] \arrow[d, "\mathrm{t}"] & \mathrm{W} \\ \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \arrow[ur, "\widetilde{\mu}"'] & \end{tikzcd}$$

If such a pair $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}, \mathrm{t}$ exists with this property then it is unique up to isomorphism in $\mathcal{C}_{\mathrm{V}_1 \times \mathrm{V}_2}$. In other words, if $\widehat{\mathrm{t}} : \mathrm{V}_1 \times \mathrm{V}_2 \to \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2}$ is another object with this universal property then there is a module isomorphism $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \cong \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2}$ such that the following diagram commutes:

$$\begin{array}{ccc} & & \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \\ & \nearrow_{\mathrm{t}} & \\ \mathrm{V}_1 \times \mathrm{V}_2 & & \cong\downarrow \\ & \searrow^{\widehat{\mathrm{t}}} & \\ & & \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2} \end{array}$$

We refer to any such universal object as a tensor product of V_1 and V_2. We will construct a specific tensor product which we denote by $\mathrm{V}_1 \otimes \mathrm{V}_2$ with the corresponding map denoted by $\otimes : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{V}_1 \otimes \mathrm{V}_2$. More generally, we seek a universal object for k-multilinear maps $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$.

Definition 26.60 *A module* $\mathrm{T} = \mathrm{T}_{\mathrm{V}_1,\cdots,\mathrm{V}_k}$ *together with a multilinear map* $\mathrm{u} : V_1 \times \cdots \times V_k \to \mathrm{T}$ *is called* ***universal for k-multilinear maps on*** $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ *if for every multilinear map* $\mu : V_1 \times \cdots \times V_k \to \mathrm{W}$ *there is a unique linear map* $\widetilde{\mu} : \mathrm{T} \to \mathrm{W}$ *such that the following diagram commutes:*

$$\begin{array}{ccc} \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & \overset{\mu}{\to} & \mathrm{W} \\ \mathrm{u}\downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{T} & & \end{array},$$

i.e. we must have $\mu = \widetilde{\mu} \circ \mathrm{u}$. *If such a universal object exists it will be called a* ***tensor product*** *of* $V_1, ..., V_k$ *and the module itself* $\mathrm{T} = \mathrm{T}_{\mathrm{V}_1,\cdots,\mathrm{V}_k}$ *is also referred to as a tensor product of the modules* $V_1, ..., V_k$.

Lemma 26.10 *If two modules* $\mathrm{T}_1, \mathrm{u}_1$ *and* $\mathrm{T}_2, \mathrm{u}_2$ *are both universal for k-multilinear maps on* $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ *then there is an isomorphism* $\Phi : \mathrm{T}_1 \to \mathrm{T}_2$ *such that* $\Phi \circ \mathrm{u}_1 = \mathrm{u}_2$*;*

$$\begin{array}{ccccc} & & \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & & \\ & \mathrm{u}_1 \swarrow & & \searrow \mathrm{u}_2 & \\ \mathrm{T}_1 & & \overset{\Phi}{\to} & & \mathrm{T}_2 \end{array}.$$

Proof. By the assumption of universality, there are maps u_1 and u_2 such that $\Phi \circ \mathrm{u}_1 = \mathrm{u}_2$ and $\bar{\Phi} \circ \mathrm{u}_2 = \mathrm{u}_1$. We thus have $\bar{\Phi} \circ \Phi \circ \mathrm{u}_1 = \mathrm{u}_1 = \mathrm{id}$ and by the uniqueness part of the universality of u_1 we must have $\bar{\Phi} \circ \Phi = \mathrm{id}$ or $\bar{\Phi} = \Phi^{-1}$. ∎

We now show the existence of a tensor product. The specific tensor product of modules $\mathrm{V}_1, ..., \mathrm{V}_k$ that we construct will be denoted by $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ and the corresponding map will be denoted by

$$\otimes^k : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k.$$

We start out by introducing the notion of a free module on an arbitrary set. If S is just some set, then we may consider the set $F_{\mathsf{R}}(S)$ all finite formal linear combinations of elements of S with coefficients from R. For example, if $a, b, c \in \mathsf{R}$

and $s_1, s_2, s_3 \in S$ then $as_1 + bs_2 + cs_3$ is such a formal linear combination. In general, an element of $F_{\mathsf{R}}(S)$ will be of the form

$$\sum_{s\in S} a_s s$$

where the coefficients a_s are elements of R and all but finitely many are 0. Thus the sums involved are always finite. Addition of two such expressions and multiplication by elements of R are defined in the obvious way;

$$b\sum_{s\in S} a_s s = \sum_{s\in S} ba_s s$$
$$\sum_{s\in S} a_s s + \sum_{s\in S} b_s s = \sum_{s\in S} (a_s + b_s)s.$$

This is all just a way of speaking of functions $a() : S \to \mathsf{R}$ with finite support. It is also just a means of forcing the element of our arbitrary set to be the "basis elements" of a modules. The resulting module $F_{\mathsf{R}}(S)$ is called the **free module generated by** S. For example, the set of all formal linear combinations of the set of symbols $\{\mathbf{i}, \mathbf{j}\}$ over the real number ring, is just a 2 dimensional vector space with basis $\{\mathbf{i}, \mathbf{j}\}$.

Let $\mathrm{V}_1, \cdots, \mathrm{V}_k$ be modules over R and let $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ denote the set of all formal linear combinations of elements of the Cartesian product $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$. For example

$$3(v_1, w) - 2(v_2, w) \in \mathrm{F}_{\mathsf{R}}(\mathrm{V}, \mathrm{W})$$

but it is **not true** that $3(v_1, w) - 2(v_2, w) = 3(v_1 - 2v_2, w)$ since (v_1, w), (v_2, w) and $(v_1 - 2v_2, w)$ are linearly independent by definition. We now define a submodule of $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$. Consider the set B of all elements of $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ which have one of the following two forms:

1. $$(\mathrm{v}_1, ..., a\mathrm{v}_i, \cdots, \mathrm{v}_k) - a(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k)$$
   for some $a \in \mathsf{R}$ and some $1 \le i \le k$ and some $(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) \in \mathrm{V}_1 \times \cdots \times \mathrm{V}_k$.

2. $$(\mathrm{v}_1, ..., \mathrm{v}_i + \mathrm{w}_i, \cdots, \mathrm{v}_k) - (\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) - (\mathrm{v}_1, ..., \mathrm{w}_i, \cdots, \mathrm{v}_k)$$
   for some $1 \le i \le k$ and some choice of $(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) \in \mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ and $\mathrm{w}_i \in \mathrm{V}_i$.

We now define $\langle B\rangle$ to be the submodule generated by B and then define the tensor product $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ of the spaces $\mathrm{V}_1, \cdots, \mathrm{V}_k$ to be the quotient module $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)/\langle B\rangle$. Let

$$\pi : \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k) \to \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)/\langle B\rangle$$

be the quotient map and define $\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k$ to be the image of $(\mathrm{v}_1, \cdots, \mathrm{v}_k) \in \mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ under this quotient map. The quotient is the tensor space we were looking for

$$\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k := \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)/\langle B \rangle.$$

To get our universal object we need to define the corresponding map. The map we need is just the composition

$$\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \hookrightarrow \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k) \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k.$$

We denote this map by $\otimes^k : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$. Notice that $\otimes^k(\mathrm{v}_1, \cdots, \mathrm{v}_k) = \mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k$. By construction, we have the following facts:

1. $$\mathrm{v}_1 \otimes \cdots \otimes a\mathrm{v}_i \otimes \cdots \otimes \mathrm{v}_k = a\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_i \otimes \cdots \otimes \mathrm{v}_k$$

   for any $a \in \mathsf{R}$, any $i \in \{1, 2, ..., k\}$ and any $(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) \in \mathrm{V}_1 \times \cdots \times \mathrm{V}_k$.

2. $$\begin{aligned} &\mathrm{v}_1 \otimes \cdots \otimes (\mathrm{v}_i + \mathrm{w}_i) \otimes \cdots \otimes \mathrm{v}_k \\ &= \mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_i \otimes \cdots \otimes \mathrm{v}_k \ + \ \mathrm{v}_1 \otimes \cdots \otimes \mathrm{w}_i \otimes \cdots \otimes \mathrm{v}_k \end{aligned}$$

   any $i \in \{1, 2, ..., k\}$ and for all choices of $\mathrm{v}_1, ..., \mathrm{v}_k$ and w_i.

   Thus $\otimes^k$ is multilinear.

Definition 26.61 *The elements in the image of π, that is, elements which may be written as $\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k$ for some v_i, are called **decomposable** .*

Exercise 26.8 *Not all elements are decomposable but the decomposable elements generate $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.*

It may be that the V_i may be modules over more that one ring. For example, any complex vector space is a module over both $\mathbb{R}$ and $\mathbb{C}$. Also, the module of smooth vector fields $\mathfrak{X}_M(U)$ is a module over $C^\infty(U)$ and a module (actually a vector space) over $\mathbb{R}$. Thus it is sometimes important to indicate the ring involved and so we write the tensor product of two R-modules V and W as $\mathrm{V} \otimes_{\mathsf{R}} \mathrm{W}$. For instance, there is a big difference between $\mathfrak{X}_M(U) \otimes_{C^\infty(U)} \mathfrak{X}_M(U)$ and $\mathfrak{X}_M(U) \otimes_{\mathbb{R}} \mathfrak{X}_M(U)$.

Now let $\otimes^k : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ be natural map defined above which is the composition of the set injection $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \hookrightarrow \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ and the quotient map $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k) \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$. We have seen that this map actually turns out to be a multilinear map.

Theorem 26.19 *Given modules* $\mathrm{V}_1, ..., \mathrm{V}_k$*, the space* $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ *together with the map* $\otimes^k$*has the following universal property:*

For any k*-multilinear map* $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to W$*, there is a unique linear map* $\widetilde{\mu} : \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \to W$ *called the universal map, such that the following diagram commutes:*

$$\begin{array}{ccc} \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & \xrightarrow{\mu} & W \\ \otimes^k \downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k & & \end{array}.$$

Thus the pair $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k, \otimes^k$ is universal for k-multilinear maps on $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ and by **??** if T, u is any other universal pair for k-multilinear map we have $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \cong \mathrm{T}$. The module $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ is called *the* **tensor product** of $\mathrm{V}_1, ..., \mathrm{V}_k$.

Proof. Suppose that $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to W$ is multilinear. Since, $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ is free there is a unique linear map $M : \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k) \to W$ such that $M((\mathrm{v}_1, ..., \mathrm{v}_k)) = \mu(\mathrm{v}_1, ..., \mathrm{v}_k)$. Clearly, this map is zero on $\langle B \rangle$ and so there is a factorization $\widetilde{\mu}$ of M through $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$. Thus we always have

$$\widetilde{\mu}(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k) = M((\mathrm{v}_1, ..., \mathrm{v}_k)) = \mu(\mathrm{v}_1, ..., \mathrm{v}_k).$$

It is easy to see that $\widetilde{\mu}$ is unique since a linear map is determined by its action on generators (the decomposable elements generate $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$) ■

Lemma 26.11 *There are the following natural isomorphisms:*

1) $(\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U} \cong \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U}) \cong \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U})$ *and under these isomorphisms* $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u} \longleftrightarrow \mathrm{v} \otimes (\mathrm{w} \otimes \mathrm{u}) \longleftrightarrow \mathrm{v} \otimes \mathrm{w} \otimes \mathrm{u}$.

2) $\mathrm{V} \otimes \mathrm{W} \cong \mathrm{W} \otimes \mathrm{V}$ *and under this isomorphism* $\mathrm{v} \otimes \mathrm{w} \longleftrightarrow \mathrm{w} \otimes \mathrm{v}$*.*

Proof. We prove (1) and leave (2) as an exercise.

