Online Supplement fo Manifolds and Differential Geometry

Jeffrey M. Lee

Contents

Chapter 1

Calculus

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.

1.1 Derivative

Modern differential geometry is based on the theory of differentiable manifolds- a natural extension of multivariable calculus. Multivariable calculus is said to be done on (or in) an n-dimensional coordinate space $\mathbb{R}^n$ (also called variously "Euclidean space" or sometimes "Cartesian space". We hope that the great majority of readers will be comfortable with standard multivariable calculus. A reader who felt the need for a review could do no better than to study the classic book "Calculus on Manifolds" by Michael Spivak. This book does multivariable calculus[1] in a way suitable for modern differential geometry. It also has the virtue of being short. On the other hand, calculus easily generalizes from $\mathbb{R}^n$ to Banach spaces (a nice class of infinite dimensional vector spaces). We will recall a few definitions and facts from functional analysis and then review highlights from differential calculus while simultaneously generalizing to Banach spaces.

A **topological vector space** over $\mathbb{R}$ is a vector space V with a topology such that vector addition and scalar multiplication are continuous. This means that the map from $\mathsf{V} \times \mathsf{V}$ to V given by $(v_1, v_2) \mapsto v_1 + v_2$ and the map from $\mathbb{R} \times \mathsf{V}$ to V given by $(a, v) \mapsto av$ are continuous maps. Here we have given $\mathsf{V} \times \mathsf{V}$ and $\mathbb{R} \times \mathsf{V}$ the product topologies.

[1] Despite the title, most of Spivak's book is about calculus rather than manifolds.

Definition 1.1 *A map between topological vector spaces which is both a continuous linear map and which has a continuous linear inverse is called a* ***toplinear isomorphism****.*

A toplinear isomorphism is then just a linear isomorphism which is also a homeomorphism.

We will be interested in topological vector spaces which get their topology from a norm function:

Definition 1.2 *A* ***norm*** *on a real vector space* V *is a map* $\|.\| : \mathsf{V} \to \mathbb{R}$ *such that the following hold true:*
i) $\|v\| \geq 0$ *for all* $v \in \mathsf{V}$ *and* $\|v\| = 0$ *only if* $v = 0$.
ii) $\|av\| = |a| \, \|v\|$ *for all* $a \in \mathbb{R}$ *and all* $v \in \mathsf{V}$.
iii) If $v_1, v_2 \in \mathsf{V}$*, then* $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ *(triangle inequality). A vector space together with a norm is called a* ***normed vector space****.*

Definition 1.3 *Let* E *and* F *be normed spaces. A linear map* $A : \mathsf{E} \longrightarrow \mathsf{F}$ *is said to be bounded if*

$$\|A(v)\| \leq C \, \|v\|$$

for all $v \in \mathsf{E}$*. For convenience, we have used the same notation for the norms in both spaces. If* $\|A(v)\| = \|v\|$ *for all* $v \in \mathsf{E}$ *we call* A *an* ***isometry****. If* A *is a toplinear isomorphism which is also an isometry we say that* A *is an* ***isometric isomorphism.***

It is a standard fact that a linear map between normed spaces is bounded if and only if it is continuous.

The standard norm for $\mathbb{R}^n$ is given by $\left\|(x^1, ..., x^n)\right\| = \sqrt{\sum_{i=1}^n (x^i)^2}$. Imitating what we do in $\mathbb{R}^n$ we can define a distance function for a normed vector space by letting $\operatorname{dist}(v_1, v_2) := \|v_2 - v_1\|$. The distance function gives a topology in the usual way. The convergence of a sequence is defined with respect to the distance function. A sequence $\{v_i\}$ is said to be a **Cauchy sequence** if given any $\varepsilon > 0$ there is an N such that $\operatorname{dist}(v_n, v_m) = \|v_n - v_m\| < \varepsilon$ whenever $n, m > N$. In $\mathbb{R}^n$ every Cauchy sequence is a convergent sequence. This is a good property with many consequences.

Definition 1.4 *A normed vector space with the property that every Cauchy sequence converges is called a complete normed vector space or a* ***Banach space****.*

Note that if we restrict the norm on a Banach space to a closed subspace then that subspace itself becomes a Banach space. This is not true unless the subspace is closed.

Consider two Banach spaces V and W. A continuous map $A : \mathsf{V} \to \mathsf{W}$ which is also a linear isomorphism can be shown to have a continuous linear inverse. In other words, A is a toplinear isomorphism.

Even though some aspects of calculus can be generalized without problems for fairly general spaces, the most general case that we shall consider is the case of a Banach space.

What we have defined are real normed vector spaces and real Banach space but there is also the easily defined notion of complex normed spaces and complex Banach spaces. In functional analysis the complex case is central but for calculus it is the real Banach spaces that are central. Of course, every complex Banach space is also a real Banach space in an obvious way. For simplicity and definiteness all normed spaces and Banach spaces in this chapter will be real Banach spaces as defined above. Given two normed spaces V and W with norms $\|.\|_1$ and $\|.\|_2$ we can form a normed space from the Cartesian product $\mathsf{V} \times \mathsf{W}$ by using the norm $\|(v, w)\| := \max\{\|v\|_1, \|w\|_2\}$. The vector space structure on $\mathsf{V} \times \mathsf{W}$ is that of the (outer) direct sum.

Two norms on V, say $\|.\|'$and $\|.\|''$ are equivalent if there exist positive constants c and C such that

$$c\|x\|' \leq \|x\|'' \leq C\|x\|'$$

for all $x \in \mathsf{V}$. There are many norms for $\mathsf{V} \times \mathsf{W}$ equivalent to that given above including

$$\|(v, w)\|' := \sqrt{\|v\|_1^2 + \|w\|_2^2}$$

and also

$$\|(v, w)\|'' := \|v\|_1 + \|w\|_2 .$$

If V and W are Banach spaces then so is $\mathsf{V} \times \mathsf{W}$ with either of the above norms. The topology induced on $\mathsf{V} \times \mathsf{W}$ by any of these equivalent norms is exactly the product topology.

Let W_1 and W_2 be subspaces of a Banach space V. We write $\mathsf{W}_1 + \mathsf{W}_2$ to indicate the subspace

$$\{v \in \mathsf{V} : v = w_1 + w_2 \text{ for } w_1 \in \mathsf{W}_1 \text{ and } w_2 \in \mathsf{W}_2\}$$

If $\mathsf{V} = \mathsf{W}_1 + \mathsf{W}_2$ then any $v \in V$ can be written as $v = w_1 + w_2$ for $w_1 \in \mathsf{W}_1$ and $w_2 \in \mathsf{W}_2$. If furthermore $\mathsf{W}_1 \cap \mathsf{W}_2 = 0$, then this decomposition is unique and we say that W_1 and W_2 are complementary. Now unless a subspace is closed it will itself not be a Banach space and so if we are given a closed subspace W_1 of V then it is ideal if there can be found a subspace W_2 which is complementary to W_1 and which is *also closed*. In this case we write $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2$. One can use the closed graph theorem to prove the following.

Theorem 1.5 *If* W_1 *and* W_2 *are complementary closed subspaces of a Banach space* V *then there is a toplinear isomorphism* $\mathsf{W}_1 \times \mathsf{W}_2 \cong \mathsf{V}$ *given by*

$$(w_1, w_2) \longleftrightarrow w_1 + w_2.$$

When it is convenient, we can identify $\mathsf{W}_1 \oplus \mathsf{W}_2$ with $\mathsf{W}_1 \times \mathsf{W}_2$.

Let E be a Banach space and $\mathsf{W} \subset \mathsf{E}$ a *closed* subspace. If there is a *closed* complementary subspace W' say that W is a **split subspace** of E. The reason why it is important for a subspace to be split is because then we can use the isomorphism $\mathsf{W} \times \mathsf{W}' \cong \mathsf{W} \oplus \mathsf{W}'$. This will be an important technical consideration in the sequel.

Definition 1.6 (Notation) *We will denote the set of all continuous (bounded) linear maps from a normed space* E *to a normed space* F *by* $L(\mathsf{E},\mathsf{F})$*. The set of all continuous linear isomorphisms from* E *onto* F *will be denoted by* $Gl(\mathsf{E},\mathsf{F})$*. In case,* $\mathsf{E}=\mathsf{F}$ *the corresponding spaces will be denoted by* $\mathfrak{gl}(\mathsf{E})$ *and* $Gl(\mathsf{E})$*.*

$Gl(\mathsf{E})$ is a group under composition and is called the **general linear group**. In the following, the symbol$\widehat{\ }$is used to indicated that the factor is omitted.

Definition 1.7 *Let* V_i*,* $i=1,...,k$ *and* W *be normed spaces. A map* $\mu:\mathsf{V}_1\times\cdots\times\mathsf{V}_k\to\mathsf{W}$ *is called* ***multilinear*** *(k-multilinear) if for each i,* $1\leq i\leq k$ *and each fixed* $(w_1,...,\widehat{w_i},...,w_k)\in\mathsf{V}_1\times\cdots\times\widehat{\mathsf{V}_i}\times\cdots\times\mathsf{V}_k$ *we have that the map*

$$v\mapsto\mu(w_1,...,\underset{i-th\ slot}{v},...,w_k),$$

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

$$\|\mu(v_1,v_2,...,v_k)\|_{\mathsf{W}}\leq C\,\|v_1\|_{\mathsf{E}_1}\|v_2\|_{\mathsf{E}_2}\cdots\|v_k\|_{\mathsf{E}_k}$$

for all $(v_1,...,v_k)\in\mathsf{E}_1\times\cdots\times\mathsf{E}_k$*.*

Now $\mathsf{V}_1\times\cdots\times\mathsf{V}_k$ is a normed space in several equivalent ways just in the same way that we defined before for the case $k=2$. The topology is the product topology.

Proposition 1.8 *A multilinear map* $\mu:\mathsf{V}_1\times\cdots\times\mathsf{V}_k\to\mathsf{W}$ *is bounded if and only if it is continuous.*

Proof. ($\Leftarrow$) We shall simplify by letting $k=2$. Let (a_1,a_2) and (v_1,v_2) be elements of $\mathsf{E}_1\times\mathsf{E}_2$ and write

$$\begin{aligned}&\mu(v_1,v_2)-\mu(a_1,a_2)\\&=\mu(v_1-a_1,v_2)+\mu(a_1,v_2-a_2).\end{aligned}$$

We then have

$$\begin{aligned}&\|\mu(v_1,v_2)-\mu(a_1,a_2)\|\\&\leq C\,\|v_1-a_1\|\,\|v_2\|+C\,\|a_1\|\,\|v_2-a_2\|\end{aligned}$$

and so if $\|(v_1,v_2)-(a_1,a_2)\|\to0$ then $\|v_i-a_i\|\to0$ and we see that

$$\|\mu(v_1,v_2)-\mu(a_1,a_2)\|\to0.$$

(Recall that $\|(v_1,v_2)\|:=\max\{\|v_1\|,\|v_2\|\}$).

($\Rightarrow$) Start out by assuming that μ is continuous at $(0,0)$. Then for $r>0$ sufficiently small, $(v_1,v_2)\in B((0,0),r)$ implies that $\|\mu(v_1,v_2)\|\leq1$ so if for $i=1,2$ we let

$$z_i:=\frac{rv_i}{\|v_1\|_i+\epsilon}\quad\text{for some }\epsilon>0$$

then $(z_1, z_2) \in B((0,0), r)$ and $\|\mu(z_1, z_2)\| \leq 1$. The case $(v_1, v_2) = (0,0)$ is trivial so assume $(v_1, v_2) \neq (0,0)$. Then we have

$$\begin{aligned} \|\mu(z_1, z_2)\| &= \left\| \mu(\frac{rv_1}{\|v_1\| + \epsilon}, \frac{rv_2}{\|v_2\| + \epsilon}) \right\| \\ &= \frac{r^2}{(\|v_1\| + \epsilon)(\|v_2\| + \epsilon)} \|\mu(v_1, v_2)\| \leq 1 \end{aligned}$$

and so $\|\mu(v_1, v_2)\| \leq r^{-2}(\|v_1\| + \epsilon)(\|v_2\| + \epsilon)$. Now let $\epsilon \to 0$ to get the result. ■

Notation 1.9 *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* $L(\mathsf{E}, ..., \mathsf{E}; \mathsf{W})$.

Notation 1.10 *For linear maps* $T : \mathsf{V} \to \mathsf{W}$ *we sometimes write* $T \cdot v$ *instead of* $T(v)$ *depending on the notational needs of the moment. In fact, a particularly useful notational device is the following: Suppose for some set* X, *we have map* $A : X \to L(\mathsf{V}, \mathsf{W})$. *Then* $A(x)(\ v)$ *makes sense but we may find ourselves in a situation where* $A|_x\ v$ *is even more clear. This latter notation suggests a family of linear maps* $\{\ A|_x\}$ *parameterized by* $x \in X$.