Elements of the form $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u}$ generate $(\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U}$ so any map that sends $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u}$ to $\mathrm{v} \otimes (\mathrm{w} \otimes \mathrm{u})$ for all $\mathrm{v}, \mathrm{w}, \mathrm{u}$ must be unique. Now we have compositions

$$(\mathrm{V} \times \mathrm{W}) \times \mathrm{U} \overset{\otimes \times \mathrm{id}_{\mathrm{U}}}{\to} (\mathrm{V} \otimes \mathrm{W}) \times \mathrm{U} \overset{\otimes}{\to} (\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U}$$

and

$$\mathrm{V} \times (\mathrm{W} \times \mathrm{U}) \overset{\mathrm{id}_{\mathrm{U}} \times \otimes}{\to} (\mathrm{V} \times \mathrm{W}) \otimes \mathrm{U} \overset{\otimes}{\to} \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U}).$$

It is a simple matter to check that these composite maps have the same universal property as the map $\mathrm{V} \times \mathrm{W} \times \mathrm{U} \overset{\otimes}{\to} \mathrm{V} \otimes \mathrm{W} \otimes \mathrm{U}$. The result now follows from the existence and essential uniqueness results proven so far (**??** and **??**). ■

We shall use the first isomorphism and the obvious generalizations to identify $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ with all legal parenthetical constructions such as $(((\mathrm{V}_1 \otimes \mathrm{V}_2) \otimes \cdots \otimes \mathrm{V}_j) \otimes \cdots) \otimes \mathrm{V}_k$ and so forth. In short, we may construct $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ by tensoring spaces two at a time. In particular we assume the isomorphisms

$$(\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k) \otimes (\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k) \cong \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \otimes \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k$$

which map $(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k) \otimes (\mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k)$ to $\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k \otimes \mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k$.

Consider the situation where we have module homomorphisms $h_i : \mathrm{W}_i \to \mathrm{V}_i$ for $1 \leq i \leq m$. We may then define a map $T(h_1, ..., h_m) : \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_m \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_m$ (by using the universal property again) so that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{W}_1 \times \cdots \times \mathrm{W}_m & \overset{h_1 \times \ldots \times h_m}{\rightarrow} & \mathrm{V}_1 \times \cdots \times \mathrm{V}_m \\ \otimes^k \downarrow & & \otimes^k \downarrow \\ \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_m & \overset{T(h_1, \ldots, h_m)}{\rightarrow} & \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_m \end{array}.$$

This is functorial in the sense that

$$T(h_1, ..., h_m) \circ T(g_1, ..., g_m) = T(h_1 \circ g_1, ..., h_m \circ g_m)$$

and $T(\mathrm{id}, ..., \mathrm{id}) = \mathrm{id}$. Also, $T(h_1, ..., h_m)$ has the following effect on decomposable elements:

$$T(h_1, ..., h_m)(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_m) = h_1(\mathrm{v}_1) \otimes \cdots \otimes h_m(\mathrm{v}_m).$$

Now we could jump the gun a bit and use the notation $h_1 \otimes \cdots \otimes h_m$ for $T(h_1, ..., h_m)$ but is this the same thing as the element of $\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_1, \mathrm{V}_1) \otimes \cdots \otimes \mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_m, \mathrm{V}_m)$ which must be denoted the same way? The answer is that in general, these are distinct objects. On the other hand, there is little harm done if context determines which of the two possible meanings we are invoking. Furthermore, we shall see than in many cases, the two meanings actually do coincide.

A basic isomorphism which is often used is the following:

Proposition 26.7 *For* $\mathsf{R}-$*modules* $\mathrm{W}, \mathrm{V}, U$ *we have*

$$\mathrm{Hom}_{\mathsf{R}}(\mathrm{W} \otimes \mathrm{V}, \mathrm{U}) \cong L(\mathrm{W}, \mathrm{V}; \mathrm{U})$$

More generally,

$$\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k, \mathrm{U}) \cong L(\mathrm{W}_1, ..., \mathrm{W}_k; \mathrm{U})$$

Proof. This is more or less just a restatement of the universal property of $\mathrm{W} \otimes \mathrm{V}$. One should check that this association is indeed an isomorphism. ∎

Exercise 26.9 *Show that if* W *is free with basis* $(f_1, ..., f_n)$ *then* W^* *is also free and has a dual basis* $f^1, ..., f^n$*, that is,* $f^i(f_j) = \delta^i_j$.

Theorem 26.20 *If* $\mathrm{V}_1, ..., \mathrm{V}_k$ *are free* $\mathsf{R}-$*modules and if* $(v^j_1, ..., v^j_{n_j})$ *is a basis for* V_j *then set of all decomposable elements of the form* $v^1_{i_1} \otimes \cdots \otimes v^k_{i_k}$ *form a basis for* $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.

Proposition 26.8 *There is a unique* R*-module map* $\iota : \mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^* \to (\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k)^*$ *such that if* $\alpha_1 \otimes \cdots \otimes \alpha_k$ *is a (decomposable) element of* $\mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^*$ *then*

$$\iota(\alpha_1 \otimes \cdots \otimes \alpha_k)(w_1 \otimes \cdots \otimes w_k) = \alpha_1(w_1) \cdots \alpha_k(w_k).$$

If the modules are all free then this is an isomorphism.

Proof. If such a map exists, it must be unique since the decomposable elements span $\mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^*$. To show existence we define a multilinear map

$$\vartheta : \mathrm{W}_1^* \times \cdots \times \mathrm{W}_k^* \times \mathrm{W}_1 \times \cdots \times \mathrm{W}_k \to \mathsf{R}$$

by the recipe

$$(\alpha_1, ..., \alpha_k, w_1, ..., w_k) \mapsto \alpha_1(w_1) \cdots \alpha_k(w_k).$$

By the universal property there must be a linear map

$$\widetilde{\vartheta} : \mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^* \otimes \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k \to \mathsf{R}$$

such that $\widetilde{\vartheta} \circ u = \vartheta$ where u is the universal map. Now define

$$\begin{aligned} &\iota(\alpha_1 \otimes \cdots \otimes \alpha_k)(w_1 \otimes \cdots \otimes w_k) \\ &:= \widetilde{\vartheta}(\alpha_1 \otimes \cdots \otimes \alpha_k \otimes w_1 \otimes \cdots \otimes w_k). \end{aligned}$$

The fact, that ι is an isomorphism in case the W_i are all free follows easily from exercise **??** and theorem **??**. Once we view an element of W_i as a functional from $\mathrm{W}_i^{**} = L(\mathrm{W}_i^*; \mathsf{R})$ we see that the effect of this isomorphism is to change the interpretation of the tensor product to the "map" tensor product in $(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k)^* = L(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k; \mathsf{R})$. Thus the basis elements match up under ι. ■

Definition 26.62 *The k-th tensor power of a module* W *is defined to be*

$$\mathrm{W}^{\otimes k} := \mathrm{W} \otimes \cdots \otimes \mathrm{W}.$$

This module is also denoted $\bigotimes^k(\mathrm{W})$*. We also, define the space of* $({}^r{}_s)$*-**tensors on*** W *:*

$$\bigotimes{}^r{}_s(\mathrm{W}) := \mathrm{W}^{\otimes r} \otimes (\mathrm{W}^*)^{\otimes s}.$$

Similarly, $\bigotimes{}_s{}^r(\mathrm{W}) := (\mathrm{W}^*)^{\otimes s} \otimes \mathrm{W}^{\otimes r}$ *is the space of* $({}_s{}^r)$*-**tensors on*** W*.*

Again, although we distinguish $\bigotimes{}^r{}_s(\mathrm{W})$ from $\bigotimes{}_s{}^r(\mathrm{W})$ we shall be able to develop things so as to use mostly the space $\bigotimes{}^r{}_s(\mathrm{W})$ and so by default we take $\bigotimes_s^r(\mathrm{W})$ to mean $\bigotimes{}^r{}_s(\mathrm{W})$.

If $(v_1, ..., v_n)$ is a basis for a module V and $(\upsilon^1, ..., \upsilon^n)$ the dual basis for V^* then a basis for $\bigotimes_s^r(\mathrm{V})$ is given by

$$\{v_{i_1} \otimes \cdots \otimes v_{i_r} \otimes \upsilon^{j_1} \otimes \cdots \otimes \upsilon^{j_s}\}$$

where the index set is the set $\mathcal{I}(r,s,n)$ defined by

$$\mathcal{I}(r,s,n) := \{(i_1,...,i_r;j_1,...,j_s) : 1 \leq i_k \leq n \text{ and } 1 \leq j_k \leq n\}.$$

Thus $\bigotimes^r_s(\mathrm{V})$ has dimension $n^r n^s$ (where $n = \dim(\mathrm{V})$).

We re-state the universal property in this special case of tensors:

Proposition 26.9 (Universal mapping property) *Given a module or vector space* V *over* R*, then* $\bigotimes^r_s(\mathrm{V})$ *has associated with it, a map*

$$\otimes^r_s : \underbrace{\mathrm{V}\times\mathrm{V}\times\cdots\times\mathrm{V}}_{r}\times\underbrace{\mathrm{V}^*\times\mathrm{V}^*\times\cdots\times\mathrm{V}^*}_{s}\to\bigotimes^r_s(\mathrm{V})$$

such that for any multilinear map $\Lambda \in T_r{}^s(\mathrm{V};\mathsf{R})$*;*

$$\Lambda : \underbrace{\mathrm{V}\times\mathrm{V}\times\cdots\times\mathrm{V}}_{r-times}\times\underbrace{\mathrm{V}^*\times\mathrm{V}^*\cdots\times\mathrm{V}^*}_{s-times}\to\mathsf{R}$$

there is a unique linear map $\widetilde{\Lambda} : \bigotimes{}^r{}_s(\mathrm{V}) \to \mathsf{R}$ *such that* $\widetilde{\Lambda}\circ\otimes^r_s = \Lambda$*. Up to isomorphism, the space* $\bigotimes{}^r{}_s(\mathrm{V})$ *is the unique space with this universal mapping property.*

Corollary 26.3 *There is an isomorphism* $(\bigotimes{}^r{}_s(\mathrm{V}))^* \cong T_r{}^s(\mathrm{V})$ *given by* $\widetilde{\Lambda} \mapsto \widetilde{\Lambda}\circ\otimes^r_s$*. (**Warning**: Notice that the* $T^r{}_s(\mathrm{V})$ *occurring here is **not** the default space* $T^r{}_s(\mathrm{V})$ *that we often denote by* $T^r_s(\mathrm{V})$*.)*

Corollary 26.4 $(\bigotimes{}^r{}_s(\mathrm{V}^*))^* = T_r{}^s(\mathrm{V}^*)$

Now along the lines of the map of proposition **??** we have a homomorphism

$$\iota^r_s : \bigotimes{}^r{}_s(\mathrm{V}) \to T^r{}_s(\mathrm{V}) \tag{26.7}$$

given by

$$\iota^r_s\big((\mathrm{v}_1\otimes\cdots\otimes\mathrm{v}_r\otimes\eta^1\otimes\cdots..\otimes\eta^s)\big)(\theta^1,...,\theta^r,\mathrm{w}_1,..,\mathrm{w}_s)$$
$$= \theta_1(\mathrm{v}_1)\theta_2(\mathrm{v}_2)\cdots\theta_l(\mathrm{v}_l)\eta^1(\mathrm{w}_1)\eta^2(\mathrm{w}_2)\cdots\eta^k(\mathrm{w}_k)$$

$\theta^1,\theta^2,...,\theta^r \in \mathrm{V}^*$ and $\mathrm{w}_1,\mathrm{w}_2,..,\mathrm{w}_k \in \mathrm{V}$. If V is a finite dimensional free module then we have $\mathrm{V} = \mathrm{V}^{**}$. This is the reflexive property. We say that V is totally reflexive if the homomorphism **??** just given is in fact an isomorphism. This happens for free modules:

Proposition 26.10 *For a finite dimensional free module* V *we have a natural isomorphism* $\bigotimes{}^r{}_s(\mathrm{V}) \cong T^r{}_s(\mathrm{V})$*. The isomorphism is given by the map* ι^r_s *(see* **??***)*

Proof. Just to get the existence of a natural isomorphism we may observe that

$$\begin{aligned}\bigotimes{}^{r}{}_{s}(\mathrm{V}) &= \bigotimes{}^{r}{}_{s}(\mathrm{V}^{**}) = (\bigotimes{}^{r}{}_{s}(\mathrm{V}^{*}))^{*}\\ &= T_{r}{}^{s}(\mathrm{V}^{*}) = L(\mathrm{V}^{*r}, \mathrm{V}^{**s}; \mathsf{R})\\ &= L(\mathrm{V}^{*r}, \mathrm{V}^{s}; \mathsf{R}) :=T^{r}{}_{s}(\mathrm{V}).\end{aligned}$$

We would like to take a more direct approach. Since V is free we may take a basis $\{f_1, ..., f_n\}$ and a dual basis $\{f^1, ..., f^n\}$ for V^*. It is easy to see that ι^r_s sends the basis elements of $\bigotimes{}^{r}{}_{s}(\mathrm{V})$ to basis elements of $T^{r}{}_{s}(\mathrm{V}; \mathsf{R})$ as for example

$$\iota^1_1 : f^i \otimes f_j \mapsto f^i \otimes f_j$$

where only the interpretation of the $\otimes$ changes. ∎

In the finite dimensional case, we will identify $\bigotimes{}^{r}{}_{s}(\mathrm{V})$ with the space $T^{r}{}_{s}(\mathrm{V};\mathsf{R}) = L(\mathrm{V}^{r*}, \mathrm{V}^{s}; \mathsf{R})$ of r,s-multilinear maps. We may freely think of a decomposable tensor $\mathrm{v}_1 \otimes ... \otimes \mathrm{v}_r \otimes \eta^1 \otimes ... \otimes \eta^s$ as a *multilinear map* by the formula

$$\begin{aligned}&(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_r \otimes \eta^1 \otimes \cdots .. \otimes \eta^s) \cdot (\theta^1, ..., \theta^r, \mathrm{w}_1, .., \mathrm{w}_s)\\ &= \theta_1(\mathrm{v}_1)\theta_2(\mathrm{v}_2) \cdots \theta_l(\mathrm{v}_l)\eta^1(\mathrm{w}_1)\eta^2(\mathrm{w}_2) \cdots \eta^k(\mathrm{w}_k)\end{aligned}$$

U) of smooth vector fields over an open set U in some manifold M. We shall see that for finite dimensional manifolds $T^r_s(\mathfrak{X}(U))$ is naturally isomorphic to the smooth sections of a so called tensor bundle. R−Algebras

Definition 26.63 *Let* R *be a commutative ring. An* R−***algebra*** $\mathfrak{A}$ *is an* R−*module that is also a ring with identity* $1_{\mathfrak{A}}$ *where the ring addition and the module addition coincide; and where*

1) $r(a_1a_2) = (ra_1)a_2 = a_1(ra_2)$ *for all* $a_1, a_2 \in \mathfrak{A}$ *and all* $r \in \mathsf{R}$,

2) $(r_1r_2)(a_1a_2) = (r_1a_1)(r_2a_2)$.

If we also have $(a_1a_2)a_3 = a_1(a_2a_3)$ *for all* $a_1, a_2, a_3 \in \mathfrak{A}$ *we call* $\mathfrak{A}$ *an associative* R−*algebra.*

Definition 26.64 *Let* $\mathfrak{A}$ *and* $\mathfrak{B}$ *be* R−*algebras. A module homomorphism* $h : \mathfrak{A} \to \mathfrak{B}$ *that is also a ring homomorphism is called an* R−***algebra homomorphism****. Epimorphism, monomorphism and isomorphism are defined in the obvious way.*

If a submodule $\mathfrak{I}$ of an algebra $\mathfrak{A}$ is also a two sided ideal with respect to the ring structure on $\mathfrak{A}$ then $\mathfrak{A}/\mathfrak{I}$ is also an algebra.

Example 26.20 *The set of all smooth functions* $C^\infty(U)$ *is an* $\mathbb{R}$−*algebra (*$\mathbb{R}$ *is the real numbers) with unity being the function constantly equal to* 1.

Example 26.21 *The set of all complex* $n \times n$ *matrices is an algebra over* $\mathbb{C}$ *with the product being matrix multiplication.*

Example 26.22 *The set of all complex $n \times n$ matrices with real polynomial entries is an algebra over the ring of polynomials $\mathbb{R}[x]$.*

Definition 26.65 *The set of all endomorphisms of an $\mathsf{R}-$module W is an $\mathsf{R}-$algebra denoted $End_{\mathsf{R}}(\mathrm{W})$ and called the **endomorphism algebra** of W. Here, the sum and scalar multiplication is defined as usual and the product is composition. Note that for $r \in \mathsf{R}$*

$$r(f \circ g) = (rf) \circ g = f \circ (rg)$$

where $f, g \in End_{\mathsf{R}}(\mathrm{W})$.

Definition 26.66 *A $\mathbb{Z}$-graded R-algebra is an R-algebra with a direct sum decomposition $\mathfrak{A} = \sum_{i \in \mathbb{Z}} \mathfrak{A}_i$ such that $\mathfrak{A}_i \mathfrak{A}_j \subset \mathfrak{A}_{i+j}$.*

Definition 26.67 *Let $\mathfrak{A} = \sum_{i \in \mathbb{Z}} \mathfrak{A}_i$ and $\mathfrak{B} = \sum_{i \in \mathbb{Z}} \mathfrak{B}_i$ be $\mathbb{Z}$-graded algebras. An R-algebra homomorphism $h : \mathfrak{A} \to \mathfrak{B}$ is a called a $\mathbb{Z}$-graded homomorphism if $h(\mathfrak{A}_i) \subset \mathfrak{B}_i$ for each $i \in \mathbb{Z}$.*

We now construct the tensor algebra on a fixed $\mathsf{R}-$module W. This algebra is important because is universal in a certain sense and contains the symmetric and alternating algebras as homomorphic images. Consider the following situation: $\mathfrak{A}$ is an $\mathsf{R}-$algebra, W an $\mathsf{R}-$module and $\phi : \mathrm{W} \to \mathfrak{A}$ a module homomorphism. If $h : \mathfrak{A} \to \mathfrak{B}$ is an algebra homomorphism then of course $h \circ \phi : \mathrm{W} \to \mathfrak{B}$ is an $\mathsf{R}-$module homomorphism.

Definition 26.68 *Let W be an $\mathsf{R}-$module. An $\mathsf{R}-$ algebra $\mathfrak{U}$ together with a map $\phi : \mathrm{W} \to \mathfrak{U}$ is called universal with respect to W if for any $\mathsf{R}-$module homomorphism $\psi : \mathrm{W} \to \mathfrak{B}$ there is a unique algebra homomorphism If $h : \mathfrak{U} \to \mathfrak{B}$ such that $h \circ \phi = \psi$.*

Again if such a universal object exists it is unique up to isomorphism. We now exhibit the construction of this type of universal algebra. First we define $T^0(\mathrm{W}) := \mathsf{R}$ and $T^1(\mathrm{W}) := \mathrm{W}$. Then we define $T^k(\mathrm{W}) := \mathrm{W}^{k\otimes} = \mathrm{W} \otimes \cdots \otimes \mathrm{W}$. The next step is to form the direct sum $T(\mathrm{W}) := \sum{}_{i=0}^{\infty} T^i(\mathrm{W})$. In order to make this a $\mathbb{Z}$-graded algebra we define $T^i(\mathrm{W}) := 0$ for $i < 0$ and then define a product on $T(\mathrm{W}) := \sum{}_{i \in \mathbb{Z}} T^i(\mathrm{W})$ as follows: We know that for $i, j > 0$ there is an isomorphism $\mathrm{W}^{i\otimes} \otimes \mathrm{W}^{j\otimes} \to \mathrm{W}^{(i+j)\otimes}$ and so a bilinear map $\mathrm{W}^{i\otimes} \times \mathrm{W}^{j\otimes} \to \mathrm{W}^{(i+j)\otimes}$ such that

$$\mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_i \times \mathrm{w}_1' \otimes \cdots \otimes \mathrm{w}_j' \mapsto \mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_i \otimes \mathrm{w}_1' \otimes \cdots \otimes \mathrm{w}_j'.$$

Similarly, we define $T^0(\mathrm{W}) \times \mathrm{W}^{i\otimes} = \mathsf{R} \times \mathrm{W}^{i\otimes} \to \mathrm{W}^{i\otimes}$ by just using scalar multiplication. Also, $\mathrm{W}^{i\otimes} \times \mathrm{W}^{j\otimes} \to 0$ if either i or j is negative. Now we may use the symbol $\otimes$ to denote these multiplications without contradiction and put then together to form an product on $T(\mathrm{W}) := \sum{}_{i \in \mathbb{Z}} T^i(\mathrm{W})$. It is now clear that $T^i(\mathrm{W}) \times T^j(\mathrm{W}) \mapsto T^i(\mathrm{W}) \otimes T^j(\mathrm{W}) \subset T^{i+j}(\mathrm{W})$ where we make the needed trivial definitions for the negative powers $T^i(\mathrm{W}) = 0$, $i < 0$. Thus $T(\mathrm{W})$ is a graded algebra.

26.12.1 Smooth Banach Vector Bundles

The tangent bundle and cotangent bundle are examples of a general object called a (smooth) vector bundle which we have previously defined in the finite dimensional case. As a sort of review and also to introduce the ideas in the case of infinite dimensional manifolds we will define again the notion of a smooth vector bundle. For simplicity we will consider only C^∞ manifold and maps in this section. Let E be a Banach space. The most important case is when E is a finite dimensional vector space and in that case we might as well take $\mathsf{E}= \mathbb{R}^n$. It will be convenient to introduce the concept of a general fiber bundle and then specialize to vector bundles. The following definition is not the most efficient logically since there is some redundancy built in but is presented in this form for pedagogical reasons.

Definition 26.69 *Let F be a smooth manifold modelled on F. A smooth* ***fiber bundle*** *$\xi = (E, \pi_E, M, F)$ with typical fiber F consists of*

1) smooth manifolds E and M referred to as the ***total space*** *and the* ***base space*** *respectively and modelled on Banach spaces $\mathsf{M}\times\mathsf{F}$ and M respectively;*

2) a smooth surjection $\pi_E : E \to M$ such that each ***fiber*** *$E_x = \pi^{-1}\{x\}$ is diffeomorphic to F;*

3) a cover of the ***base space*** *M by domains of maps $\phi_\alpha : E_{U_\alpha} := \pi_E^{-1}(U_\alpha) \to U_\alpha \times F$, called* ***bundle charts,*** *which are such that the following diagram commutes:*

$$\begin{array}{ccccc} E_{U_\alpha} & & \overset{(\pi_{U_\alpha},\Phi_\alpha)}{\rightarrow} & & U_\alpha \times F \\ \pi_E & \searrow & & \swarrow & \\ & & U_\alpha & & \end{array}.$$

Thus each ϕ_α is of the form $(\pi_{U_\alpha}, \Phi_\alpha)$ where $\pi_{U_\alpha} := \pi_E| U_\alpha$ and $\Phi_\alpha : E_{U_\alpha} \to F$ is a smooth submersion.