Definition 1.11 *A multilinear map* $\mu : \mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ *is called* ***symmetric*** *if for any* $v_1, v_2, ..., v_k \in \mathsf{V}$ *we have that*

$$\mu(v_{\sigma(1)}, v_{\sigma(2)}, ..., v_{\sigma(k)}) = \mu(v_1, v_2, ..., v_k)$$

for all permutations σ *on the letters* $\{1, 2, ..., k\}$. *Similarly,* μ *is called* ***skew-symmetric*** *or* ***alternating if***

$$\mu(v_{\sigma(1)}, v_{\sigma(2)}, ..., v_{\sigma(k)}) = \operatorname{sgn}(\sigma)\mu(v_1, v_2, ..., v_k)$$

for all permutations σ. *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 if W is complete, that is, if W is a Banach space then the space $L(\mathsf{V}, \mathsf{W})$ is a Banach space in its own right with norm given by

$$\|A\| = \sup_{v \in \mathsf{V}, v \neq 0} \frac{\|A(v)\|_{\mathsf{W}}}{\|v\|_{\mathsf{V}}} = \sup\{\|A(v)\|_{\mathsf{W}} : \|v\|_{\mathsf{V}} = 1\}.$$

Similarly, the spaces $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ are also Banach spaces normed by

$$\|\mu\| := \sup\{\|\mu(v_1, v_2, ..., v_k)\|_{\mathsf{W}} : \|v_i\|_{\mathsf{E}_i} = 1 \text{ for } i = 1, ..., k\}$$

There is a natural linear bijection $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}, \mathsf{W})$ given by $T \mapsto \iota T$ where

$$(\iota T)(v_1)(v_2) = T(v_1, v_2)$$

and we identify the two spaces and write T instead of $\iota\ T$. We also have $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^3(\mathsf{V}; \mathsf{W})$ 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. It is also not hard to show that the isomorphism above is continuous and norm preserving, that is, ι is an isometric isomorphism.

We now come the central definition of differential calculus.

Definition 1.12 *A map $f : \mathsf{V} \supset U \to \mathsf{W}$ between normed spaces and defined on an open set $U \subset \mathsf{V}$ is said to be **differentiable at** $p \in U$ if and only if there is a bounded linear map $A_p \in L(\mathsf{V}, \mathsf{W})$ such that*

$$\lim_{\|h\| \to 0} \frac{\|f(p+h) - f(p) - A_p \cdot h\|}{\|h\|} = 0$$

Proposition 1.13 *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 $p + h \in U$ we have

$$\begin{aligned} A_p \cdot h - B_p \cdot h = &-(f(p+h) - f(p) - A_p \cdot h) \\ &+ (f(p+h) - f(p) - B_p \cdot h). \end{aligned}$$

Taking norms, dividing by $\|h\|$ and taking the limit as $\|h\| \to 0$ we get

$$\|A_p h - B_p h\| / \|h\| \to 0$$

Now let $h \neq 0$ be arbitrary and choose $\epsilon > 0$ small enough that $p + \epsilon h \in U$. Then we have

$$\|A_p(\epsilon h) - B_p(\epsilon h)\| / \|\epsilon h\| \to 0.$$

But, by linearity $\|A_p(\epsilon h) - B_p(\epsilon h)\| / \|\epsilon h\| = \|A_p h - B_p h\| / \|h\|$ which doesn't even depend on ϵ so in fact $\|A_p h - B_p h\| = 0$. ■

If a function f is differentiable at p, then the linear map A_p which exists by definition and is unique by the above result, will be denoted by $Df(p)$. The linear map $Df(p)$ is called the **derivative** of f at p. We will also use the notation $Df|_p$ or sometimes $f'(p)$. We often write $Df|_p \cdot h$ instead of $Df(p)(h)$.

It is not hard to show that the derivative of a constant map is constant and the derivative of a (bounded) linear map is the very same linear map.

If we are interested in differentiating "in one direction" then we may use the natural notion of directional derivative. A map $f : \mathsf{V} \supset U \to \mathsf{W}$ has a directional derivative $D_h f$ at p in the direction h if the following limit exists:

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

In other words, $D_h f(p) = \frac{d}{dt}\big|_{t=0} f(p+th)$. But a function may have a directional derivative in every direction (at some fixed p), that is, for every $h \in \mathsf{V}$ and yet still not be differentiable at p in the sense of definition 1.12.

Notation 1.14 *The directional derivative is written as $(D_h f)(p)$ and, in case f is actually differentiable at p, this is equal to $Df|_p h = Df(p) \cdot h$ (the proof is easy). Note carefully that $D_x f$ should not be confused with $Df|_x$.*

Let us now restrict our attention to complete normed spaces. From now on $\mathsf{V}, \mathsf{W}, \mathsf{E}$ etc. will refer to Banach spaces. If it happens that a map $f : U \subset \mathsf{V} \to \mathsf{W}$ 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 $p \mapsto Df(p)$. This map is called the derivative of f. If this map itself is differentiable at some $p \in \mathsf{V}$ then its derivative at p is denoted $DDf(p) = D^2 f(p)$ or $D^2 f\big|_p$ and is an element of $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}; \mathsf{W})$ which is called the second derivative at p. If in turn $D^2 f\big|_p$ exist for all p throughout U then we have a map $D^2 f : U \to L^2(\mathsf{V}; \mathsf{W})$ called the second derivative. Similarly, we may inductively define $D^k f\big|_p \in L^k(\mathsf{V}; \mathsf{W})$ and $D^k f : U \to L^k(\mathsf{V}; \mathsf{W})$ whenever f is nice enough that the process can be iterated appropriately.

Definition 1.15 *We say that a map $f : U \subset \mathsf{V} \to \mathsf{W}$ is C^r-differentiable on U if $D^r f|_p \in L^r(\mathsf{V}, \mathsf{W})$ exists for all $p \in U$ and if $D^r f$ 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).*

To complete the notation we let C^0 indicate mere continuity. The reader should not find it hard to see that a bounded multilinear map is C^∞.

Definition 1.16 *A bijection f between open sets $U_\alpha \subset \mathsf{V}$ and $U_\beta \subset \mathsf{W}$ is called a C^r-**diffeomorphism** if and only if 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.*

Definition 1.17 *Let U be open in V. A map $f : U \to \mathsf{W}$ is called a* ***local*** *C^r* ***diffeomorphism*** *if and only if for every $p \in U$ there is an open set $U_p \subset U$ with $p \in U_p$ such that $f|_{U_p} : U_p \to f(U_p)$ is a C^r-diffeomorphism.*

We will sometimes think of the derivative of a curve[2] $c : I \subset \mathbb{R} \to \mathsf{E}$ at $t_0 \in I$, as a velocity vector and so we are identifying $Dc|_{t_0} \in L(\mathbb{R}, \mathsf{E})$ with $Dc|_{t_0} \cdot 1 \in \mathsf{E}$. Here the number 1 is playing the role of the unit vector in $\mathbb{R}$. Especially in this context we write the velocity vector using the notation $\dot{c}(t_0)$.

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

$$\int_{[a,b]} f := \sum_{i=0}^{k-1} f(t_i) \Delta t_i$$

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

where $\Delta t_i := t_{i+1} - t_i$. 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. The norm on $\mathcal{B}(a,b,\mathsf{V})$ is given by $\|f\| = \sup_{a\leq t\leq b}\|f(t)\|$. 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 f we have

$$\left|\int_{[a,b]} f\right| \leq (b-a)\,\|f\|_{\infty}.$$

The elements of $\bar{\mathcal{S}}(a,b,\mathsf{V})$ are referred to as **regulated** maps. In the limit, this bound persists and so is valid for all $f \in \bar{\mathcal{S}}(a,b,\mathsf{V})$. 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 1.18 *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*

$$\int_{[a,b]} \ell \circ f = \ell\left(\int_{[a,b]} f\right)$$

Proof. This is obvious for step functions. The general result follows by taking a limit for a sequence of step functions converging to f in $\bar{\mathcal{S}}(a,b,\mathsf{V})$. ■

The following is a version of the **mean value theorem**:

Theorem 1.19 *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 $c(t) = (1-t)x + ty$ then

$$f(y) - f(x) = \int_0^1 Df(c(t))dt \cdot (y-x).$$

Notice that in the previous theorem we have $\int_0^1 Df(c(t))dt \in L(\mathsf{V},\mathsf{W})$.

A subset U of a Banach space (or any vector space) is said to be convex if it has the property that whenever x and y are contained in U then so are all points of the line segment $l_{xy} := \{(1-t)x + ty : 0 \leq t \leq 1\}$.

Corollary 1.20 *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 $x, y \in U$ we have*

$$\|f(y) - f(x)\| \leq C_{x,y}\,\|y - x\|$$

where $C_{x,y}$ is the supremum over all values taken by f on point of the line segment l_{xy} (see above).

Let $f : U \subset \mathsf{E} \to \mathsf{F}$ be a map and suppose that we have a splitting $\mathsf{E} = \mathsf{E}_1 \times \mathsf{E}_2$. Let (x, y) denote a generic element of $\mathsf{E}_1 \times \mathsf{E}_2$. Now for every $(a, b) \in U \subset \mathsf{E}_1 \times \mathsf{E}_2$ the partial maps $f_{a,} : y \mapsto f(a, y)$ and $f_{,b} : x \mapsto f(x, b)$ are defined in some neighborhood of b (resp. a). Notice the logical placement of commas in this notation. We define the partial derivatives, when they exist, by $D_2 f(a, b) := Df_{a,}(b)$ and $D_1 f(a, b) := Df_{,b}(a)$. These are, of course, linear maps.

$$D_1 f(a, b) : \mathsf{E}_1 \to \mathsf{F}$$
$$D_2 f(a, b) : \mathsf{E}_2 \to \mathsf{F}$$

Remark 1.21 *It is useful to notice that if we consider that maps $\iota_{a,} : x \mapsto (a, x)$ and $\iota_{,b} : x \mapsto (x, a)$ then $D_2 f(a, b) = D(f \circ \iota_{a,})(b)$ and $D_1 f(a, b) = D(f \circ \iota_{,b})(a)$.*

The partial derivative can exist even in cases where f might not be differentiable in the sense we have defined. This is a slight generalization of the point made earlier: f might be differentiable only in certain directions without being fully differentiable in the sense of 1.12. On the other hand, we have

Proposition 1.22 *If f has continuous partial derivatives $D_i f(x, y) : \mathsf{E}_i \to \mathsf{F}$ near $(x, y) \in \mathsf{E}_1 \times \mathsf{E}_2$ then $Df(x, y)$ exists and is continuous. In this case, we have for $\mathrm{v} = (\mathrm{v}_1, \mathrm{v}_2)$,*

$$\begin{aligned} &Df(x, y) \cdot (\mathrm{v}_1, \mathrm{v}_2) \\ &= D_1 f(x, y) \cdot \mathrm{v}_1 + D_2 f(x, y) \cdot \mathrm{v}_2. \end{aligned}$$

Clearly we can consider maps on several factors $f :: \mathsf{E}_1 \times \mathsf{E}_2 \cdots \times \mathsf{E}_n \to \mathsf{F}$ and then we can define partial derivatives $D_i f : \mathsf{E}_i \to \mathsf{F}$ for $i = 1, ..., n$ in the obvious way. Notice that the meaning of $D_i f$ depends on how we factor the domain. For example, we have both $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ and also $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a map. Then we for $a \in U$ we define

$$\begin{aligned} (\partial_i f)(a) &:= (D_i f)(a) \cdot \mathsf{e} \\ &= \lim_{h \to 0} \left[\frac{f(a^1, ..., a^2 + h, ..., a^n) - f(a^1, ..., a^n)}{h} \right] \end{aligned}$$

where e is the standard basis vector in $\mathbb{R}$. The function $\partial_i f$ is defined where the above limit exists. If we have named the standard coordinates on $\mathbb{R}^n$ say as $(x^1, ..., x^n)$ then it is common to write $\partial_i f$ as

$$\frac{\partial f}{\partial x^i}$$

Note that in this setting, the linear map $(D_i f)(a)$ is often identified with the number $\partial_i f(a)$.

Now let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be a map that is differentiable at $a = (a^1, ..., a^n) \in \mathbb{R}^n$. The map f is given by m functions $f^i : U \to \mathbb{R}$, $1 \leq i \leq m$ such that

$f(u) = (f^1(u), ..., f^n(u))$. The above proposition have an obvious generalization to the case where we decompose the Banach space into more than two factors as in $\mathbb{R}^m = \mathbb{R} \times \cdots \times \mathbb{R}$ and we find that if all partials $\frac{\partial f^i}{\partial x^j}$ are continuous in U then f is C^1.

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:

$$J_a(f) := \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(a) & \frac{\partial f^1}{\partial x^2}(a) & \cdots & \frac{\partial f^1}{\partial x^n}(a) \\ \frac{\partial f^2}{\partial x^1}(a) & & & \frac{\partial f^2}{\partial x^n}(a) \\ \vdots & & \ddots & \\ \frac{\partial f^m}{\partial x^1}(a) & & & \frac{\partial f^m}{\partial x^n}(a) \end{pmatrix}.$$

The rank of this matrix is called the rank of f at a. If $n = m$ 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.