Definition 26.70 *The family of bundle charts whose domains cover the base space of a fiber bundle as in the above definition is called a* ***bundle atlas****.*

For all $x \in U_\alpha$, each restriction $\Phi_{\alpha,x} := \Phi_\alpha|_{E_x}$ is a diffeomorphism onto $\mathbf{F}$. Whenever we have two bundle charts $\phi_\alpha = (\pi_E, \Phi_\alpha)$ and $\phi_\beta = (\pi_E, \Phi_\beta)$ such that $U_\alpha \cap U_\beta \neq \emptyset$ then for every $x \in U_\alpha \cap U_\beta$ we have the diffeomorphism $\Phi_{\alpha\beta,x} = \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1} : F \to F$. Thus we have map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{Diff}(F)$ given by $g_{\alpha\beta}(x) := \Phi_{\alpha\beta,x}$. Notice that $g_{\beta\alpha} \circ g_{\alpha\beta}^{-1} = \mathrm{id}$. The maps so formed satisfy the following cocycle conditions:

$$g_{\gamma\beta} \circ g_{\alpha\gamma} = g_{\alpha\beta} \text{ whenever } U_\alpha \cap U_\beta \cap U_\gamma \neq \emptyset.$$

Let ξ be as above and let U be open in M. Suppose we have a smooth map $\phi : E_U \to U \times F$ such that the diagram

$$\begin{array}{ccccc} E_U & & \overset{\phi}{\rightarrow} & & U \times F \\ \pi_U & \searrow & & \swarrow & \\ & & U & & \end{array}$$

where $E_U := \pi_E^{-1}(U)$ as before. We call ϕ a trivialization and even if ϕ was not one of the bundle charts of a given bundle atlas, it must have the form (π_E, Φ) and we may enlarge the atlas by including this map. We do not wish to consider the new atlas as determining a new bundle so instead we say that the new atlas is equivalent. There is a unique maximal atlas for the bundle which just contains every possible trivialization.

Now given any open set U in the base space M of a fiber bundle $\xi = (E, \pi_E, M, F)$ we may form the restriction $\xi| U$ which is the fiber bundle $(\pi_E^{-1}(U), \pi_E| U, U, F)$. To simplify notation we write $E_U := \pi_E^{-1}(U)$ and $\pi_E| U := \pi_U$. This is a special case of the notion of a pull back bundle.

One way in which vector bundles differ from general fiber bundles is with regard to the existence on global sections. A vector bundle always has at least one global section. Namely, the **zero section** $0_E : M \to E$ which is given by $x \mapsto 0_x \in E_x$. Our main interest at this point is the notion of a vector bundle. Before we proceed with our study of vector bundles we include one example of fiber bundle that is not a vector bundle.

Example 26.23 *Let M and F be smooth manifolds and consider the projection map $pr_1 : M \times F \to M$. This is a smooth fiber bundle with typical fiber F and is called a **product bundle** or a **trivial bundle**.*

Example 26.24 *Consider the tangent bundle $\tau_M : TM \to M$ of a smooth manifold modelled on $\mathbb{R}^n$. This is certainly a fiber bundle (in fact, a vector bundle) with typical fiber $\mathbb{R}^n$ but we also have the **bundle of nonzero vectors** $\pi : TM^\times \to M$ defined by letting $TM^\times := \{v \in TM : v \neq 0\}$ and $\pi := \tau_M|_{TM^\times}$. This bundle may have no global sections.*

Remark 26.14 *A "structure" on a fiber bundle is determined by requiring that the atlas be paired down so that the transition maps all have values in some subset G (usually a subgroup) of $\mathrm{Diff}(F)$. Thus we speak of a $G-$**atlas** for $\xi = (E, \pi_E, M, F)$. In this case, a trivialization $(\pi_E, \Phi) : E_U \to U \times F$ is compatible with a given $G-$atlas $\mathcal{A}(G)$ if $\Phi_{\alpha,x} \circ \Phi_x^{-1} \in G$ and $\Phi_x \circ \Phi_{\alpha,x}^{-1} \in G$ for all $(\pi_E, \Phi_\alpha) \in \mathcal{A}(G)$. The set of all trivializations (bundle charts) compatible which a given $G-$atlas is a maximal $G-$atlas and is called a G-**structure.** Clearly, any $G-$atlas determines a G-structure. A fiber bundle $\xi = (E, \pi_E, M, F)$ with a G-atlas is called a G-bundle and any two G-atlases contained in the same maximal G-atlas are considered equivalent and determine the same G-bundle. We will study this idea in detail after we have introduced the notion of a Lie group.*

We now introduce our current object of interest.

Definition 26.71 *A (real) **vector bundle** is a fiber bundle $\xi = (E, \pi_E, M, \mathsf{E})$ with typical fiber a (real) Banach space E such that for each pair of bundle chart domains U_α and U_β with nonempty intersection, the map*

$$g_{\alpha\beta} : x \mapsto \Phi_{\alpha\beta,x} := \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1}$$

is a C^∞ morphism $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathsf{E})$. If E is finite dimensional, say $\mathsf{E} = \mathbb{R}^n$, then we say that $\xi = (E, \pi_E, M, \mathbb{R}^n)$ has rank n.

So if $v_x \in \pi_E^{-1}(x) \subset E_{U_\alpha}$ then $\phi_\alpha(v_x) = (x, \Phi_{\alpha,x}(v_x))$ for $\Phi_{\alpha,x} : E_x \to \mathsf{E}$ a diffeomorphism. Thus we can transfer the vector space structure of V to each fiber E_x in a well defined way since $\Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1} \in \mathrm{GL}(\mathsf{E})$ for any x in the intersection of two VB-chart domains U_α and U_β. Notice that we have $\phi_\alpha \circ \phi_\beta^{-1}(x, \mathsf{v}) = (x, g_{\alpha\beta}(x) \cdot \mathsf{v})$ where $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathsf{E})$ is differentiable and is given by $g_{\alpha\beta}(x) = \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1}$. Notice that $g_{\alpha\beta}(x) \in \mathrm{GL}(\mathsf{E})$.

A **complex vector bundle** is defined in an analogous way. For a complex vector bundle the typical fiber is a complex vector space (Banach space) and the transition maps have values in $\mathrm{GL}(\mathsf{E};\mathbb{C})$.

The set of all sections of real (resp. complex) vector bundle is a vector space over $\mathbb{R}$ (resp. $\mathbb{C}$) and a module over the ring of smooth real valued (resp. complex valued) functions.

Remark 26.15 *If E is of finite dimension then the smoothness of the maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathsf{E})$ is automatic.*

Definition 26.72 *The maps $g_{\alpha\beta}(x) := \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1}$ are called **transition maps**.*

The transition maps always satisfy the following **cocycle condition**:

$$g_{\gamma\beta}(x) \circ g_{\beta\alpha}(x) = g_{\gamma\alpha}(x)$$

In fact, these maps encode the entire vector bundle up to isomorphism:

Remark 26.16 *The following definition assumes the reader knows the definition of a Lie group and has a basic familiarity with Lie groups and Lie group homomorphisms. We shall study Lie groups in Chapter **??**. The reader may skip this definition.*

Definition 26.73 *Let G be a Lie subgroup of $\mathrm{GL}(\mathsf{E})$. We say that $\pi_E : E \to M$ has a **structure group** G if there is a cover by trivializations (vector bundle charts) $\phi_\alpha : E_{U_\alpha} \to U_\alpha \times \mathsf{E}$ such that for every non-empty intersection $U_\alpha \cap U_\beta$, the transition maps $g_{\alpha\beta}$ take values in G.*

Remark 26.17 *Sometimes it is convenient to define the notion of vector bundle chart in a slightly different way. Notice that U_α is an open set in M and so ϕ_α is not quite a chart for the total space manifold E. But by choosing a possibly smaller open set inside U_α we may assume that U_α is the domain of an admissible chart U_α, ψ_α for M. Then we can compose to get a map $\widetilde{\phi}_\alpha : E_{U_\alpha} \to \psi_\alpha(U_\alpha) \times \mathsf{E}$. The maps now can serve as admissible charts for the differentiable manifold E. This leads to an alternative definition of VB-chart which fits better with what we did for the tangent bundle and cotangent bundle:*

Definition 26.74 (Type II vector bundle charts) *A **(type II) vector bundle chart** on an open set $V \subset E$ is a fiber preserving diffeomorphism $\phi : V \to O\times\mathsf{E}$ which covers a diffeomorphism $\underline{\phi}{:}\pi_E(V) \to O$ in the sense that the following diagram commutes*

$$\begin{array}{cccc} & V & \xrightarrow{\phi} & O\times\mathsf{E} \\ \pi_E & \downarrow & & \downarrow pr_1 \\ & \pi_E(V) & \xrightarrow[\underline{\phi}]{} & O \end{array}$$

and which is a linear isomorphism on each fiber.

Example 26.25 *The maps $T\psi_\alpha : TU_\alpha \to \psi_\alpha(U_\alpha) \times \mathsf{E}$ are (type II) VB-charts and so not only give TM a differentiable structure but also provide TM with a vector bundle structure. Similar remarks apply for T^*M.*

Example 26.26 *Let E be a vector space and let $E = M\times\mathsf{E}$. The using the projection $pr_1 : M\times\mathsf{E}\to M$ we obtain a vector bundle. A vector bundle of this simple form is called a trivial vector bundle.*

Define the sum of two section s_1 and s_2 by $(s_1 + s_2)(p) := s_1(p) + s_2(p)$. For any $f \in C^\infty(U)$ and $s \in \Gamma(U, E)$ define a section fs by $(fs)(p) = f(p)s(p)$. Under these obvious definitions $\Gamma(U, E)$ becomes a $C^\infty(U)$-module.

The the appropriate morphism in our current context is the vector bundle morphism:

Definition 26.75 ***Definition 26.76*** *Let (E, π_E, M) and $(F, \pi_F, N\ )$ be vector bundles. A **vector bundle morphism** $(E, \pi_E, M) \to (F, \pi_F, N)$ is a pair of maps $f : E \to F$ and $f_0 : M \to N$ such that*

1. ***Definition 26.77*** *1) The following diagram commutes:*

$$\begin{array}{cccccc} & E & \xrightarrow{f} & F & \\ \pi_E & \downarrow & & \downarrow & \pi_F \\ & M & \xrightarrow[f_0]{} & N & \end{array}$$

and $f|_{E_p}$ is a continuous linear map from E_p into $F_{f_0(p)}$ for each $p \in M$.

2) For each $x_0 \in M$ there exist VB-charts $(\pi_E, \Phi) : E_U \to U\times\mathsf{E}$ and $(\pi_E, \Phi'_\alpha) : F_{U'} \to U'\times\mathsf{E}'$ with $x_0 \in U$ and $f_0(U) \subset V$ such that

$$x \mapsto \Phi'|_{F_{f(x)}} \circ f_0 \circ \Phi|_{E_x}$$

is a smooth map from U into $\mathrm{GL}(\mathsf{E},\mathsf{E}')$.