1.1.1 Chain Rule, Product rule and Taylor's Theorem

Theorem 1.23 (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 $D(g \circ f) = Dg(f(p)) \circ Dg(p)$. In other words, if $v \in \mathsf{E}_1$ then

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

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 $O_1(v)$, $O_2(v)$ etc. to mean functions such that $O_i(v) \to 0$ as $\|v\| \to 0$. Let $y = f(p)$. Since f is differentiable at p we have

$$f(p+h) = y + Df|_p \cdot h + \|h\| O_1(h) := y + \Delta y$$

and since g is differentiable at y we have $g(y+\Delta y) = Dg|_y \cdot (\Delta y) + \|\Delta y\| O_2(\Delta y)$. Now $\Delta y \to 0$ as $h \to 0$ and in turn $O_2(\Delta y) \to 0$ hence

$$\begin{aligned} g \circ f(p+h) &= g(y + \Delta y) \\ &= Dg|_y \cdot (\Delta y) + \|\Delta y\| O_2(\Delta y) \\ &= Dg|_y \cdot (Df|_p \cdot h + \|h\| O_1(h)) + \|h\| O_3(h) \\ &= Dg|_y \cdot Df|_p \cdot h + \|h\| Dg|_y \cdot O_1(h) + \|h\| O_3(h) \\ &= Dg|_y \cdot Df|_p \cdot h + \|h\| O_4(h) \end{aligned}$$

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 on U_1 by

$$m_{f,g} : p \mapsto (Dg(f(p), Df(p)).$$

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:

compositions of C^r 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. ■

The following lemma is useful for calculations and may be used without explicit mention:

Lemma 1.24 *Let $f : U \subset \mathsf{V} \to \mathsf{W}$ be twice differentiable at $x_0 \in U \subset \mathsf{V}$; then the map $D_v f : x \mapsto Df(x) \cdot v$ is differentiable at x_0 and its derivative at x_0 is given by*

$$D(D_v f)|_{x_0} \cdot h = D^2 f(x_0)(h, v).$$

Proof. The map $D_v f : x \mapsto Df(x) \cdot v$ is decomposed as the composition

$$x \overset{Df}{\mapsto} Df|_x \overset{R^v}{\mapsto} Df|_x \cdot v$$

where $R^v : L(\mathsf{V}, \mathsf{W}) \mapsto \mathsf{W}$ is the map $(A, b) \mapsto A \cdot b$. The chain rule gives

$$\begin{aligned} D(D_v f)(x_0) \cdot h &= DR^v(Df|_{x_0}) \cdot D(Df)|_{x_0} \cdot h) \\ &= DR^v(Df(x_0)) \cdot (D^2 f(x_0) \cdot h). \end{aligned}$$

But R^v is linear and so $DR^v(y) = R^v$ for all y. Thus

$$\begin{aligned} D(D_v f)|_{x_0} \cdot h &= R^v(D^2 f(x_0) \cdot h) \\ &= (D^2 f(x_0) \cdot h) \cdot v = D^2 f(x_0)(h, v). \end{aligned}$$

$$D(D_v f)|_{x_0} \cdot h = D^2 f(x_0)(h, v).$$

■

Theorem 1.25 *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:*

$$D^2 f(p)(w, v) = D^2 f(p)(v, w).$$

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

Proof. Let $p \in U$ and define an affine map $A : \mathbb{R}^2 \to \mathsf{V}$ by $A(s,t) := p + sv + tw$. By the chain rule we have

$$\frac{\partial^2 (f \circ A)}{\partial s \partial t}(0) = D^2(f \circ A)(0) \cdot (\mathbf{e}_1, \mathbf{e}_2) = D^2 f(p) \cdot (v, w)$$

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 1.26 *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 $\varrho(f_1, f_2) : x \mapsto \varrho(f_1(x), f_2(x))$. Furthermore,*

$$D\varrho(f_1, f_2)|_x \cdot v = \varrho(Df_1|_x \cdot v\ ,\ f_2(x)) + \varrho(f_1(x)\ ,\ Df_2|_x \cdot v).$$

In particular, if F is a Banach 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

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

Recall that for a fixed x, higher derivatives $D^p f|_x$ are symmetric multilinear maps. For the following let $(y)^k$ denote $(y, y, ..., y)$ where the y is repeated k times. With this notation we have the following version of Taylor's theorem.

Theorem 1.27 (Taylor's theorem) *Given Banach spaces V and W, a C^r function $f : U \to \mathsf{W}$ and a line segment $t \mapsto (1-t)x + ty$ 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*

$$f(x+y) = f(x) + \frac{1}{1!}\, Df|_x \cdot y + \frac{1}{2!}\, D^2 f\big|_x \cdot (y)^2 + \cdots + \frac{1}{(k-1)!}\, D^{k-1} f\big|_x \cdot (y)^{(k-1)}$$
$$+ \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} D^k f(x + ty) \cdot (y)^k dt$$

The proof is by induction and follows the usual proof closely. See [**?**]. 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.

1.1.2 Local theory of differentiable 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. The so called hard inverse mapping theorems such as that of Nash and Moser require special estimates and are constructed to apply only in a very controlled situation.

Definition 1.28 *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* $f(U)$*. If* f *is a* C^r *diffeomorphism near* p *for all* $p \in \operatorname{dom}(f)$ *then we say that* f *is a* ***local*** C^r ***diffeomorphism****.*

Definition 1.29 *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* $d(f(x_1), f(x_2)) \leq kd(x_1, x_2)$ *for all* $x_1, x_2 \in X$*. If* $0 < k < 1$ *the map is called a* ***contraction mapping*** *(with constant* k*) and is said to be* k***-contractive****.*

The following technical result has numerous applications and uses the idea of iterating a map. **Warning**: For this next theorem f^n will denote the $n-$fold composition $f \circ f \circ \cdots \circ f$ rather than an $n-$fold product.

Proposition 1.30 (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*

$$d(f(x), f(y)) \leq kd(x, y)$$

for some fixed $0 \leq k < 1$*. 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*.*

2) for any $y \in F$ *the sequence* $y_n := f^n(y)$ *converges to the fixed point* x_0 *with the error estimate* $d(y_n, x_0) \leq \frac{k^n}{1-k} d(y_1, x_0)$*.*

Proof. Let $y \in F$. By iteration

$$d(f^n(y), f^{n-1}(y)) \leq kd(f^{n-1}(y), f^{n-2}(y)) \leq \cdots \leq k^{n-1}d(f(y), y)$$

as follows:

$$\begin{aligned} d(f^{n+j+1}(y), f^n(y)) &\leq d(f^{n+j+1}(y), f^{n+j}(y)) + \cdots + d(f^{n+1}(y), f^n(y)) \\ &\leq (k^{j+1} + \cdots + k)d(f^n(y), f^{n-1}(y)) \\ &\leq \frac{k}{1-k} d(f^n(y), f^{n-1}(y)) \\ &\leq \frac{k^n}{1-k} d(f^1(y), y)) \end{aligned}$$

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 were 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 1.31 *Note that a Lipschitz map f may not satisfy the hypotheses of the last theorem even if $k < 1$ since an open U is not a complete metric space unless $U = \mathsf{E}$.*

Definition 1.32 *A continuous map $f : U \to \mathsf{E}$ such that $L_f := \mathrm{id}_U - f$ is injective has a inverse G_f (not necessarily continuous) 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 1.33 *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

$$\begin{aligned} \|R_f(y_1) - R_f(y_2)\| &= \|f(x_1) - f(x_2)\| \leq k \|x_1 - x_2\| \\ &\leq k \|y_1 - R_f(y_1) - (y_2 - R_f(y_2))\| \\ &\leq k \|y_1 - y_2\| + k \|R_f(y_1) - R_f(y_2)\| . \end{aligned}$$

Solving this inequality we get

$$\|R_f(y_1) - R_f(y_2)\| \leq \frac{k}{1-k} \|y_1 - y_2\| .$$

■

Lemma 1.34 *The space $Gl(\mathsf{E},\mathsf{F})$ 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 $A\in Gl(\mathsf{E})$ then $A^{-1}=\lim_{N\to\infty}\sum_{n=0}^{N}(\mathrm{id}-A)^n$.*

Proof. Let $A_0\in GL(\mathsf{E},\mathsf{F})$. The map $A\mapsto A_0^{-1}\circ A$ is continuous and maps $GL(\mathsf{E},\mathsf{F})$ onto $GL(\mathsf{E},\mathsf{F})$. It follows that we may assume that $\mathsf{E}=\mathsf{F}$ and $A_0=\mathrm{id}_{\mathsf{E}}$. Our task is to show that elements of $\mathrm{L}(\mathsf{E},\mathsf{E})$ that are close enough to id_{E} are in fact elements of $GL(\mathsf{E})$. For this we show that

$$\|\mathrm{id}-A\|<1$$

implies that $A\in GL(\mathsf{E})$. 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 1.35 *The map $\mathcal{I}:Gl(\mathsf{E},\mathsf{F})\to Gl(\mathsf{E},\mathsf{F})$ given by taking inverses is a C^∞ map and the derivative of $\mathcal{I}:g\mapsto g^{-1}$ at some $g_0\in Gl(\mathsf{E},\mathsf{F})$ is the linear map given by the formula: $D\mathcal{I}|_{g_0}:\mathrm{A}\mapsto -g_0^{-1}\mathrm{A}g_0^{-1}$.*

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 $D\mathrm{inv}_{\mathsf{E}}|_{h_0}:\mathrm{A}\mapsto -A$. Now by the chain rule we have

$$\begin{aligned}
(D\,\mathrm{inv}|_{h_0})\cdot\mathrm{A}&=D(R_{h_0^{-1}}\circ\mathrm{inv}_{\mathsf{E}}\circ L_{h_0^{-1}})\cdot\mathrm{A}\\
&=\left(R_{h_0^{-1}}\circ D\,\mathrm{inv}_{\mathsf{E}}|_{\mathrm{id}}\circ L_{h_0^{-1}}\right)\cdot\mathrm{A}\\
&=R_{h_0^{-1}}\circ(-\mathrm{A})\circ L_{h_0^{-1}}=-h_0^{-1}\mathrm{A}h_0^{-1}
\end{aligned}$$

so the result is true for an arbitrary $h_0 \in \mathrm{GL}(\mathsf{E},\mathsf{F})$. Thus we are reduced to showing that $D\mathrm{inv}_{\mathsf{E}}|_{\mathrm{id}} : A \mapsto -A$. 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 (**??**) the above difference quotient becomes

$$\begin{aligned}
&\lim_{\|\mathrm{A}\|\to 0} \frac{\left\|(\mathrm{id}+\mathrm{A})^{-1} - \mathrm{id}+\mathrm{A}\right\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \frac{\|\sum_{n=0}^{\infty}(\mathrm{id}-(\mathrm{id}+\mathrm{A}))^n - \mathrm{id}+\mathrm{A}\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \frac{\|\sum_{n=0}^{\infty}(-\mathrm{A})^n - \mathrm{id}+\mathrm{A}\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \frac{\|\sum_{n=2}^{\infty}(-\mathrm{A})^n\|}{\|\mathrm{A}\|} \leq \lim_{\|\mathrm{A}\|\to 0} \frac{\sum_{n=2}^{\infty}\|\mathrm{A}\|^n}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\|\to 0} \sum_{n=1}^{\infty}\|\mathrm{A}\|^n = \lim_{\|\mathrm{A}\|\to 0} \frac{\|\mathrm{A}\|}{1-\|\mathrm{A}\|} = 0.
\end{aligned}$$

■

Theorem 1.36 (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 $x_0 \in U$ and that $f'(x_0) = Df|_x : \mathsf{E} \to \mathsf{F}$ is a continuous linear isomorphism. Then there exists an open set $V \subset U$ with $x_0 \in V$ 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 $Df^{-1}|_y = (Df|_{f^{-1}(y)})^{-1}$.*

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 $x_0 = 0$, $f(0) = 0$ 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

$$\|Dg|_x\| < \frac{1}{2}$$

for $x \in B(0,2r)$. The mean value theorem now tells us that $\|g(x_2) - g(x_1)\| \leq \frac{1}{2}\|x_2 - x_1\|$ 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 1.30 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 $f:\overline{B}(0,r/2)\rightarrow f(\overline{B}(0,r/2))$ is invertible. But f^{-1} is also continuous since

$$\begin{aligned}\left\|f^{-1}(y_2)-f^{-1}(y_1)\right\| &= \|x_2-x_1\| \\ &\leq \|f(x_2)-f(x_1)\|+\|g(x_2)-g(x_1)\| \\ &\leq \|f(x_2)-f(x_1)\|+\frac{1}{2}\|x_2-x_1\| \\ &= \|y_2-y_1\|+\frac{1}{2}\left\|f^{-1}(y_2)-f^{-1}(y_1)\right\|\end{aligned}$$