Notation 26.6 *Each of the following is a valid way to refer to a vector bundle morphism:*

1) $(f, f_0) : (E, \pi_E, M, \mathsf{E}) \to (F, \pi_F, N, \mathsf{F})$

2) $f : (E, \pi_E, M) \to (F, \pi_F, N)$ *(the map* f_0 *is induced and hence understood)*

3) $f : \xi_1 \to \xi_2$ *(this one is concise and fairly exact once it is set down that* $\xi_1 = (E_1, \pi_1, M)$ *and* $\xi_2 = (E_2, \pi_2, M)$*)*

4) $f : \pi_E \to \pi_F$

5) $E \xrightarrow{f} F$

Remark 26.18 *There are many variations of these notations in use and the reader would do well to get used to this kind of variety. Actually, there will be much more serious notational difficulties in store for the novice. It has been said that notation is one of the most difficult aspects of differential geometry. On the other hand, once the underlying geometric picture has been properly understood , one may "see through" the notation. Drawing diagrams while interpreting equations is often a good idea.*

Definition 26.78 ***Definition 26.79*** *If* f *is an (linear) isomorphism on each fiber* E_p *then we say that* f *is a* ***vector bundle isomorphism*** *and the two bundles are considered equivalent.*

Notation 26.7 *If* $\widetilde{f}$ *is a VB morphism from a vector bundle* $\pi_E : E \to M$ *to a vector bundle* $\pi_F : F \to M$ *we will sometimes write this as* $\widetilde{f} : \pi_E \to \pi_F$ *or* $\pi_E \xrightarrow{\widetilde{f}} \pi_F$*.*

Definition 26.80 *A vector bundle is called* ***trivial*** *if there is a there is a vector bundle isomorphism onto a trivial bundle:*

$$\begin{array}{ccccc} & & \cong & & \\ E & & \to & & M \times \mathsf{E} \\ \pi_E & \searrow & & \swarrow & pr_1 \\ & & M & & \end{array}$$

Now we make the observation that a section of a trivial bundle is in a sense, nothing more than a vector-valued function since all sections $s \in \Gamma(M, M \times \mathsf{E})$ are of the form $p \to (p, f(p))$ for a unique function $f \in C^\infty(M, \mathsf{E})$. It is common to identify the function with the section.

Now there is an important observation to be made here; a trivial bundle always has a section which never takes on the value zero. There reason is that we may always take a trivialization $\phi : E \to M \times \mathsf{E}$ and then transfer the obvious nowhere-zero section $p \mapsto (p, 1)$ over to E. In other words, we let $s_1 : M \to E$ be defined by $s_1(p) = \phi^{-1}(p, 1)$. We now use this to exhibit a very simple example of a non-trivial vector bundle:

Example 26.27 (Möbius bundle) *Let E be the quotient space of $[0,1]\times\mathbb{R}$ under the identification of $(0,t)$ with $(1,-t)$. The projection $[0,1]\times\mathbb{R}\to[0,1]$ becomes a map $E\to S^1$ after composition with the quotient map:*

$$\begin{array}{ccc}[0,1]\times\mathbb{R} & \to & [0,1]\\ \downarrow & & \downarrow\\ E & \to & S^1\end{array}$$

Here the circle arises as $[0,1]/\sim$ where we have the induced equivalence relation given by taking $0\sim 1$ in $[0,1]$. The familiar Mobius band has an interior which is diffeomorphic to the Mobius bundle.

Now we ask if it is possible to have a nowhere vanishing section of E. It is easy to see that sections of E correspond to continuous functions $f:[0,1]\to\mathbb{R}$ such that $f(0)=-f(1)$. But then continuity forces such a function to take on the value zero which means that the corresponding section of E must vanish somewhere on $S^1=[0,1]/\sim$. Of course, examining a model of a Mobius band is even more convincing; any nonzero section of E could be, if such existed, normalized to give a map from S^1 to the boundary of a Möbius band which only went around once, so to speak, and inspection of a model would convince the reader that this is impossible.

Let $\xi_1=(E_1,\pi_1,M,\mathsf{E}_1)$ and $\xi_2=(E_2,\pi_2,M,\mathsf{E}_2)$ be vector bundles locally isomorphic to $\mathsf{M}\times\ \mathsf{E}_1$ and $\mathsf{M}\times\ \mathsf{E}_2$ respectively. We say that the sequence of vector bundle morphisms

$$0\to\xi_1\xrightarrow{f}\xi_2$$

is **exact** if the following conditions hold:

1. There is an open covering of M by open sets U_α together with trivializations $\phi_{1,\alpha}:\pi_1^{-1}(U_\alpha)\to U_\alpha\times\ \mathsf{E}_1$ and $\phi_{2,\alpha}:\pi_2^{-1}(U_\alpha)\to U_\alpha\times\ \mathsf{E}_2$ such that $\mathsf{E}_2=\mathsf{E}_1\times\mathsf{F}$ for some Banach space F;

2. the diagram below commutes for all α :

$$\begin{array}{ccccc} & \pi_1^{-1}(U_\alpha) & \to & \pi_2^{-1}(U_\alpha) & \\ \phi_{1,\alpha} & \downarrow & & \downarrow & \phi_{2,\alpha}\\ & U_\alpha\times\mathsf{E}_1 & \to & U_\alpha\times\mathsf{E}_1\times\mathsf{F} & \end{array}$$

Definition 26.81 *A **subbundle** of a vector bundle $\xi=(E,\pi,M)$ is a vector bundle of the form $\xi=(L,\pi|_L,M)$ where $\pi|_L$ is the restriction to $L\subset E$, and where $L\subset E$ is a submanifold such that*

$$0\to\xi|_L\to\xi$$

is exact. Here, $\xi|_L\to\xi$ is the bundle map given by inclusion: $L\hookrightarrow E$.

*Equivalently, $\pi|_L:L\to M$ is a **subbundle** if $L\subset E$ is a submanifold and there is a splitting $\mathsf{E}=\mathsf{E}_1\times\mathsf{F}$ such that for each $p\in M$ there is a bundle chart $\phi:\pi^{-1}U\to U\times\mathsf{E}$ with $p\in U$ and $\phi((\pi^{-1}U)\cap L)=U\times\mathsf{E}_1\times\{0\}$.*

Definition 26.82 *The chart ϕ from the last definition is said to be adapted to the subbundle.*

Notice that if $L \subset E$ is as in the previous definition then $\pi|_L : L \to M$ is a vector bundle with VB-atlas given by the various VB-charts U, ϕ restricted to $(\pi^{-1}U) \cap S$ and composed with projection $U \times \mathsf{E}_1 \times \{0\} \to U \times \mathsf{E}_1$ so $\pi|_L$ is a bundle locally isomorphic to $M \times \mathsf{E}_1$. The fiber of $\pi|_L$ at $p \in L$ is $L_p = E_p \cap L$. Once again we remind the reader of the somewhat unfortunate fact that although the bundle includes and is indeed determined by the map $\pi|_L : L \to M$ we often refer to L itself as the subbundle. In order to help the reader see what is going on here lets us look at how the definition of subbundle looks if we are in the finite dimensional case. We take $\mathsf{M} = \mathbb{R}^n$, $\mathsf{E} = \mathbb{R}^m$ and $\mathsf{E}_1 \times \mathsf{F}$ is the decomposition $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{m-k}$. Thus the bundle $\pi : E \to M$ has rank m (i.e. the typical fiber is $\mathbb{R}^m$) while the subbundle $\pi|_L : L \to M$ has rank k. The condition described in the definition of subbundle translates into there being a VB-chart $\phi : \pi^{-1}U \to U \times \mathbb{R}^k \times \mathbb{R}^{m-k}$ with $\phi((\pi^{-1}U) \cap L) = U \times \mathbb{R}^k \times \{0\}$. What if our original bundle was the trivial bundle $pr_1 : U \times \mathbb{R}^m \to U$? Then the our adapted chart must be a map $U \times \mathbb{R}^m \to U \times \mathbb{R}^k \times \mathbb{R}^{m-k}$ which must have the form $(x, v) \mapsto (x, f(x)v, 0)$ where for each x the $f(x)$ is a linear map $\mathbb{R}^m \to \mathbb{R}^k$.

26.12.2 Formulary

We now define the pseudogroup(s) relevant to the study of foliations. Let $\mathsf{M} = \mathsf{E} \times \mathsf{F}$ be a (split) Banach space. Define $\Gamma_{\mathsf{M},\mathsf{F}}$ to be the set of all diffeomorphisms from open subsets of $\mathsf{E} \times \mathsf{F}$ to open subsets of $\mathsf{E} \times \mathsf{F}$ of the form

$$\Phi(\mathsf{x}, \mathsf{y}) = (f(\mathsf{x}, \mathsf{y}), g(\mathsf{y})).$$

In case M is n dimensional and $\mathsf{M} = \mathbb{R}^n$ is decomposed as $\mathbb{R}^k \times \mathbb{R}^q$ we write $\Gamma_{\mathsf{M},\mathsf{F}} = \Gamma_{n,q}$. We can then the following definition:

Definition 26.83 *A $\Gamma_{\mathsf{M},\mathsf{F}}$ structure on a manifold M modelled on $\mathsf{M} = \mathsf{E} \times \mathsf{F}$ is a maximal atlas of charts satisfying the condition that the overlap maps are all members of $\Gamma_{\mathsf{M},\mathsf{F}}$.*

1) $\iota_{[X,Y]} = \mathcal{L}_X \circ \iota_Y + \iota_Y \circ \mathcal{L}_X$
2) $\mathcal{L}_{fX}\omega = f\mathcal{L}_X\omega + df \wedge \iota_X\omega$ for all $\omega \in \Omega(M)$
3) $d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge (d\beta)$
4) $\frac{d}{dt}\mathrm{Fl}_t^{X*}Y = \mathrm{Fl}_t^{X*}(L_X Y)$
5) $[X, Y] = \sum_{i,j=1}^m \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \right) \frac{\partial}{\partial x^i}$

Example 26.28 (Frame bundle) *Let M be a smooth manifold of dimension n. Let $F_x(M)$ denote the set of all bases (frames) for the vector space T_xM. Now let $F(M) := \bigcup_{x \in M} F_x(M)$. Define the natural projection $\pi : F(M) \to M$ by $\pi(\mathbf{f}) = x$ for all frames $\mathbf{f} = (f_i)$ for the space T_xM. It can be shown that $F(M)$ has a natural smooth structure. It is also a $\mathrm{GL}(n, \mathbb{R})$-bundle whose typical* *fiber is also* $\mathrm{GL}(n,\mathbb{R})$*. The bundle charts are built using the charts for* M *in the following way: Let* U_α, ψ_α *be a chart for* M*. Any frame* $\mathbf{f} = (f_i)$ *at some point* $x \in U_\alpha$ *may be written as*

$$f_i = \sum c_i^j \left.\frac{\partial}{\partial x^j}\right|_x .$$

We then map $\mathbf{f}$ *to* $(x, (c_i^j)) \in U_\alpha \times \mathrm{GL}(n,\mathbb{R})$*. This recipe gives a map* $\pi^{-1}(U_\alpha) \to U_\alpha \times \mathrm{GL}(n,\mathbb{R})$ *which is a bundle chart.*

Definition 26.84 *A bundle morphism* $(f, f_0) : \xi_1 \to \xi_2$ *from one fiber bundle* $\xi_1 = (E_1, \pi_{E_1}, M_1, F_1)$ *to another* $\xi_2 = (E_2, \pi_{E_2}, M_2, F_2)$ *is a pair of maps* (f, f_0) *such that the following diagram commutates*

$$\begin{array}{ccc} E_1 & \xrightarrow{f} & E_2 \\ \pi_{E_1}\downarrow & & \pi_{E_2}\downarrow \\ M_1 & \xrightarrow{f_0} & M_2 \end{array}$$

In case $M_1 = \quad M_2$ *and* $f_0 = \mathrm{id}_M$ *we call* f *a strong bundle morphism. In the latter case if* $f : E_1 \to E_2$ *is also a diffeomorphism then we call it a* ***bundle isomorphism.***

Definition 26.85 *Let* ξ_1 *and* ξ_2 *be fiber bundles with the same base space* M*. If there exists a bundle isomorphism* $(f, \mathrm{id}_M) : \xi_1 \to \xi_2$ *we say that* ξ_1 *and* ξ_2 *are* ***isomorphic as fiber bundles over*** M *and write* $\xi_1 \overset{fib}{\cong} \xi_2$*.*

Now given any open set U in the base space M of a fiber bundle $\xi = (E, \pi_E, M, F)$ we may form the restriction $\xi|\, U$ which is the fiber bundle $(\pi_E^{-1}(U), \pi_E|\, U, U, F)$. To simplify notation we write $E_U := \pi_E^{-1}(U)$ and $\pi_E|\, U := \pi_U$.

Example 26.29 *Let* M *and* F *be smooth manifolds and consider the projection map* $pr_1 : M \times F \to M$*. This is a smooth fiber bundle with typical fiber* F *and is called a* ***product bundle*** *or a* ***trivial bundle****.*

A fiber bundle which is isomorphic to a product bundle is also called a trivial bundle. The definition of a fiber bundle ξ with typical fiber F includes the existence of a cover of the base space by a family of open sets $\{U_\alpha\}_{\alpha \in A}$ such that $\xi|\, U_\alpha \overset{fib}{\cong} U \times F$ for all $\alpha \in A$. Thus, fiber bundles as we have defined them, are all locally trivial.

Misc

$1-$form $\theta = \sum e_j \theta^i$ which takes any vector to itself:

$$\begin{aligned} \theta(v_p) &= \sum e_j(p)\theta^i(v_p) \\ &= \sum v^i e_j(p) = v_p \end{aligned}$$

Let us write $d^\nabla\theta = \frac{1}{2}\sum e_k \otimes T^k_{ij}\theta^i \wedge \theta^j = \frac{1}{2}\sum e_k \otimes \tau^k$. If ∇ is the Levi Civita connection on M then consider the projection $P^\wedge : E \otimes TM \otimes T^*M$ given by $P^\wedge T(\xi, v) = T(\xi, v) - T(v, \xi)$. We have

$$\begin{aligned}\nabla e_j &= \omega^k_j e_k = e\omega \\ \nabla\theta^j &= -\omega^j_k \theta^k\end{aligned}$$

$\nabla_\xi(e_j \otimes \theta^j)$
$P^\wedge(\nabla_\xi\theta^j)(v) = -\omega^j_k(\xi)\theta^k(v) + \omega^j_k(v)\theta^k(\xi) = -\omega^j_k \wedge \theta^k$

Let $T(\xi, v) = \nabla_\xi(e_i \otimes \theta^j)(v)$
$= (\nabla_\xi e_i) \otimes \theta^j(v) + e_i \otimes (\nabla_\xi\theta^j)(v) = \omega^k_i(\xi)e_k \otimes \theta^j(v) + e_i \otimes (-\omega^j_k(\xi)\theta^k(v))$
$= \omega^k_i(\xi)e_k \otimes \theta^j(v) + e_i \otimes (-\omega^k_j(\xi)\theta^j(v)) = e_k \otimes (\omega^k_i(\xi) - \omega^k_j(\xi))\theta^j(v)$
Then

$$\begin{aligned}(P^\wedge T)(\xi, v) &= T(\xi, v) - T(v, \xi) \\ &= (\nabla e_j) \wedge \theta^j + e_j \otimes d\theta^j \\ &= d^\nabla(e_j \otimes \theta^j)\end{aligned}$$

$$\begin{aligned}d^\nabla\theta &= d^\nabla \sum e_j\theta^j \\ &= \sum(\nabla e_j) \wedge \theta^j + \sum e_j \otimes d\theta^j \\ &= \sum(\sum_k e_k \otimes \omega^k_j) \wedge \theta^j + \sum e_k \otimes d\theta^k \\ &= \sum_k e_k \otimes (\sum_j \omega^k_j \wedge \theta^j + d\theta^k)\end{aligned} \tag{26.8}$$

So that $\sum_j \omega^k_j \wedge \theta^j + d\theta^k = \frac{1}{2}\tau^k$. Now let $\sigma = \sum f^j e_j$ be a vector field

$$\begin{aligned}d^\nabla d^\nabla\sigma = d^\nabla\left(d^\nabla \sum e_j f^j\right) &= d^\nabla\left(\sum(\nabla e_j)f^j + \sum e_j \otimes df^j\right) \\ &\left(\sum(\nabla e_j)df^j + \sum(d^\nabla\nabla e_j)f^j + \sum \nabla e_j df^j + \sum e_j \otimes ddf^j\right)\end{aligned}$$

$$\sum f^j(d^\nabla\nabla e_j) = \sum f^j$$

So we seem to have a map $f^j e_j \mapsto \Omega^k_j f^j e_k$.

$$\begin{aligned}e_r\Omega^r_j &= d^\nabla\nabla e_j = d^\nabla(e_k\omega^k_j) \\ &= \nabla e_k \wedge \omega^k_j + e_k d\omega^k_j \\ &= e_r\omega^r_k \wedge \omega^k_j + e_k d\omega^k_j \\ &= e_r\omega^r_k \wedge \omega^k_j + e_r d\omega^r_j \\ &= e_r(d\omega^r_j + \omega^r_k \wedge \omega^k_j)\end{aligned}$$

$$d^{\nabla}\nabla e = d^{\nabla}(e\omega) = \nabla e \wedge \omega + ed\omega$$

From this we get $0 = d(A^{-1}A)A^{-1} = (dA^{-1})AA^{-1} + A^{-1}dAA^{-1}$
$dA^{-1} = A^{-1}dAA^{-1}$

$$\begin{aligned}
\Omega^r_j &= d\omega^r_j + \omega^r_k \wedge \omega^k_j \\
\Omega &= d\omega + \omega \wedge \omega \\
\Omega' &= d\omega' + \omega' \wedge \omega' \\
\Omega' &= d\left(A^{-1}\omega A + A^{-1}dA\right) + \left(A^{-1}\omega A + A^{-1}dA\right) \wedge \left(A^{-1}\omega A + A^{-1}dA\right) \\
&= d\left(A^{-1}\omega A\right) + d\left(A^{-1}dA\right) + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1} \wedge \omega A + A^{-1}dA \wedge A^{-1}dA \\
&= d\left(A^{-1}\omega A\right) + dA^{-1} \wedge dA + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1}\omega A + A^{-1}dA \wedge A^{-1}dA \\
&= dA^{-1}\omega A + A^{-1}d\omega A - A^{-1}\omega dA + dA^{-1} \wedge dA + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1} \wedge \omega A + A^{-1}dA \wedge A^{-1}dA \\
&= A^{-1}d\omega A + A^{-1}\omega \wedge \omega A \\
\Omega' &= A^{-1}\Omega A
\end{aligned}$$

$\omega' = A^{-1}\omega A + A^{-1}dA$

These are interesting equations let us approach things from a more familiar setting so as to interpret what we have.

26.13 Curvature

An important fact about covariant derivatives is that they don't need to commute. If $\sigma : M \to E$ is a section and $X \in \mathfrak{X}(M)$ then $\nabla_X\sigma$ is a section also and so we may take it's covariant derivative $\nabla_Y\nabla_X\sigma$ with respect to some $Y \in \mathfrak{X}(M)$. In general, $\nabla_Y\nabla_X\sigma \neq \nabla_X\nabla_Y\sigma$ and this fact has an underlying geometric interpretation which we will explore later. A measure of this lack of commutativity is the **curvature operator** which is defined for a pair $X, Y \in \mathfrak{X}(M)$ to be the map $F(X,Y) : \Gamma(E) \to \Gamma(E)$ defined by

$$F(X,Y)\sigma := \nabla_X\nabla_Y\sigma - \nabla_Y\nabla_X\sigma - \nabla_{[X,Y]}\sigma.$$

or

$$[\nabla_X, \nabla_Y]\sigma - \nabla_{[X,Y]}\sigma$$

26.14 Group action

$\rho : G \times M \to M$

$\rho(g,x)$ as gx then require that

Exercise 26.10 *1) $g_1(g_2x) = (g_1g_2)x$ for all $g_1, g_2 \in G$ and for all $x \in M$*
2) $ex = x$ for all $x \in M$,
3) the map (x, g)

26.15 Notation and font usage guide

Category	Space or object	Typical elements	Typical morphisms
Vector Spaces	V , W , $\mathbb{R}^n$	v,w, x, y	A, B, K, λ, L
Banach Spaces	$\mathsf{E,F,M,N,V,W}$, $\mathbb{R}^n$	$\mathsf{v,w,x,y}$ etc.	A, B, K, λ, L
Open sets in vector spaces	U, V, O, U_α	$\mathsf{p,q,x,y,v,w}$	f, g, φ, ψ
Differentiable manifolds	M, N, P, Q	p, q, x, y	f, g, φ, ψ
Open sets in manifolds	U, V, O, U_α	p, q, x, y	$f, g, \varphi, \psi, \mathtt{x}_\alpha$
Bundles	$E \to M$	v, w, ξ, p, q, x	$(\bar{f}, f), (g, id), h$
Sections of bundles	$\Gamma(M, E)$	$s, s_1, \sigma, ...$	f^*
Sections over open sets	$\Gamma(U, E) = \mathcal{S}_M^E(U)$	$s, s_1, \sigma, ...$	f^*
Lie Groups	G, H, K	g, h, x, y	h, f, g
Lie Algebras	$\mathfrak{g}, \mathfrak{h}, \mathfrak{k}, \mathfrak{a}, \mathfrak{b}$	v, x, y, z, ξ	h, g, df, dh
Fields	$\mathbb{F}, \mathbb{R}, \mathbb{C}, \mathbb{K}$	t, s, x, y, z, r	f, g, h
Vector Fields	$\mathfrak{X}_M(U)$,$\mathfrak{X}(M)$	X, Y, Z	f^*, f_*??