Thus $\left\|f^{-1}(y_2)-f^{-1}(y_1)\right\|\leq 2\|y_2-y_1\|$ 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

$$\begin{aligned}&\left\|f^{-1}(y_2)-f^{-1}(y_1)-(Df(x_1))^{-1}\cdot\Delta y\right\| \\ &=\left\|\Delta x-(Df(x_1))^{-1}(f(x_2)-f(x_1))\right\| \\ &=\left\|\{(Df(x_1))^{-1}Df(x_1)\}\Delta x-\{(Df(x_1))^{-1}Df(x_1)\}(Df(x_1))^{-1}(f(x_2)-f(x_1))\right\| \\ &\leq B\left\|Df(x_1)\Delta x-(f(x_2)-f(x_1))\right\|\leq o(\Delta x)=o(\Delta y)\ \text{(by continuity)}.\end{aligned}$$

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 1.35. This last step is left to the reader. ■

Corollary 1.37 *Let $U\subset\mathsf{E}$ be an open set. Suppose that $f:U\rightarrow\mathsf{F}$ is differentiable with $Df(p):\mathsf{E}\rightarrow\mathsf{F}$ a (bounded) linear isomorphism for each $p\in U$. Then f is a local diffeomorphism.*

Example 1.38 *Consider the map $\phi:\mathbb{R}^2\rightarrow\mathbb{R}^2$ given by*

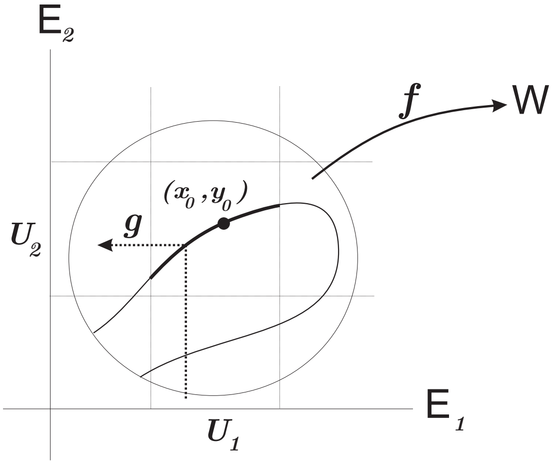
$$\phi(x,y):=(x^2-y^2,2xy)$$

The derivative is given by the matrix

$$\begin{bmatrix}2x & -2y\\ 2y & 2x\end{bmatrix}$$

which is invertible for all $(x,y)\neq(0,0)$. Thus, by the inverse mapping theorem, the restriction of ϕ to a sufficiently small open disk centered at any point but the origin will be a diffeomorphism. We may say that the restriction $\phi|_{\mathbb{R}^2\backslash\{(0,0)\}}$ is a local diffeomorphism. However, notice that $\phi(x,y)=\phi(-x,-y)$ so generically ϕ is a 2-1 map and in particular is not a (global) diffeomorphism.

The next theorem is basic for differentiable manifold theory.



Theorem 1.39 (Implicit Mapping Theorem) *Let* $\mathsf{E}_1, \mathsf{E}_2$ *and* W *be Banach spaces and* $O \subset \mathsf{E}_1 \times \mathsf{E}_2$ *open. Let* $f : O \to \mathsf{W}$ *be a* C^r*mapping such that* $f(x_0, y_0) = 0$*. If* $D_2 f(x_0, y_0) : \mathsf{E}_2 \to \mathsf{W}$ *is a continuous linear isomorphism then there exists open sets* $U_1 \subset \mathsf{E}_1$ *and* $U_2 \subset \mathsf{E}_2$ *such that* $U_1 \times U_2 \subset O$ *with* $x_0 \in U_1$ *and* C^r *mapping* $g : U_1 \to U_2$ *with* $g(x_0) = y_0$ *such that for all* $(x, y) \in U_1 \times U_2$*. We can take* U_1 *to be connected*

$$f(x, y) = 0 \text{ if and only if } y = g(x).$$

The function g in the theorem satisfies $f(x, g(x)) = 0$ which says that graph of g is contained in $(U_1 \times U_2) \cap f^{-1}(0)$ but the conclusion of the theorem is stronger since it says that in fact the graph of g is exactly equal to $(U_1 \times U_2) \cap f^{-1}(0)$.

Proof of the implicit mapping theorem. Let $F : O \to \mathsf{E}_1 \times \mathsf{W}$ be defined by

$$F(x, y) = (x, f(x, y)).$$

Notice that $DF|_{(x_0,y_0)}$ has the form

$$\begin{bmatrix} \text{id} & 0 \\ D_1 f(x_0, y_0) & D_2 f(x_0, y_0) \end{bmatrix}$$

and it is easily seen that this is a toplinear isomorphism from $\mathsf{E}_1 \times \mathsf{E}_2$ to $\mathsf{E}_1 \times \mathsf{W}$. Thus by the inverse mapping theorem there is an open set $O' \subset O$ containing (x_0, y_0) such that F restricted to O' is a diffeomorphism. Now take open sets U_1 and U_2 so that $(x_0, y_0) \in U_1 \times U_2 \subset O'$ and let $\psi := F|_{U_1 \times U_2}$. Then ψ is a diffeomorphism and, being a restriction of F, we have $\psi(x, y) = (x, f(x, y))$ for all $(x, y) \in U_1 \times U_2$. Now ψ^{-1} must have the form $\psi^{-1}(x, w) = (x, h(x, w))$ where $h : \psi(U_1 \times U_2) \to U_2$. Note that $\psi(U_1 \times U_2) = U_1 \times h(U_1 \times U_2)$.

Let $g(x) := h(x, 0)$. Then $(x, 0) = \psi \circ \psi^{-1}(x, 0) = \psi \circ (x, h(x, 0)) = (x, f(x, h(x, 0)))$ so that in particular $0 = f(x, h(x, 0)) = f(x, g(x))$ from which we now see that $\text{graph}(g) \subset (U_1 \times U_2) \cap f^{-1}(0)$.

We now show that $(U_1 \times U_2) \cap f^{-1}(0) \subset \mathrm{graph}(g)$. Suppose that for some $(x, y) \in U_1 \times U_2$ we have $f(x, y) = 0$. Then $\psi(x, y) = (x, 0)$ and so

$$\begin{aligned}(x, y) &= \psi^{-1} \circ \psi(x, y) \\ &= \psi^{-1}(x, 0) = (x, h(x, 0)) \\ &= (x, g(x))\end{aligned}$$

from which we see that $y = g(x)$ and thus $(x, y) \in \mathrm{graph}(g)$. ■

The simplest situation is that of a function $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(a, b) = 0$ and $D_2f(a, b) \neq 0$. Then the implicit mapping theorem gives a function g so that $f(x, g(x)) = 0$ for all x sufficiently near a. Note, however, the following exercise:

Exercise 1.40 *Find a function $f : \mathbb{R}^2 \to \mathbb{R}$ with $D_2f(0, 0) = 0$ and a continuous function g with $f(x, g(x)) = 0$ for all x sufficiently near a. Thus we see that the implicit mapping theorem gives sufficient but not necessary conditions for the existence of a function g with the property $f(x, g(x)) = 0$.*

1.1.3 Immersion

Theorem 1.41 *Let E, and F be Banach spaces. Let U be and open subset of E with $0 \in U$, and let $f : U \to \mathsf{E} \times \mathsf{F}$ be a smooth map with $f(0) = (0, 0)$. If $Df(0) : \mathsf{E} \to \mathsf{E} \times \mathsf{F}$ is of the form $x \mapsto (\alpha(x), 0)$ for a continuous linear isomorphism $\alpha : \mathsf{E} \to \mathsf{E}$ then there exists a diffeomorphism g from an open neighborhood V of $(0, 0) \in \mathsf{E}\times\mathsf{F}$ onto an an open neighborhood W of $(0, 0) \in \mathsf{E}\times\mathsf{F}$ such that $g \circ f : f^{-1}(V) \to W$ is of the form $a \mapsto (a, 0)$.*

Proof. Define $\phi : U \times \mathsf{F} \to \mathsf{E} \times \mathsf{F}$ by $\phi(x, y) := f(x) + (0, y)$. Note that $\phi(x, 0) = f(x)$ and $D\phi(0, 0) = (\alpha,\mathrm{id}_{\mathsf{F}}) : x \mapsto (\alpha(x), x)$ and this is clearly a continuous linear isomorphism. The inverse mapping theorem there is a local inverse for ϕ say $g : V \to W$. Thus if a is in U' we have

$$g \circ f(a) = g(\phi(a, 0)) = (a, 0)$$

■

Corollary 1.42 *If U is an open neighborhood of $0 \in \mathbb{R}^k$ and $f : U \subset \mathbb{R}^k \to \mathbb{R}^n$ is a smooth map with $f(0) = 0$ such that $Df(0)$ has rank k then there is an open neighborhood V of $0 \in \mathbb{R}^n$, an open neighborhood of W of $0 \in \mathbb{R}^n$ and a diffeomorphism $g : V \to W$ such that that $g \circ f : f^{-1}(V) \to W$ is of the form $(a^1, ..., a^k) \mapsto (a^1, ..., a^k, 0, ..., 0)$.*

Proof. Since $Df(0)$ has rank k there is a linear map $A : \mathbb{R}^n \to \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$ such that $A \circ Df(0)$ is of the form $x \mapsto (\alpha(x), 0)$ for a linear isomorphism $\alpha : \mathbb{R}^k \to \mathbb{R}^k$. But $A \circ Df(0) = D(A \circ f)(0)$ so apply the previous theorem to $A \circ f$. ■

1.1.4 Submersion

Theorem 1.43 *Let E_1, E_2 and F be Banach space and let U be an open subset of a point $(a_1, a_2) \in \mathsf{E}_1 \times \mathsf{E}_2$. If $f : U \to \mathsf{F}$ be a smooth map with $f(a_1, a_2) = 0$. If the partial derivative $D_1 f(a_1, a_2) : \mathsf{E}_1 \to \mathsf{F}$ is an continuous linear isomorphism then there exist a diffeomorphism $h : V_1 \times V_2 \to U_1$ where $U_1 \subset U$ is an open neighborhood of (a_1, a_2) and V_1 and V_2 are open in neighborhoods of $0 \in \mathsf{E}_1$ and $0 \in \mathsf{E}_2$ respectively such that the composite map $f \circ h$ is of the form $(x, y) \mapsto x$.*

Proof. Clearly we make assume that $(a_1, a_2) = (0, 0)$. Let $\phi : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1 \times \mathsf{E}_2$ be defined by $\phi(x, y) := (f(x, y), y)$. In matrix format the derivative of ϕ at $(0, 0)$ is

$$\begin{pmatrix} D_1 f & D_2 f \\ 0 & \mathrm{id} \end{pmatrix}$$

and so is a continuous linear isomorphism. The inverse mapping theorem provides a local inverse h of ϕ. We may arrange that the domain of ϕ is of the form $V_1 \times V_2$ with image inside U. Now suppose that $\phi(b_1, b_2) = (x, y) \in V_1 \times V_2$. Then $(x, y) = (f(b_1, b_2), b_2)$ so $x = f(b_1, b_2)$ and $y = b_2$. Then since

$$f \circ h(x, y) = f(b_1, b_2) = x$$

we see that $f \circ h$ has the required form. ■

Corollary 1.44 *If U is an open neighborhood of $0 \in \mathbb{R}^n$ and $f : U \subset \mathbb{R}^n \to \mathbb{R}^k$ is a smooth map with $f(0) = 0$ and if the partial derivative $D_1 f(0, 0)$ is a linear isomorphism then there exist a diffeomorphism $h : V \subset \mathbb{R}^n \to U_1$ where V is an open neighborhood of $0 \in \mathbb{R}^n$ and U is an open is an open neighborhood of $0 \in \mathbb{R}^k$ respectively such that the composite map $f \circ h$ is of the form*

$$\left(a^1, ..., a^n\right) \mapsto \left(a^1, ..., a^k\right)$$

1.1.5 Constant Rank Theorem

If the reader thinks about what is meant by local immersion and local submersion they 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 will allow. Now rank is only 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. We will state the following theorem only for the finite dimensional case. However there is a way to formulate and prove a version for infinite dimensional Banach spaces that can be found in [**?**].

Theorem 1.45 (The Rank Theorem) *Let $f : (\mathbb{R}^n, p) \to (\mathbb{R}^m, q)$ be a local map such that Df has constant rank r in an open set containing p. Then there are local diffeomorphisms $g_1 : (\mathbb{R}^n, p) \to (\mathbb{R}^n, q)$ and $g_2 : (\mathbb{R}^m, q) \to (\mathbb{R}^m, 0)$ 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

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

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 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 rank$(f) =$ 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. ■

1.1.6 Existence and uniqueness for differential equations

Theorem 1.46 *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 $t \in (-a, a)$.*

Theorem 1.47 *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 $v \in \mathsf{E}$.