So, as we said, after imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers. We will envision there to be a copy $\mathbb{R}_p^n$ of $\mathbb{R}^n$ at each of its points $p \in \mathbb{R}^n$. The elements of $\mathbb{R}_p^n$ are to be thought of as the vectors based at p, that is, the "tangent vectors". These tangent spaces are related to each other by the obvious notion of vectors being parallel (this is exactly what is not generally possible for tangents spaces of a manifold). For the standard basis vectors e_j (relative to the coordinates x_i) taken as being based at p we often write $\left.\frac{\partial}{\partial x_i}\right|_p$ and this has the convenient second interpretation as a differential operator acting on C^∞functions defined near $p \in \mathbb{R}^n$. Namely,

$$\left.\frac{\partial}{\partial x_i}\right|_p f = \frac{\partial f}{\partial x_i}(p).$$

An n-tuple of C^∞functions $X^1, ..., X^n$ defines a C^∞vector field $X = \sum X^i \frac{\partial}{\partial x_i}$ whose value at p is $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$. Thus a vector field assigns to each p in its domain, an open set U, a vector $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$ at p. We may also think of vector field as a differential operator via

$$f \mapsto Xf \in C^\infty(U)$$
$$(Xf)(p) := \sum X^i(p) \frac{\partial f}{\partial x_i}(p)$$

Example 26.30 *$X = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}$ is a vector field defined on $U = \mathbb{R}^2 - \{0\}$ and $(Xf)(x, y) = y\frac{\partial f}{\partial x}(x, y) - x\frac{\partial f}{\partial y}(x, y)$.*

Notice that we may certainly add vector fields defined over the same open set as well as multiply by functions defined there:

$$(fX + gY)(p) = f(p)X(p) + g(p)X(p)$$

The familiar expression $df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n$ has the intuitive interpretation expressing how small changes in the variables of a function give rise to small changes in the value of the function. Two questions should come to mind. First, "what does 'small' mean and how small is small enough?" Second, "which direction are we moving in the coordinate" space? The answer to these questions lead to the more sophisticated interpretation of df as being a linear functional on each tangent space. Thus we must choose a direction v_p at $p \in \mathbb{R}^n$ and then $df(v_p)$ is a number depending linearly on our choice of vector v_p. The definition is determined by $dx_i(e_j) = \delta_{ij}$. In fact, this shall be the basis of our definition of df at p. We want

$$Df|_p \left(\left. \frac{\partial}{\partial x_i} \right|_p \right) := \frac{\partial f}{\partial x_i}(p).$$

Now any vector at p may be written $v_p = \sum_{i=1}^n v^i \left. \frac{\partial}{\partial x_i} \right|_p$ which invites us to use v_p as a differential operator (at p):

$$v_p f := \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p) \in \mathbb{R}$$

This consistent with our previous statement about a vector field being a differential operator simply because $X(p) = X_p$ is a vector at p for every $p \in U$. This is just the directional derivative. In fact we also see that

$$\begin{aligned} Df|_p (v_p) &= \sum_j \frac{\partial f}{\partial x_j}(p) dx_j \left(\sum_{i=1}^n v^i \left. \frac{\partial}{\partial x_i} \right|_p \right) \\ &= \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p) = v_p f \end{aligned}$$

so that our choices lead to the following definition:

Definition 26.86 *Let f be a C^∞function on an open subset U of $\mathbb{R}^n$. By the symbol df we mean a family of maps $Df|_p$ with p varying over the domain U of f and where each such map is a linear functional of tangent vectors based at p given by $Df|_p (v_p) = v_p f = \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p)$.*

Definition 26.87 *More generally, a smooth 1-form α on U is a family of linear functionals $\alpha_p : T_p\mathbb{R}^n \to \mathbb{R}$ with $p \in U$ which is smooth is the sense that $\alpha_p(\left. \frac{\partial}{\partial x_i} \right|_p)$ is a smooth function of p for all i.*

From this last definition it follows that if $X = X^i \frac{\partial}{\partial x_i}$ is a smooth vector field then $\alpha(X)(p) := \alpha_p(X_p)$ defines a smooth function of p. Thus an alternative way to view a $1-$form is as a map $\alpha : X \mapsto \alpha(X)$ which is defined on vector fields and linear over the algebra of smooth functions $C^\infty(U)$:

$$\alpha(fX + gY) = f\alpha(X) + g\alpha(Y).$$

Fixing a problem. It is at this point that we want to destroy the privilege of the rectangular coordinates and express our objects in an arbitrary coordinate system smoothly related to the existing coordinates. This means that for any two such coordinate systems, say $u^1, ..., u^n$ and $y^1, ..., y^n$ we want to have the ability to express fields and forms in either system and have for instance

$$X^i_{(y)} \frac{\partial}{\partial y_i} = X = X^i_{(u)} \frac{\partial}{\partial u_i}$$

for appropriate functions $X^i_{(y)}$, $X^i_{(u)}$. This equation only makes sense on the overlap of the domains of the coordinate systems. To be consistent with the chain rule we must have

$$\frac{\partial}{\partial y^i} = \frac{\partial u^j}{\partial y^i} \frac{\partial}{\partial u^j}$$

which then forces the familiar transformation law:

$$\sum \frac{\partial u^j}{\partial y^i} X^i_{(y)} = X^i_{(u)}$$

We think of $X^i_{(y)}$ and $X^i_{(u)}$ as referring to, or representing, the same geometric reality from the point of view of two different coordinate systems. No big deal right? Well, how about the fact that there is this underlying abstract space that we are coordinatizing? That too is no big deal. We were always doing it in calculus anyway. What about the fact that the coordinate systems aren't defined as a 1-1 correspondence with the points of the space unless we leave out some points in the space? For example, polar coordinates must exclude the positive x-axis and the origin in order to avoid ambiguity in θ and have a nice open domain. Well if this is all fine then we may as well imagine other abstract spaces that support coordinates in this way. This is manifold theory. We don't have to look far for an example of a manifold other than Euclidean space. Any surface such as the sphere will do. We can talk about 1-forms like say $\alpha = \theta d\phi + \phi \sin(\theta) d\theta$, or a vector field tangent to the sphere $\theta \sin(\phi) \frac{\partial}{\partial \theta} + \theta^2 \frac{\partial}{\partial \phi}$ and so on (just pulling things out of a hat). We just have to be clear about how these arise and most of all how to change to a new coordinate expression for the same object. This is the approach of tensor analysis. An object called a 2-tensor T is represented in two different coordinate systems as for instance

$$\sum T^{ij}_{(y)} \frac{\partial}{\partial y^i} \otimes \frac{\partial}{\partial y^j} = \sum T^{ij}_{(u)} \frac{\partial}{\partial u^i} \otimes \frac{\partial}{\partial u^j}$$

where all we really need to know for many purposes the transformation law

$$T^{ij}_{(y)} = \sum_{r,s} T^{rs}_{(u)} \frac{\partial y^i}{\partial u^r} \frac{\partial y^i}{\partial u^s}.$$

Then either expression is referring to the same abstract tensor T. This is just a preview but it highlight the approach wherein a transformation laws play a defining role.

In order to understand modern physics and some of the best mathematics it is necessary to introduce the notion of a space (or spacetime) which only locally has the (topological) features of a vector space like $\mathbb{R}^n$. Examples of two dimensional manifolds include the sphere or any of the other closed smooth surfaces in $\mathbb{R}^3$ such a torus. These are each locally like $\mathbb{R}^2$ and when sitting in space in a nice smooth way like we usually picture them, they support coordinates systems which allow us to do calculus on them. The reader will no doubt be comfortable with the idea that it makes sense to talk about directional rates of change in say a temperature distribution on a sphere representing the earth.

For a higher dimensional example we have the $3-$sphere S^3 which is the hypersurface in $\mathbb{R}^4$ given by the equation $x^2 + y^2 + z^2 + w^2 = 1$.

For various reasons, we would like coordinate functions to be defined on open sets. It is not possible to define nice coordinates on closed surfaces like the sphere which are defined on the whole surface. By nice we mean that together the coordinate functions, say, θ, ϕ should define a 1-1 correspondence with a subset of $\mathbb{R}^2$ which is continuous and has a continuous inverse. In general the best we can do is introduce several coordinate systems each defined on separate open subsets which together cover the surface. This will be the general idea for all manifolds.

Now suppose that we have some surface S and two coordinate systems

$$(\theta, \phi) : U_1 \to \mathbb{R}^2$$
$$(u, v) : U_2 \to \mathbb{R}^2$$

Imagine a real valued function f defined on S (think of f as a temperature or something). Now if we write this function in coordinates (θ, ϕ) we have f represented by a function of two variables $f_1(\theta, \phi)$ and we may ask if this function is differentiable or not. On the other hand, f is given in (u, v) coordinates by a representative function $f_2(u, v)$. In order that our conclusions about differentiability at some point $p \in U_1 \cap U_2 \subset S$ should not depend on what coordinate system we use we had better have the coordinate systems themselves related differentiably. That is, we want the coordinate change functions in both directions to be differentiable. For example we may then relate the derivatives as they appear in different coordinates by chain rules expressions like

$$\frac{\partial f_1}{\partial \theta} = \frac{\partial f_2}{\partial u} \frac{\partial u}{\partial \theta} + \frac{\partial f_2}{\partial v} \frac{\partial v}{\partial \theta}$$

which have validity on coordinate overlaps. The simplest and most useful condition to require is that coordinates systems have C^∞ coordinate changes on the overlaps.

Definition 26.88 *A set M is called a C^∞ differentiable manifold of dimension n if M is covered by the domains of some family of coordinate mappings or charts $\{\mathtt{x}_\alpha : U_\alpha \to \mathbb{R}^n\}_{\alpha\in A}$ where $\mathtt{x}_\alpha = (x_\alpha^1, x_\alpha^2, ...x_\alpha^n)$. We require that the coordinate change maps $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ are continuously differentiable any number of times on their natural domains in $\mathbb{R}^n$. In other words, we require that the functions*

$$\begin{aligned} x^1 &= x_\beta^1(x_\alpha^1, ..., x_\alpha^n) \\ x_\beta^2 &= x_\beta^2(x_\alpha^1, ..., x_\alpha^n) \\ &\vdots \\ x_\beta^n &= x_\beta^n(x_\alpha^1, ..., x_\alpha^n) \end{aligned}$$

together give a C^∞ bijection where defined. The α and β are just indices from some set A and are just a notational convenience for naming the individual charts.

Note that we are employing the same type of abbreviations and abuse of notation as is common is every course on calculus where we often write things like $y = y(x)$. Namely, $(x_\alpha^1, ..., x_\alpha^n)$ denotes both an n-tuple of coordinate functions and an element of $\mathbb{R}^n$. Also, $x_\beta^1 = x_\beta^1(x_\alpha^1, ..., x_\alpha^n)$ etc. could be thought of as an abbreviation for a functional relation which when evaluated at a point p on the manifold reads

$$(x_\beta^1(p), ...x_\beta^n(p)) = \mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}(x_\alpha^1(p), ..., x_\alpha^n(p)).$$

A function f on M will be deemed to be C^r if its representatives f_α are all C^r for every coordinate system $\mathtt{x}_\alpha = (x_\alpha^1, ..., x_\alpha^n)$ whose domain intersects the domain of f. Now recall our example of temperature on a surface. For an arbitrary pair of coordinate systems $\mathtt{x} = (x^1, ..., x^n)$ and $\mathtt{y} = (y^1, ..., y^n)$ the functions $f_1 := f \circ \mathtt{x}^{-1}$ and $f_2 := f \circ \mathtt{y}^{-1}$ represent the same function f with in the coordinate domains but the expressions $\frac{\partial f_1}{\partial x^i}$ and $\frac{\partial f_2}{\partial y^i}$ are not equal and do not refer to the same physical or geometric reality. The point is simply that because of our requirements on the smooth relatedness of our coordinate systems we know that on the overlap of the two coordinate systems if $f \circ \mathtt{x}^{-1}$ has continuous partial derivatives up to order k then the same will be true of $f \circ \mathtt{y}^{-1}$.

Also we have the following notations

$C^\infty(U)$ or $\mathcal{F}(U)$	Smooth functions on U
$C_c^\infty(U)$ or $\mathcal{D}(U)$	"...." with compact support in U
T_pM	Tangent space at p
TM, with $\tau_M : TM \to M$	Tangent bundle of M
$T_pf : T_pM \to T_{f(p)}N$	Tangent map of $f : M \to N$ at p
$Tf : TM \to TN$	Tangent map of $f : M \to N$
T_p^*M	Cotangent space at p
T^*M, with $\pi_M : T^*M \to M$	Cotangent bundle of M
$J_x(M,N)_y$	k-jets of maps $f :: M, x \to N, y$
$\mathfrak{X}(U), \mathfrak{X}_M(U)$ (or $\mathfrak{X}(M)$)	Vector field over U (or over M)
$(\mathtt{x}, U), (\mathtt{x}_\alpha, U_\alpha), (\psi_\beta, U_\beta), (\varphi, U)$	Typical charts
$T_s^r(\mathrm{V})$ r-contravariant s-covariant	Tensors on V
$\mathfrak{T}_s^r(M)$ r-contravariant s-covariant	Tensor fields on M

d	exterior derivative, differential
∇	covariant derivative

M, g	Riemannian manifold with metric tensor g
M, ω	Symplectic manifold with symplectic form ω
$L(\mathrm{V}, \mathrm{W})$	Linear maps from V to W. Assumed bounded if V, W are Banach
$L_s^r(\mathrm{V}, \mathrm{W})$	r-contravariant, s-covariant multilinear maps $\mathrm{V}^{*r} \times \mathrm{V}^s \to \mathrm{W}$

For example, let V be a vector space with a basis $\mathbf{f}_1, ..., \mathbf{f}_n$ and dual basis $\mathbf{f}^1, ..., \mathbf{f}^n$ for V^*. Then we can write an arbitrary element $v \in \mathrm{V}$ variously by

$$
\begin{aligned}
v &= v^1\mathbf{f}_1 + \cdots + v^n\mathbf{f}_n = \sum v^i\mathbf{f}_i \\
&= (\mathbf{f}_1, ..., \mathbf{f}_n) \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}
\end{aligned}
$$

while $\alpha \in \mathrm{V}^*$ would usually be written as one of the following

$$
\begin{aligned}
\alpha &= a_1\mathbf{f}^1 + \cdots + a_n\mathbf{f}^n = \sum \alpha_i\mathbf{f}^i \\
&= (a_1, ..., a_n) \begin{pmatrix} \mathbf{f}^1 \\ \vdots \\ \mathbf{f}^n \end{pmatrix}.
\end{aligned}
$$

We also sometimes use the convention that when an index is repeated once up and once down then a summation is implied. For example, $\alpha(v) = a_iv^i$ means $\alpha(v) = \sum_i a_iv^i$.

Another useful convention that we will use often is that when we have a list of objects $(o_1, ..., o_N)$ then $(o_1, ..., \widehat{o_i}, ..., o_N)$ will mean the same list with the i-th object omitted.

Finally, in some situations a linear function A of a variable, say h, is written as Ah or $A \cdot h$ instead of $A(h)$. This notation is particularly useful when we have a family of linear maps depending on a point in some parameter space. For example, the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ at a point $x \in \mathbb{R}^n$ is a linear map $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ and as such we may apply it to a vector $h \in \mathbb{R}^n$.

But to write $Df(x)(h)$ is a bit confusing and so we write $Df(x) \cdot h$ or $Df|_x h$ to clarify the different roles of the variables x and h. As another example, if $x \to A(x)$ is an $m \times n$ matrix valued function we might write $A_x h$ for the matrix multiplication of $A(x)$ and $h \in \mathbb{R}^n$.

In keeping with this we will later think of the space $L(\mathrm{V}, \mathrm{W})$ of linear maps from V to W as being identified with $\mathrm{W} \otimes \mathrm{V}^*$ rather than $\mathrm{V}^* \otimes \mathrm{W}$ (this notation will be explained in detail). Thus $(w \otimes \alpha)(v) = w\alpha(v) = \alpha(v)w$. This works nicely since if $w = (w^1, ..., w^m)^t$ is a column and $\alpha = (a_1, ..., a_n)$ then the linear transformation $w \otimes \alpha$ defined above has as matrix

Chapter 27

Bibliography

Bibliography

[A] J. F. Adams, Stable Homotopy and Generalized Homology, Univ. of Chicago Press, 1974.

[Arm] M. A. Armstrong, Basic Topology, Springer-Verlag, 1983.

[At] M. F. Atiyah, K-Theory, W.A.Benjamin, 1967.

[A,B,R] Abraham, R., Marsden, J.E., and Ratiu, T., *Manifolds, tensor analysis, and applications*, Addison Wesley, Reading, 1983.

[Arn] Arnold, V.I., *Mathematical methods of classical mechanics*, Graduate Texts in Math. **60**, Springer-Verlag, New York, 2nd edition (1989).

[A] Alekseev, A.Y., *On Poisson actions of compact Lie groups on symplectic manifolds*, J. Diff. Geom. **45** (1997), 241-256.

[Bott and Tu]

[Bry]

[Ben] D. J. Benson, *Representations and Cohomology, Volume II: Cohomology of Groups and Modules,* Cambridge Univ. Press, 1992.

[1] R. Bott and L. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag GTM 82,1982.

[Bre] G. Bredon, *Topology and Geometry*, Springer-Verlag GTM 139, 1993.

[Chav1]

[Chav2]

[Drin] Drinfel'd, V.G., *On Poisson homogeneous spaces of Poisson-Lie groups*, Theor. Math. Phys. **95** (1993), 524-525.

[Dieu] J. Dieudonn´e, *A History of Algebraic and Differential Topology* 1900-1960, Birkh¨auser,1989.

[Do] A. Dold, *Lectures on Algebraic Topology*, Springer-Verlag, 1980.

[Dug] J. Dugundji, *Topology*, Allyn & Bacon, 1966.

[Eil,St] S. Eilenberg and N. Steenrod, *Foundations of Algebraic Topology*, Princeton Univ. Press, 1952.

[Fen] R. Fenn, *Techniques of Geometric Topology*, Cambridge Univ. Press, 1983.

[Fr,Q] M. Freedman and F. Quinn, *Topology of 4-Manifolds*, Princeton Univ. Press, 1990.

[Fult] W. Fulton, *Algebraic Topology: A First Course*, Springer-Verlag, 1995.

[G1] Guillemin, V., and Sternberg, S., *Convexity properties of the moment mapping*, Invent. Math. 67 (1982), 491-513.

[G2] Guillemin, V., and Sternberg, S., *Symplectic Techniques in Physics*, Cambridge Univ. Press, Cambridge, 1984.

[Gu,Hu,We] Guruprasad, K., Huebschmann, J., Jeffrey, L., and Weinstein, A., *Group systems, groupoids, and moduli spaces of parabolic bundles*, Duke Math. J. 89 (1997), 377-412.

[Gray] B. Gray, *Homotopy Theory*, Academic Press, 1975.

[Gre,Hrp] M. Greenberg and J. Harper, *Algebraic Topology: A First Course*, Addison-Wesley, 1981.

[2] P. J. Hilton, *An Introduction to Homotopy Theory*, Cambridge University Press, 1953.

[Hilt2] P. J. Hilton and U. Stammbach, *A Course in Homological Algebra*, Springer-Verlag, 1970.

[Huss] D. Husemoller, *Fibre Bundles*, McGraw-Hill, 1966 (later editions by Springer-Verlag).

[Hu] Huebschmann, J., *Poisson cohomology and quantization*, J. Reine Angew. Math. **408** (1990), 57-113.

[KM]

[Kirb,Seib] R. Kirby and L. Siebenmann, *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*, Ann. of Math.Studies 88, 1977.

[L1] Lang, S. *Foundations of Differential Geometry*, Springer-Verlag GTN vol 191

[M,T,W] Misner,C. Wheeler, J. and Thorne, K. *Gravitation*, Freeman 1974

[Mil] Milnor, J., *Morse Theory*, Annals of Mathematics Studies **51**, Princeton U. Press, Princeton, 1963.

[MacL] S. MacLane, *Categories for the Working Mathematician*, Springer-Verlag GTM 5, 1971.

[Mass] W. Massey, *Algebraic Topology: An Introduction*, Harcourt, Brace & World, 1967 (reprinted by Springer-Verlag).

[Mass2] W. Massey, *A Basic Course in Algebraic Topology*, Springer-Verlag, 1993.

[Maun] C. R. F. Maunder, *Algebraic Topology*, Cambridge Univ. Press, 1980 (reprinted by Dover Publications).

[Miln1] J. Milnor, *Topology from the Differentiable Viewpoint*, Univ. Press of Virginia, 1965.

[Mil,St] J. Milnor and J. Stasheff, *Characteristic Classes*, Ann. of Math. Studies 76, 1974.

[Roe] Roe,J. Elliptic Operators, *Topology and Asymptotic methods*, Longman, 1988

[Spv] Spivak, M. *A Comprehensive Introduction to Differential Geometry*, (5 volumes) Publish or Perish Press, 1979.

[St] Steenrod, N. *Topology of fiber bundles*, Princeton University Press, 1951.

[Va] Vaisman, I., *Lectures on the Geometry of Poisson Manifolds*, Birkhäuser, Basel, 1994.

[We1] Weinstein, A., *Lectures on Symplectic Manifolds*, Regional conference series in mathematics **29**, Amer. Math. Soc.,Providence,1977.

[We2] Weinstein, A., *The local structure of Poisson manifolds*, J. Diff. Geom. **18** (1983), 523–557.

[We3] Weinstein, A., *Poisson structures and Lie algebras*, Astérisque, hors série (1985), 421–434.

[We4] Weinstein, A., *Groupoids: Unifying Internal and External Symmetry*, Notices of the AMS, July 1996.

[3] J. A. Álvarez López, *The basic component of the mean curvature of Riemannian foliations,* Ann. Global Anal. Geom. **10**(1992), 179–194.

[BishCr] R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds*, New York:Academic Press, 1964.

[Chavel] I. Chavel, *Eigenvalues in Riemannian Geometry,* Orlando: Academic Press, 1984.

[Cheeger] J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in analysis (Papers dedicated to Salomon Bochner, 1969), pp. 195–199, Princeton, N. J.: Princeton Univ. Press, 1970.

[ChEbin] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, Amsterdam: North-Holland, 1975.

[Cheng] S.Y. Cheng, *Eigenvalue Comparison Theorems and its Geometric Applications* Math. Z. **143**, 289–297, (1975).

[El-KacimiHector] A. El Kacimi–Alaoui and G. Hector, *Décomposition de Hodge basique pour un feuilletage Riemannien,* Ann. Inst. Fourier, Grenoble **36**(1986), no. 3 , 207–227.

[Gray2] A. Gray *Comparison Theorems for the Volumes of Tubes as Generalizations of the Weyl Tube Formula* Topology **21, no. 2**, 201–228, (1982).

[HeKa] E. Heintz and H. Karcher, *A General Comparison Theorem with Applications to Volume estimates for Submanifolds* Ann. scient. Ěc. Norm Sup., 4^e sėrie t. **11**, 451–470, (1978).

[KamberTondeur] F. W. Kamber and Ph. Tondeur, *De Rham-Hodge theory for Riemannian foliations*, Math. Ann. **277**(1987), 415–431.

[Lee] J. Lee, *Eigenvalue Comparison for Tubular Domains* Proc.of the Amer. Math. Soc. **109 no. 3**(1990).

[Min-OoRuhTondeur] M. Min-Oo, E. A. Ruh, and Ph. Tondeur, *Vanishing theorems for the basic cohomology of Riemannian foliations*, J. reine angew. Math. **415**(1991), 167–174.

[Molino] P. Molino, *Riemannian foliations,* Progress in Mathematics, Boston:Birkhauser, 1988.

[NishTondeurVanh] S. Nishikawa, M. Ramachandran, and Ph. Tondeur, *The heat equation for Riemannian foliations,* Trans. Amer. Math. Soc. **319**(1990), 619–630.

[O'Neill] B. O'Neill, *Semi-Riemannian Geometry*, New York:Academic Press, 1983.

[PaRi] E. Park and K. Richardson, *The Basic Laplacian of a Riemannian foliation*, Amer. J. Math. **118**(1996), no. 6, pp. 1249–1275.

[Ri1] K. Richardson, *The asymptotics of heat kernels on Riemannian foliations*, to appear in Geom. Funct. Anal.

[Ri2] K. Richardson, *Traces of heat kernels on Riemannian foliations*, preprint.

[Tondeur1] Ph. Tondeur, *Foliations on Riemannian manifolds,* New York:Springer Verlag, 1988.

[Tondeur2] Ph. Tondeur, *Geometry of Foliations*, Monographs in Mathematics, vol. 90, Basel: Birkhäuser, 1997.