1.2 Naive Functional Calculus.

We have recalled the basic definitions of the directional derivative of a map such as $f : \mathbb{R}^n \to \mathbb{R}^m$. This is a good starting point for making the generalizations to come but let us think about a bit more about our "directions" h and "points" p. In both cases these refer to $n-$tuples in $\mathbb{R}^n$. The values taken by the function are also tuples ($m-$tuples in this instance). From one point of view a $n-$tuple is just a function whose domain is the finite set $\{1, 2, ..., n\}$. For instance, the n-tuple $h = (h^1, ..., h^n)$ is just the function $i \mapsto h^i$ which may as well have been written $i \mapsto h(i)$. This suggests that we generalize to functions whose domain is an infinite set. A sequence of real numbers is just such an example but so is any real (or complex) valued function. This brings us to the notion of a function space. An example of a function space is $C([0, 1])$, the space of continuous functions on the unit interval $[0, 1]$. So, whereas an element of $\mathbb{R}^3$, say $(1, \pi, 0)$ has 3 components or entries, an element of $C([0, 1])$, say $(t \mapsto \sin(2\pi t))$ has a continuum of "entries". For example, the $1/2$ entry of the latter element is $\sin(2\pi(1/2)) = 0$. So one approach to generalizing the usual setting of calculus might be to consider replacing the space of $n-$tuples $\mathbb{R}^n$ by a space of functions. Now we are interested in differentiating functions whose arguments are themselves functions. This type of function is sometimes called a functional. We shall sometimes follow the tradition of writing $F[f]$ instead of $F(f)$. Some books even write $F[f(x)]$. Notice that this is *not* a composition of functions. A simple example of a functional on $C([0, 1])$ is

$$F[f] = \int_0^1 f^2(x)dx.$$

We may then easily define a formal notion of directional derivative:

$$(D_h F)\,[f] = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\left(F[f + \epsilon h] - F[f]\right)$$

where h is some function which is the "direction vector". This also allows us to define the differential δF which is a linear map on the functions space given at f by $\left.\delta F\right|_f h = (D_h F)\,[f]$. We use a δ instead of a d to avoid confusion between dx^i and δx^i which comes about when x^i is simultaneously used to denote a number and also a function of , say, t.

It will become apparent that choosing the right function space for a particular problem is highly nontrivial and in each case the function space must be given an appropriate topology. In the following few paragraphs our discussion will be informal and we shall be rather cavalier with regard to the issues just mentioned. After this informal presentation we will develop a more systematic approach (Calculus on Banach spaces).

The following is another typical example of a functional defined on the space $C^1([0,1])$ of continuously differentiable functions defined on the interval $[0,1]$:

$$S[c] := \int_0^1 \sqrt{1 + (dc/dt)^2}\,dt$$

The reader may recognize this example as the arc length functional. The derivative *at* the function c in the *direction of* a function $h \in C^1([0,1])$ would be given by

$$\left.\delta S\right|_c(h) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(S[c + \varepsilon h] - S[c]\right).$$

It is well known that if $\left.\delta S\right|_c(h) = 0$ for every h then c is a linear function; $c(t) = at + b$. The condition $\left.\delta S\right|_c(h) = 0 = 0$ (for all h) is often simply written as $\delta S = 0$. We shall have a bit more to say about this notation shortly. For examples like this one, the analogy with multi-variable calculus is summarized as

$$\begin{aligned} \text{The index or argument becomes continuous: } & i \rightsquigarrow t \\ d\text{-tuples become functions: } & x^i \rightsquigarrow c(t) \\ \text{Functions of a vector variable become functionals of functions: } & f(\vec{x}) \rightsquigarrow S[c] \end{aligned}$$

Here we move from $d-$tuples (which are really functions with finite domain) to functions with a continuous domain. The function f of x becomes a functional S of functions c.

We now exhibit a common example from the mechanics which comes from considering a bead sliding along a wire. We are supposed to be given a so called "Lagrangian function" $L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ which will be the basic ingredient in building an associated functional. A typical example is of the form $L(x, v) = \frac{1}{2}mv^2 - V(x)$. Define the action functional S by using L as follows: For a given function $t \longmapsto q(t)$ defined on $[a, b]$ let

$$S[q] := \int_a^b L(q(t), \dot{q}(t))dt.$$

We have used x and v to denote variables of L but since we are eventually to plug in $q(t), \dot{q}(t)$ we could also follow the common tradition of denoting these variables by q and $\dot{q}$ but then it must be remembered that we are using these symbols in two ways. In this context, one sometimes sees something like following expression for the so-called variation

$$\delta S = \int \frac{\delta S}{\delta q(t)} \delta q(t) dt \tag{1.1}$$

Depending on one's training and temperament, the meaning of the notation may be a bit hard to pin down. First, what is the meaning of δq as opposed to, say, the differential dq? Second, what is the mysterious $\frac{\delta S}{\delta q(t)}$? A good start might be to go back and settle on what we mean by the differential in ordinary multivariable calculus. For a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ we take df to just mean the map

$$df : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

given by $df(p, h) = f'(p)h$. We may also fix p and write $df|_p$ or $df(p)$ for the linear map $h \mapsto df(p, h)$. With this convention we note that $dx^i\big|_p (h) = h^i$ where $h = (h^1, ..., h^d)$. Thus applying both sides of the equation

$$df|_p = \sum \frac{\partial f}{\partial x^i}(p)\ dx^i\big|_p \tag{1.2}$$

to some vector h we get

$$f'(p)h = \sum \frac{\partial f}{\partial x^i}(p) h^i. \tag{1.3}$$

In other words, $df|_p = D_h f(p) = \nabla f \cdot h = f'(p)$. Too many notations for the same concept. Equation 1.2 is clearly very similar to $\delta S = \int \frac{\delta S}{\delta q(t)} \delta q(t) dt$ and so we expect that δS is a linear map and that $t \mapsto \frac{\delta S}{\delta q(t)}$ is to δS as $\frac{\partial f}{\partial x^i}$ is to df:

$$\begin{aligned} df &\rightsquigarrow \delta S \\ \frac{\partial f}{\partial x^i} &\rightsquigarrow \frac{\delta S}{\delta q(t)}. \end{aligned}$$

Roughly, $\frac{\delta S}{\delta q(t)}$ is taken to be whatever function (or distribution) makes the equation 1.1 true. We often see the following type of calculation

$$\begin{aligned} \delta S &= \delta \int L dt \\ &= \int (\frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q}) dt \\ &= \int \{ \frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \} \delta q dt \end{aligned} \tag{1.4}$$

from which we are to conclude that

$$\frac{\delta S}{\delta q(t)} = \frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}$$

Actually, there is a subtle point here in that we must restrict δS to variations for which the integration by parts is justified. We can make much better sense of things if we have some notion of derivative for functionals defined on some function space. There is also the problem of choosing an appropriate function space. On the one hand, we want to be able to take (ordinary) derivatives of these functions since they may appear in the very definition of S. On the other hand, we must make sense out of limits so we must pick a space of functions with a tractable and appropriate topology. We will see below that it is very desirable to end up with what is called a Banach space. Often one is forced to deal with more general topological vector spaces. Let us ignore all of these worries for a bit longer and proceed formally. If δS is somehow the variation due to a variation $h(t)$ of $q(t)$ then it depends on both the starting position in function space (namely, the function $q(.)$) and also the direction in function space that we move (which is the function $h(.)$). Thus we interpret $\delta q = h$ as some appropriate function and then interpret δS as short hand for

$$\begin{aligned}\left.\delta S\right|_{q(.)} h &:= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(S[q + \varepsilon h] - S[q]) \\ &= \int (\frac{\partial L}{\partial q} h + \frac{\partial L}{\partial \dot{q}} \dot{h}) dt\end{aligned}$$

Note: Here and throughout the book the symbol ":= " is sued to indicate equality by definition.

If we had been less conventional and more cautious about notation we would have used c for the function which we have been denoting by $q : t \mapsto q(t)$. Then we could just write $\left.\delta S\right|_c$ instead of $\left.\delta S\right|_{q(.)}$. The point is that the notation $\left.\delta S\right|_q$ might leave one thinking that $q \in \mathbb{R}$ (which it is under one interpretation!) but then $\left.\delta S\right|_q$ would make no sense. It is arguably better to avoid letting q refer both to a number and to a function even though this is quite common. At any rate, from here we restrict attention to "directions" h for which $h(a) = h(b) = 0$ and use integration by parts to obtain

$$\left.\delta S\right|_{q(.)} h = \int \{\frac{\partial L}{\partial x^i}(q(t), \dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t))\} h^i(t) dt.$$

So it seems that the function $E(t) := \frac{\partial L}{\partial q}(q(t), \dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}(q(t), \dot{q}(t))$ is the right candidate for the $\frac{\delta S}{\delta q(t)}$. However, once again, we must restrict to h which vanish at the boundary of the interval of integration. On the other hand, this family is large enough to force the desired conclusion. Despite this restriction the function $E(t)$ is clearly important. For instance, if $\left.\delta S\right|_{q(.)} = 0$ (or even $\left.\delta S\right|_{q(.)} h = 0$ for all functions that vanish at the end points) then we may conclude easily that $E(t) \equiv 0$. This gives an equation (or system of equations) known as the Euler-Lagrange equation for the function $q(t)$ corresponding to the action functional

S :

$$\frac{\partial L}{\partial q}(q(t),\dot{q}(t))-\frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q(t),\dot{q}(t))=0$$

Exercise 1.48 *Replace $S[c]=\int L(c(t),\dot{c}(t))dt$ by the similar function of several variables $S(c_1,...c_N)=\sum L(c_i,\triangle c_i)$. Here $\triangle c_i := c_i - c_{i-1}$ (taking $c_0 = c_N$) and L is a differentiable map $\mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$. What assumptions on $c = (c_1,...c_N)$ and $h = (h_1,...h_N)$ justify the following calculation?*

$$\begin{aligned}
dS|_{(c_1,...c_N)}\, h &= \sum \frac{\partial L}{\partial c_i}h^i + \frac{\partial L}{\partial \triangle c_i}\triangle h^i \\
&= \sum \frac{\partial L}{\partial c_i}h^i + \sum \frac{\partial L}{\partial \triangle c_i}h^i - \sum \frac{\partial L}{\partial \triangle c_i}h^{i-1} \\
&= \sum \frac{\partial L}{\partial c_i}h^i + \sum \frac{\partial L}{\partial \triangle c_i}h^i - \sum \frac{\partial L}{\partial \triangle c_{i+1}}h^i \\
&= \sum \frac{\partial L}{\partial c_i}h^i - \sum (\frac{\partial L}{\partial \triangle c_{i+1}} - \frac{\partial L}{\partial \triangle c_i})h^i \\
&= \sum \{\frac{\partial L}{\partial c_i}h^i - (\triangle \frac{\partial L}{\partial \triangle c_i})\}h^i \\
&= \sum \frac{\partial S}{\partial c_i}h^i.
\end{aligned}$$

The upshot of our discussion is that the δ notation is just an alternative notation to refer to the differential or derivative. Note that q^i might refer to a coordinate or to a function $t \mapsto q^i(t)$ and so dq^i is the usual differential and maps $\mathbb{R}^d$ to $\mathbb{R}$ whereas $\delta x^i(t)$ is either taken as a variation function $h^i(t)$ as above or as the map $h \mapsto \delta q^i(t)(h) = h^i(t)$. In the first interpretation $\delta S = \int \frac{\delta S}{\delta q^i(t)}\delta q^i(t)dt$ is an abbreviation for $\delta S(h) = \int \frac{\delta S}{\delta q^i(t)}h^i(t)dt$ and in the second interpretation it is the map $\int \frac{\delta S}{\delta q^i(t)}\delta q^i(t)dt : h \mapsto \int \frac{\delta S}{\delta q^i(t)}(\delta q^i(t)(h))dt = \int \frac{\delta S}{\delta q^i(t)}h^i(t)dt$. The various formulas make sense in either case and both interpretations are ultimately equivalent. This much the same as taking the dx^i in $df = \frac{\partial f}{\partial x i}dx^i$ to be components of an arbitrary vector $(dx^1,...,dx^d)$ or we may take the more modern view that dx^i is a linear map given by $dx^i : h \mapsto h^i$. If this seems strange recall that x^i itself is also interpreted both as a number and as a coordinate function.

Example 1.49 *Let $F[c] := \int_{[0,1]} c^2(t)dt$ as above and let $c(t) = t^3$ and $h(t) =$*

$\sin(t^4)$. *Then*

$$\begin{aligned}\delta F|_c(h) &= D_h F(c) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(F[c+\varepsilon h] - F[c]\right)\\ &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} F[c+\varepsilon h]\\ &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}\int_{[0,1]}(c(t)) + \varepsilon h(t))^2 dt\\ &= 2\int_{[0,1]} c(t)h(t)dt = 2\int_0^1 t^3\sin(\pi t^4)dx\\ &= \frac{1}{\pi}\end{aligned}$$

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

Exercise 1.50 *Notice that for a smooth function* $s : \mathbb{R} \to\mathbb{R}$ *we may write*

$$\frac{\partial s}{\partial x^i}(x_0) = \lim_{h\to 0}\frac{s(x_0 + he_i) - s(x_0)}{h}$$

$$\text{where } e_i = (0, ..., 1, ...0)$$

Consider the following similar statement which occurs in the physics literature quite often.

$$\frac{\delta S}{\delta c(t)} = \lim_{h\to 0}\frac{S[c + h\delta_t] - S[c]}{h}$$

Here δ_t *is the Dirac delta function (distribution) with the defining property* $\int \delta_t\phi = \phi(t)$ *for all continuous* ϕ. *To what extent is this rigorous? Try a formal calculation using this limit to determine* $\frac{\delta S}{\delta c(t)}$ *in the case that*

$$S(c) := \int_0^1 c^3(t)dt.$$

1.2.1 Lagrange Multipliers and Ljusternik's Theorem

`Note: This section is under construction.`

The next example show how to use Lagrange multipliers to handle constraints.

Example 1.51 *Let* E *and* F *and* F_0 *be as in the previous example. We define two functionals*

$$\begin{aligned}\mathcal{F}[f] &:= \int_D \nabla f\cdot\nabla f dx\\ \mathcal{C}[f] &= \int_D f^2 dx\end{aligned}$$

We want a necessary condition on f such that f extremizes $\mathcal{D}$ subject to the constraint $\mathcal{C}[f] = 1$. The method of Lagrange multipliers applies here and so we have the equation $D\mathcal{F}|_f = \lambda\, D\mathcal{C}|_f$ which means that

$$\langle \frac{\delta \mathcal{F}}{\delta f}, h\rangle = \lambda \langle \frac{\delta \mathcal{C}}{\delta f}, h\rangle \text{ for all } h \in C_c^2(D)$$

or

$$\frac{\delta \mathcal{F}}{\delta f} = \lambda \frac{\delta \mathcal{C}}{\delta f}$$

After determining the functional derivatives we obtain

$$-\nabla^2 f = \lambda f$$

This is not a very strong result since it is only a necessary condition and only hints at the rich spectral theory for the operator ∇^2.

Theorem 1.52 *Let E and F be Banach spaces and $U \subset \mathsf{E}$ open with a differentiable map $f : U \to \mathsf{F}$. If for $x_0 \in U$ with $y_0 = f(x_0)$ we have that $Df|_{x_0}$ is onto and $\ker Df|_{x_0}$ is complemented in E then the set $x_0 + \ker Df|_{x_0}$ is tangent to the level set $f^{-1}(y_0)$ in the following sense: There exists a neighborhood $U' \subset U$ of x_0 and a homeomorphism $\phi : U' \to V$ where V is another neighborhood of x_0 and where $\phi(x_0 + h) = x_0 + h + \varepsilon(h)$ for some continuous function ε with the property that*

$$\lim_{h \to 0} \frac{\|\varepsilon(h)\|}{\|h\|} = 0.$$

Proof. $Df|_{x_0}$ is surjective. Let $K := \ker Df|_{x_0}$ and let L be the complement of K in E. This means that there are projections $p : \mathsf{E} \to K$ and $q : \mathsf{E} \to L$

$$p^2 = p \text{ and } q^2 = q$$
$$p + q = id$$

Let $r > 0$ be chosen small enough that $x_0 + B_r(0) + B_r(0) \subset U$. Define a map

$$\psi : K \cap B_r(0) \times L \cap B_r(0) \to \mathsf{F}$$

by $\psi(h_1, h_2) := f(x_0 + h_1 + h_2)$ for $h_1 \in K \cap B_r(0)$ and $h_2 \in L \cap B_r(0)$. We have $\psi(0, 0) = f(x_0) = y_0$ and also one may verify that ψ is C^1 with $\partial_1 \psi = Df(x_0)|\,K = 0$ and $\partial_2 \psi = Df(x_0)|\,L$. Thus $\partial_2 \psi : L \to \mathsf{F}$ is a continuous isomorphism (use the open mapping theorem) and so we have a continuous linear inverse $(\partial_2 \psi)^{-1} : \mathsf{F} \to L$. We may now apply the implicit function theorem to the equation $\psi(h_1, h_2) = y_0$ to conclude that there is a locally unique function $\varepsilon : K \cap B_\delta(0) \to L$ for small $\delta > 0$ (less than r) such that

$$\psi(h, \varepsilon(h)) = y_0 \text{ for all } h \in K \cap B_\delta(0)$$
$$\varepsilon(0) = 0$$
$$D\varepsilon(0) = -\left(\partial_2 \psi\right)^{-1} \circ \partial_1 \psi|_{(0,0)}$$

But since $\partial_1 \psi = Df(x_0)| K = 0$ this last expression means that $D\varepsilon(0) = 0$ and so

$$\lim_{h \to 0} \frac{\|\varepsilon(h)\|}{\|h\|} = 0$$

Clearly the map $\phi : (x_0 + K \cap B_\delta(0)) \to \mathsf{F}$ defined by $\phi(x_0 + h) := x_0 + h + \varepsilon(h)$ is continuous and also since by construction $y_0 = \psi(h, \varepsilon(h)) = \phi(x_0 + h + \varepsilon(h))$ we have that ϕ has its image in $f^{-1}(y_0)$. Let the same symbol ϕ denote the map $\phi : (x_0 + K \cap B_\delta(0)) \to f^{-1}(y_0)$ which only differs in its codomain. Now h and $\varepsilon(h)$ are in complementary subspaces and so ϕ must be injective. Thus its restriction to the set $V := \{x_0 + h + \varepsilon(h) : h \in K \cap B_\delta(0)$ is invertible and in fact we have $\phi^{-1}(x_0 + h + \varepsilon(h)) = x_0 + h$. That V is open follows from the way we have used the implicit function theorem. Now recall the projection p. Since the range of p is K and its kernel is L we have that $\phi^{-1}(x_0 + h + \varepsilon(h)) = x_0 + p(h + \varepsilon(h))$ and we see that ϕ^{-1} is continuous on V. Thus ϕ (suitably restricted) is a homeomorphism of $U' := x_0 + K \cap B_\delta(0)$ onto $V \subset f^{-1}(y_0)$. We leave it to the reader to provide the easy verification that ϕ has the properties claimed by statement of the theorem. ■

1.3 Problem Set

1. Find the matrix that represents (with respects to standard bases) 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

   $$F[f](x) = \int_0^1 k(x,y) \, [f(y)]^2 \, dy$$

   where $k(x,y)$ is a bounded continuous function on $[0,1] \times [0,1]$.

3. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable and define $F : C[0,1] \to C[0,1]$ by

   $$F(g) := f \circ g$$

   Show that F is differentiable and $DF|_g : C[0,1] \to C[0,1]$ is the linear map given by $\left(DF|_g \cdot u \right)(t) = f'(g(t)) \cdot u(t)$.

4. a) Let U be an open subset of $\mathbb{R}^n$ with $n > 1$. Show that if $f : U \to \mathbb{R}$ is continuously differentiable then f cannot be injective. Hint: Use the mean value theorem and the implicit mapping theorem.

   b) Show that if U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^k$ is continuously differentiable then f cannot be injective unless $k \geq n$. Hint: Look for a way to reduce it to part (a).

5. Let $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be C^∞ and define
$$S[c] = \int_0^1 L(c(t), c'(t), t)dt$$
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
$$DS|_c \cdot b = \int_0^1 \langle g_c(t), b(t)\rangle dt$$
or in other words,
$$DS|_c \cdot b = \int_0^1 \sum_{i=1}^n g_c^i(t)b^i(t)dt.$$

6. In the last problem, if we had not 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).

   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
$$\mathcal{B}_{\vec{x}_0\vec{x}_1} = \{c : c(0) = \vec{x}_0 \text{ and } c(1) = \vec{x}_1\}$$
$$\mathcal{B} = \{c : c(0) = 0 \text{ and } c(1) = 0\}$$
   Then we have $T_{c_0}\mathcal{B}_{\vec{x}_0\vec{x}_1} \cong \mathcal{B}$. Thus we should consider $DS|_{c_0} : \mathcal{B} \to \mathcal{B}$.

7. Let $\mathrm{Fl}_t(.)$ be defined by $\mathrm{Fl}_t(x) = (t + 1)x$ for $t \in (-1/2, 1/2)$ and $x \in \mathbb{R}^n$. Assume that the map is jointly C^1 in both variable. Find the derivative of
$$f(t) = \int_{D(t)} (tx)^2\, dx$$
at $t = 0$, where $D(t) := \mathrm{Fl}_t(D)$ the image of the disk $D = \{|x| \le 1\}$.

   Hint: Find the Jacobian $J_t := \det[DFl_t(x)]$ and then convert the integral above to one just over $D(0) = D$.

8. Let $M_{n\times n}(\mathbb{R})$ be the vector space of $n \times n$ matrices with real entries and let $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ be the determinant map. The derivative at the identity element I should be a linear map $D\det(I) : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$. Show that $D\det(I)\cdot B = Tr(B)$. More generally, show that $D\det(A)\cdot B = Tr((\operatorname{cof} A)^t\, B)$ where $\operatorname{cof} A$ is the matrix of cofactors of A.

   What is $\frac{\partial}{\partial x_{ij}} \det X$ where $X = (x_{ij})$?

9. 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.

10. Show that if F is any closed subset of $\mathbb{R}^n$ there is a C^∞-function f whose zero set $\{x : f(x) = 0\}$ is exactly F.

11. Let U be an open set in $\mathbb{R}^n$. For $f \in C^k(U)$ and $S \subset U$ a compact set, let $\|f\|_k^S := \sum_{|\alpha| \leq k} \sup_{x \in S} \left| \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x) \right|$. a) Show that (1) $\|rf\|_k^S = |r| \, \|f\|_k^S$ for any $r \in \mathbb{R}$, (2) $\|f_1 + f_2\|_k^S \leq \|f_1\|_k^S + \|f_2\|_k^S$ for any $f_1, f_2 \in C^k(U)$, (3) $\|fg\|_k^S \leq \|f\|_k^S \, \|g\|_k^S$ for $f, g \in C^k(U)$.
b) Let $\{K_i\}$ be a compact subsets of U such that $U = \bigcup_i K_i$. Show that $d(f, g) := \sum_{i=0}^\infty \frac{1}{2^i} \frac{\|f-g\|_k^{K_i}}{1+\|f-g\|_k^{K_i}}$ defines a complete metric space structure on $C^k(U)$.

12. Let E and F be real Banach spaces. A function $f : \mathsf{E} \to \mathsf{F}$ is said to be homogeneous of degree k if $f(rx) = rf(x)$ for all $r \in \mathbb{R}$ and $x \in \mathsf{E}$. Show that if f is homogeneous of degree k and is differentiable, then $Df(v) \cdot v = kf(v)$.

13. Show that the implicit mapping theorem implies the inverse mapping theorem. Hint: Consider $g(x, y) = f(x) - y$ for $f : U \to \mathsf{F}$.

Chapter 2

The Language of Category Theory

Category theory provides a powerful means of organizing our thinking in mathematics. Some readers may be put off by the abstract nature of category theory. To such readers I can only say that it is not really difficult to catch on to the spirit of category theory and the payoff in terms of organizing mathematical thinking is considerable. I encourage these readers to give it a chance. In any case, it is not strictly necessary for the reader to be completely at home with category theory before going further into the book. In particular, physics and engineering students may not be used to this kind of abstraction and should simply try to gradually become accustomed to the language. Feel free to defer reading this appendix on Category theory until it seems necessary.

Roughly speaking, category theory is an attempt at clarifying structural similarities that tie together different parts of mathematics. A category has "objects" and "morphisms". The prototypical category is just the category **Set** which has for its objects ordinary *sets* and for its morphisms *maps* between sets. The most important category for differential geometry is what is sometimes called the "smooth category" consisting of smooth manifolds and smooth maps. (The definition of these terms is given in the text proper but roughly speaking smooth means differentiable.)

Now on to the formal definition of a category.

Definition 2.1 *A **category** $\mathfrak{C}$ is a collection of objects* $\mathrm{Ob}(\mathfrak{C}) = \{X, Y, Z, ...\}$ *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 morphisms in a category $\mathfrak{C}$ will be denoted* $\mathrm{Mor}(\mathfrak{C})$. *In addition, a category is required to have a **composition** law which is defined as 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 2.2 (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 2.3 (Cat2) *The composition law is associative:* $f \circ (g \circ h) = (f \circ g) \circ h$*.*

Axiom 2.4 (Cat3) *Each set of morphisms of the form* $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ *must contain a necessarily element* id_X*, the identity element, 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 2.5 *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 by* $X \xrightarrow{f} Y$*.*

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 extra structure on the sets and the maps, this will be almost the only type of category we shall need to talk about. On the other hand there are plenty of interesting categories of this type. Examples include the following.

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. $\mathbf{Lin}_{\mathbb{F}}$: The objects are vector spaces over the field $\mathbb{F}$ and the morphisms are linear maps. This category is referred to as the linear category or the vector space category (over the field $\mathbb{F}$).

4. **Top**: The objects are topological spaces and the morphisms are continuous maps.

5. $\mathbf{Man}^r$**:** 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 2.6 *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":*

$$\begin{tikzcd}
X_1 \arrow[r, "f_1"] \arrow[d, "g_X"'] & Y_1 \arrow[d, "g_Y"] \\
X_2 \arrow[r, "f_2"'] & Y_2
\end{tikzcd}$$

Similarly, if $h \circ f = g$ *we say that the diagram*

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[dr, "g"'] & Y \arrow[d, "h"] \\
 & Z
\end{tikzcd}$$

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 that begin and end at the same object are equal.*

Definition 2.7 *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* ***monomorphism****. Similarly, if $f : X \to Y$ is "right cancellable" we call f an* ***epimorphism****. A morphism that is both a monomorphism and an epimorphism is called an* ***isomorphism.*** *If the category needs to be specified then we talk about a* ***$\mathfrak{C}$-monomorphism,*** ***$\mathfrak{C}$-epimorphism*** *and so on).*

In some cases we will use other terminology. For example, an isomorphism in the smooth category is called a **diffeomorphism**. In the linear category, we speak of linear maps and linear isomorphisms. Morphisms which comprise $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ are also called endomorphisms and so we also write $\mathrm{End}_{\mathfrak{C}}(X) := \mathrm{Hom}_{\mathfrak{C}}(X, X)$. The set of all isomorphisms in $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ is sometimes denoted by $\mathrm{Aut}_{\mathfrak{C}}(X)$ and these morphisms are called **automorphisms**.

We single out the following: In many categories like the above we can form a new category that uses the notion of pointed space and pointed map. For example, we have the "pointed topological category" . A pointed topological space is an topological space X together with a distinguished point p. Thus a typical object in the pointed topological category would be written (X, p). A morphism $f : (X, p) \to (W, q)$ is a continuous map such that $f(p) = q$.

2.0.1 Functors

A **functor** $\mathcal{F}$ is a pair of maps, both denoted by the same letter $\mathcal{F}$, that map objects and morphisms from one category to those of another

$$\mathcal{F} : \mathrm{Ob}(\mathfrak{C}_1) \to \mathrm{Ob}(\mathfrak{C}_2)$$
$$\mathcal{F} : \mathrm{Mor}(\mathfrak{C}_1) \to \mathrm{Mor}(\mathfrak{C}_2)$$

such that composition and identity morphisms are respected: This means that for a morphism $f : X \to Y$, the morphism

$$\mathcal{F}(f) : \mathcal{F}(X) \to \mathcal{F}(Y)$$

is a morphism in the second category and we must have

1. $\mathcal{F}(\mathrm{id}_{\mathfrak{C}_1}) = \mathrm{id}_{\mathfrak{C}_2}$

2. If $f : X \to Y$ and $g : Y \to Z$ then $\mathcal{F}(f) : \mathcal{F}(X) \to \mathcal{F}(Y)$, $\mathcal{F}(g) : \mathcal{F}(Y) \to \mathcal{F}(Z)$ and
$$\mathcal{F}(g \circ f) = \mathcal{F}(g) \circ \mathcal{F}(f).$$

Example 2.8 *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*

$$\begin{aligned} \ell &\mapsto \ell_{\mathbb{C}} \\ \mathrm{V} &\mapsto \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V} \end{aligned}$$

define a functor from $\mathbf{Lin}_{\mathbb{R}}$ *to* $\mathbf{Lin}_{\mathbb{C}}$*.*

Remark 2.9 *In practice, complexification amounts to simply allowing complex scalars. For instance, we might just write cv instead of* $c \otimes v$*.*

Actually, what we have defined here is a **covariant functor**. A **contravariant functor** is defined similarly except that the order of composition is reversed so that instead of **Funct2** above we would have $\mathcal{F}(g \circ f) = \mathcal{F}(f) \circ \mathcal{F}(g)$. An example of a contravariant functor is the dual vector space functor which is a functor from the category of vector spaces $\mathbf{Lin}_{\mathbb{R}}$ to itself which sends each space V to its dual V^* and each linear map to its dual (or transpose). Under this functor a morphism

$$\mathrm{V} \xrightarrow{L} \mathrm{W}$$

is sent to the morphism

$$\mathrm{V}^* \xleftarrow{L^*} \mathrm{W}^*$$

Notice the arrow reversal.

Remark 2.10 *One of the most important functors for our purposes is the* ***tangent functor*** *defined in chapter* **??***. Roughly speaking this functor replaces differentiable maps and spaces by their linear parts.*

Example 2.11 *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:*

$$\begin{tikzcd} V \arrow[d, "A"'] & & W^* \arrow[d, "A^*"'] & & V^{**} \arrow[d, "A^{**}"'] \\ W & \mapsto & V^* & \mapsto & W^{**} \end{tikzcd}$$

2.0.2 Natural transformations

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 $\mathfrak{C}_1$, the following diagram commutes:

$$\begin{tikzcd}
\mathcal{F}_1(X) \arrow[r, "\mathcal{T}(X)"] \arrow[d, "\mathcal{F}_1(f)"'] & \mathcal{F}_2(X) \arrow[d, "\mathcal{F}_2(f)"] \\
\mathcal{F}_1(Y) \arrow[r, "\mathcal{T}(Y)"'] & \mathcal{F}_2(Y)
\end{tikzcd}$$

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 $** : \mathbf{Lin}_{\mathbb{R}} \to \mathbf{Lin}_{\mathbb{R}}$:

$$\begin{tikzcd}
\mathrm{V} \arrow[r, "\iota(\mathrm{V})"] \arrow[d, "f"'] & \mathrm{V}^{**} \arrow[d, "f^{**}"] \\
\mathrm{W} \arrow[r, "\iota(\mathrm{W})"'] & \mathrm{W}^{**}
\end{tikzcd} .$$

The map $\mathrm{V} \to \mathrm{V}^{**}$ sends a vector to a linear function $\widetilde{v} : \mathrm{V}^* \to \mathbb{R}$ defined by $\widetilde{v}(\alpha) := \alpha(v)$ (the hunted becomes the hunter 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 $\mathcal{F}_1(X)$ is naturally isomorphic to $\mathcal{F}_2(X)$. 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^* where the construction of such an isomorphism involves an arbitrary choice of a basis.

Chapter 3

Overview of Classical Physics

3.0.3 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 Newtons and decomposes into elementary units as kg$\times$m$\times$sec^{-2}.

3.0.4 Newtons 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 3.1 *A map $g : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R} \times \mathbb{R}^3$ is called a Galilean transformation if and only if 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 if and only if $(T, \mathbf{X}) = g(t, \mathbf{x})$ 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 $\mathbf{p} = (m_1x_1, m_1y_1, m_1z_1, ..., m_nx_n, m_ny_n, m_nz_n) \in \mathbb{R}^{3n}$. 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 (Newtons in the SI system). If all bodies involved are taken into account then the force **F** cannot depend explicitly on time as can be deduced by the assumption that the form taken by **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 **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 **F** will not be invariant under all spatial rotations of coordinate axes since now there is a preferred direction (up-down).

3.0.5 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 **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:

$$\begin{aligned}\delta S(\mathbf{q}(t)) &= \delta \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}})dt \\ &\int_{t_1}^{t_2} \left[\delta\mathbf{q}\cdot \frac{\partial}{\partial \mathbf{q}} L(\mathbf{q}, \dot{\mathbf{q}}) + \delta\dot{\mathbf{q}}\cdot \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}})\right] dt \\ &= \int_{t_1}^{t_2} \delta\mathbf{q}\cdot \left(\frac{\partial}{\partial \mathbf{q}} L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}})\right) dt + \left[\delta\mathbf{q}\cdot \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}})\right]_{t_1}^{t_2}\end{aligned}$$

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 3.2 *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

$$\begin{aligned}
\frac{d}{dt}E &= \frac{d}{dt}\sum \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L(\mathbf{q}, \dot{\mathbf{q}}) \\
&= \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i - \dot{q}^i\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{d}{dt}L(\mathbf{q}, \dot{\mathbf{q}}) \\
&= \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i - \dot{q}^i\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i}\dot{q}^i - \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i \\
&= 0 \text{ by the Euler Lagrange equations.} \qquad (3.1)
\end{aligned}$$

Conclusion 3.3 *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 3.4 *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

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}m\,|\dot{\mathbf{x}}|^2$$

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

$$\begin{aligned}\frac{d}{ds}L(e^{sA}\mathbf{x},e^{sA}\dot{\mathbf{x}}) &= \frac{d}{dt}(\frac{1}{2}m\left|e^{sA}\dot{\mathbf{x}}\right|^2 - V(e^{sA}\mathbf{x})) \\ &= \frac{d}{dt}(\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2 - V(\mathbf{x})) = 0\end{aligned}$$

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

$$\begin{aligned}\left[A\mathbf{x}\cdot\frac{\partial}{\partial\dot{\mathbf{x}}}(\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2 - V(\mathbf{x}))\right]_{t_1}^{t_2} &= 0 \\ [mA\mathbf{x}\cdot\dot{\mathbf{x}}]_{t_1}^{t_2} &= 0\end{aligned}$$

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 3.5 *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 3.6 *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

$$\begin{aligned} H(\mathbf{q}, \mathbf{p}) &= \sum p_i \dot{q}^i - L(\mathbf{q}, \dot{\mathbf{q}}) \\ p_i &= \frac{\partial L}{\partial \dot{q}^i} \end{aligned}$$

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

$$\begin{aligned} \frac{dq^i}{dt} &= \frac{\partial H}{\partial p_i} \\ \frac{dp^i}{dt} &= -\frac{\partial H}{\partial q^i} \end{aligned}$$

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 **??** 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 3.7 *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****.*

3.0.6 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.

2. A somewhat general 1-dimensional system is given by a Lagrangian of the form
$$L = \frac{1}{2}a(q)\dot{q}^2 - V(q) \tag{3.2}$$
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

$$mD_t^2\mathbf{X} = \mathbf{F}(\mathbf{X}, \mathbf{V}) \qquad (3.3)$$

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 $Q \in \mathrm{SO}(3)$ 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:

$$\begin{aligned} \mathbf{x} &\leftrightarrows \mathsf{x} \in \mathfrak{so}(3) \text{ and} \\ \mathbf{X} &\leftrightarrows \mathsf{X} \ \in \mathfrak{so}(3) \text{ etc.} \end{aligned}$$

	$\mathbb{R}^3$		$\mathfrak{so}(3)$
Inertial frame	$\mathbf{x}$	$\leftrightarrows$	x
Moving frame	$\mathbf{X}$	$\leftrightarrows$	X

Then we have the following chart showing how various operations match up:

$$\begin{array}{ccc}
\mathbf{x} = \mathrm{O}\mathbf{X} & \leftrightarrows & \mathsf{x} = \mathsf{O}\mathsf{X}\mathsf{O}^t \\
\mathbf{v}_1 \times \mathbf{v}_2 & \leftrightarrows & [\mathsf{v}_1, \mathsf{v}_2] \\
\mathbf{v} = \dot{\mathbf{x}} & \leftrightarrows & \mathsf{v}{=}\dot{\mathsf{x}} \\
\mathbf{V} = D_t\mathbf{X} = \dot{\mathbf{X}} + \Omega(t)\mathbf{X} & \leftrightarrows & \mathsf{V}{=}D_t\mathsf{X} = \dot{\mathsf{X}}{+}[\Omega(t), \mathsf{X}] \\
\ell = \mathbf{x} \times m\dot{\mathbf{x}} & \leftrightarrows & \mathsf{l} = [\mathsf{x}, m\dot{\mathsf{x}}] \\
\ell = \mathrm{O}\mathbf{L} & \leftrightarrows & \mathsf{l} = \mathsf{O}\mathsf{L}\mathsf{O}^t = [\mathsf{V}, \Omega(t)] \\
D_t\mathbf{L} = \dot{\mathbf{L}}{+}\Omega(t) \times \mathbf{L} & \leftrightarrows & D_t\mathsf{L} = \dot{\mathsf{L}} + [\Omega(t), \mathsf{L}]
\end{array}$$

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.

3.0.7 The Basic Ideas of Relativity

3.0.8 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

$$S = \int\int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi, \partial\phi)d^3xdt = \int_{V\times I\subset\mathbb{R}^4} \mathcal{L}(\phi, \partial\phi)d^4x$$

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

$$\begin{array}{lll}
i & \rightarrowtail\rightarrowtail\rightarrowtail & \vec{x} \\
q & \rightarrowtail\rightarrowtail\rightarrowtail & \phi \\
q^i & \rightarrowtail\rightarrowtail\rightarrowtail & \phi(., \vec{x}) \\
q^i(t) & \rightarrowtail\rightarrowtail\rightarrowtail & \phi(t, \vec{x}) = \phi(x) \\
p^i(t) & \rightarrowtail\rightarrowtail\rightarrowtail & \partial_t\phi(t, \vec{x}) + \nabla_{\vec{x}}\phi(t, \vec{x}) = \partial\phi(x) \\
L(q, p) & \rightarrowtail\rightarrowtail\rightarrowtail & \int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi, \partial\phi)d^3x \\
S = \int L(\mathbf{q}, \dot{\mathbf{q}})dt & \rightarrowtail\rightarrowtail\rightarrowtail & S = \int\int \mathcal{L}(\phi, \partial\phi)d^3xdt
\end{array}$$

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. In other words, a system with an infinite number of 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.

3.0.9 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 $\nabla \cdot \overrightarrow{\mathbf{j}} = \mathrm{div}(\overrightarrow{\mathbf{j}})$ 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.

3.0.10 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 3.8 *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 3.9 *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

$$\begin{aligned} \operatorname{curl} \tilde{\mathbf{B}} &= 0 \\ \operatorname{div} \tilde{\mathbf{E}} &= 4\pi\rho. \end{aligned}$$

Chapter 4

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 $F_\mu^\nu = \eta^{s\nu} F_{\mu s}$ and $F^{\mu\nu} = \eta^{s\mu} F_s^\nu$. 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

$$S(\tilde{\mathbf{x}}) = \int_a^b (-mc\frac{d\tau}{dt} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt})dt$$

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 4.1 *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 $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ 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 we 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 4.2 *Apply Gauss' theorem and Stokes' theorem to the first two Maxwell's equations to get the integral forms of the equations. 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 can 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 instead 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

$$\begin{aligned} \operatorname{curl}\tilde{\mathbf{B}} &= 0 \\ \operatorname{div}\tilde{\mathbf{E}} &= 4\pi\rho. \end{aligned}$$

Accordingly, if M only one isomorphism of Banach spaces is implicated by an atlas.

For completeness we add one more technical definoition. If a C^r differentiable manifold is specified by an atlas with values in a fixed Banach space E we say that it is **modeled** on E or that it is an E-manifold (of class C^r).

Finite dimensional manifolds are far more commonly studied than those modeled on more general Banach spaces and in the literature the notion of a manifold is usually defined in such a way as to be automatically finite dimensional. Furthermore manifolds are usual required to satisfy further topological requirements. This is discussed in the following.

Remark 4.3 *In the literature the definition of a finite dimensional differentiable manifold usually includes the requirement that the topology is Hausdorff and second coutable. Recall that our definition of paracompact requires the space to be Hausdorff. The main reason that second coutability is assumed because second coutability implies paracompactness and paracompactness allows the construction of the so called smooth partitions of unity (discussed later). For finite dimensional manifolds, paracompact is equiavalent to each connected component being second countable. A finite dimensional paracompact manifold with a finite or countably infinite number of components would also be second coutable. Some authors take the alternative route of defining a finite dimensional differentiable manifold to be paracompact. However, this leave open the possibility of an uncountable number of connected components and this would create problems for the celebrated theorem of Sard.*

On the other hand, an infinite dimensional Banach space need not be paracompact itself and even when it is we do not seem to automatically get the existence of smooth (C^r with $r > 1$) partitions of unity. It seems then that as long as differentiable manifolds modeled on general Banach spaces are under consideration the motivation for putting extra topological conditions into the very definition of a differentiable manifold is undercut.

4.0.11 Maxwell's equations.

$\mathbb{R}^{3,1}$ is just $\mathbb{R}^4$ but with the action of the symmetry group $O(1,3)$.

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

$$F = B + E \wedge dt.$$

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:

$$\begin{aligned} 0 &= dF = d(B + E \wedge dt) = dB + dE \wedge dt \\ &= (\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}) + d\,(E_x dx + E_y dy + E_z dz) \wedge dt \\ &= (\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}) dx \wedge dy \wedge dz + \frac{\partial B}{\partial t} \wedge dt + \\ &+ \left[(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}) dy \wedge dz + (\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}) dz \wedge dx + (\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}) dy \wedge dx \right] \wedge dt \end{aligned}$$

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 $dd = 0$ since we assuming that there is a potential A! This equation does not involve the scalar product structure encoded by the matrix η.

As for the second pair of Maxwell's equations, they too combine to give a single equation. The appropriate star operator is given by

Definition 4.4 *Define $\epsilon(\mu)$ to be entries of the diagonal matrix $\eta = diag(-1, 1, 1, 1)$. 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 $\Omega^k(\mathbb{R}^4) \to \Omega^{4-k}(\mathbb{R}^4)$. More simply and explicitly*

Exercise 4.5 *Show that* $* \circ *$ *acts on* $\Omega^k(\mathbb{R}^4)$ *by* $(-1)^{k(4-k)+1}$.

Exercise 4.6 *Show that if* F *is the electromagnetic field tensor defined above then*

$$*F = (*F)_{\mu\nu}\, dx^\mu \wedge dx^\nu$$

where $(*F)_{\mu\nu}$ *are the components of the matrix*

$$(*F)_{\mu\nu} = \begin{bmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & E_z & -E_y \\ -B_y & -E_z & 0 & E_x \\ -B_z & E_y & -E_x & 0 \end{bmatrix}$$

Now let J be the differential 1-form constructed from the 4-current $\mathsf{j} = (\rho, \tilde{\mathbf{j}})$ introduced in section 3.0.10 by letting $(j_0, j_1, j_2, j_3) = (-\rho, \tilde{\mathbf{j}})$ and then setting $J = j_\mu dx^\mu$.

Now we add the second equation

$$*d * F = J.$$

Exercise 4.7 *Show that the single differential form equation* $*d * F = J$ *is equivalent to Maxwell's second two equations*

$$\begin{aligned} \operatorname{curl}(\tilde{\mathbf{B}}) &= \frac{\partial \tilde{\mathbf{E}}}{\partial t} + \tilde{\mathbf{j}} \\ \operatorname{div}(\tilde{\mathbf{E}}) &= \rho. \end{aligned}$$

In summary, we have that Maxwell's 4 equations (in free space) in the formalism of differential forms and the Hodge star operator are simply the pair

$$\begin{aligned} dF &= 0 \\ *d * F &= J. \end{aligned}$$

The first equation is equivalent to Maxwell's first two equations and interestingly does not involve the metric structure of space $\mathbb{R}^3$ or the metric structure of spacetime $\mathbb{R}^{1,3}$. The second equation above is equivalent to Maxwell's second two equations an through the star operator essentiality involves the Metric structure of $\mathbb{R}^{1,3}$.

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 $\nabla \cdot \overrightarrow{\mathbf{j}} = \mathrm{div}(\overrightarrow{\mathbf{j}})$ 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”.

Chapter 5

Time Dependent Fields

Definition 5.1 *A C^∞ **time dependent vector field** on M is a C^∞ 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^∞ vector field.*

Similarly we can consider time dependent functions and tensors fields.

Definition 5.2 *Let X be a time dependent vector field. A curve $c : (a,b) \to M$ is called an **integral curve** of X if and only if*

$$\dot{c}(t) = X(t, c(t)) \text{ for all } t \in (a,b).$$

One can study time dependent vector fields by studying their so called **suspensions**. Let $pr_1 : (a,b) \times M \to (a,b)$ and $pr_2 : (a,b) \times M \to M$ be the projection maps. Let $\widetilde{X} \in \mathfrak{X}((a,b) \times M)$ be defined by $\widetilde{X}(t,p) = (\frac{\partial}{\partial t}, X(t,p)) \in T_t(a,b) \times T_pM = T_{(t,p)}((a,b) \times M)$. The vector field $\widetilde{X}$ is called the suspension of X. It can be checked quite easily that if $\widetilde{c}$ is an integral curve of $\widetilde{X}$ then $c := pr_2 \circ \widetilde{c}$ is an integral curve of the time dependent field X. This allows us to apply what we know about integral curves to the time dependent case.

Definition 5.3 *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 that 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 5.4 *Let X be a time dependent vector field. Suppose that $X_t \in \mathfrak{X}(M)$ for each t and that $X : (a,b) \times M \to TM$ is continuous. Then $\Phi^X_{t,s}$ is C^∞ and we have $\Phi^X_{s,a} \circ \Phi^X_{a,t} = \Phi^X_{s,t}$ whenever defined.*

Exercise 5.5 *If $\Phi^X_{t,s}$ is the evolution operator of X then the flow of the suspension $\widetilde{X}$ is given by*

$$\Phi(t,(s,p)) := (t+s, \Phi^X_{t+s,s}(p))$$

Let $\phi_t(p) := \Phi_{0,t}(p)$. Is it true that $\phi_s \circ \phi_t(p) = \phi_{s+t}(p)$? The answer is that in general this equality does *not* hold. The evolution of a time dependent vector field does *not* give rise to a local 1-parameter group of diffeomorphisms. On the other hand, we do have

$$\Phi_{s,r} \circ \Phi_{r,t} = \Phi_{s,t}$$

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.

Definition 5.6 *A time dependent vector field X is called complete if $\Phi^X_{t,s}(p)$ is defined for all $s,t \in \mathbb{R}$ and $p \in M$.*

If f is a time dependent function and Y a time dependent field then $f_t := f(t,.)$ and $Y_t := Y(t,.)$ are vector fields on M for each t. We often omit the t subscript in order to avoid clutter. For example, in the following $(\Phi^X_{t,s})^*f$, $(\Phi^X_{t,s})^*Y$ would be interpreted to mean $(\Phi^X_{t,s})^*f_t$ and $(\Phi^X_{t,s})^*Y_t$. Also, X_tf must mean $X_tf_t \in C^\infty(M)$ while $Xf \in C^\infty((a,b)\times M)$. In the following we consider complete time dependent fields so as to minimize worries about the domain of $\Phi^X_{t,s}$ and $(\Phi^X_{t,s})^*$.

Theorem 5.7 *Let X and Y be* complete *smooth time dependent vector fields and let $f : \mathbb{R}\times M \to \mathbb{R}$ be smooth time dependant function. We have the following formulas:*

$$\frac{d}{dt}(\Phi^X_{t,s})^*f = (\Phi^X_{t,s})^*(X_tf + \frac{\partial f}{\partial t})$$

and

$$\frac{d}{dt}(\Phi^X_{t,s})^*Y = (\Phi^X_{t,s})^*([X_t, Y_t] + \frac{\partial Y}{\partial t}).$$

Proof. Let f_t denote the function $f(t,.)$ as explained above. Consider the map $(u,v) \mapsto \Phi^X_{u,s}f_v$. In the following, we suppress X in expressions like $\Phi^X_{s,t}$ writing simply $\Phi_{s,t}$. If we let $u(t)=t, v(t)=t$ and compose, then by the chain rule

$$\begin{aligned}
\frac{d}{dt}\left(\Phi^*_{u,s}f_v\right)(p) &= \left.\frac{\partial}{\partial u}\right|_{(u,v)=(t,t)}\left(\Phi^*_{u,s}f_v\right)(p) + \left.\frac{\partial}{\partial v}\right|_{(u,v)=(t,t)}\left(\Phi^*_{u,s}f_v\right)(p)\\
&= \left.\frac{d}{du}\right|_{u=t}\left(\Phi^*_{u,s}f_t\right)(p) + \left.\frac{d}{dv}\right|_{v=t}\left(\Phi^*_{t,s}f_v\right)(p)\\
&= \left.\frac{d}{du}\right|_{u=t}\left(f_t\circ\Phi_{u,s}\right)(p) + \left(\Phi^*_{t,s}\frac{\partial f}{\partial t}\right)(p)\\
&= df_t\cdot\left.\frac{d}{du}\right|_{u=t}\Phi_{u,s}(p) + \left(\Phi^*_{t,s}\frac{\partial f}{\partial t}\right)(p)\\
&= df_t\cdot X_t(\Phi_{t,s}(p)) + \left(\Phi^*_{t,s}\frac{\partial f}{\partial t}\right)(p)\\
&= \left(X_t f\right)\left(\Phi_{t,s}(p)\right) + \left(\Phi^*_{t,s}\frac{\partial f}{\partial t}\right)(p)\\
&= \Phi^*_{t,s}\left(X_t f\right)(p) + \left(\Phi^*_{t,s}\frac{\partial f}{\partial t}\right)(p) = (\Phi_{t,s})^*(X_t f + \frac{\partial f}{\partial t})(p)
\end{aligned}$$

Note that a similar but simpler proof shows that if $f \in C^\infty(M)$ then

$$\frac{d}{dt}\Phi_{t,s}{}^* f = \Phi_{t,s}{}^*(X_t f) \tag{*}$$

Claim:

$$\frac{d}{dt}\Phi_{s,t}{}^* f = -X_t\{\Phi^*_{s,t}f\} \tag{**}$$

Proof of claim: Let $g := \Phi^*_{s,t}f$. Fix p and consider the map $(u,v) \mapsto \left(\Phi^*_{s,u}\Phi^*_{v,s}g\right)(p)$. If we let $u(t) = t, v(t) = t$ then the composed map $t \mapsto \left(\Phi_{s,t}{}^*\Phi^*_{t,s}g\right)(p) = p$ is constant. Thus by the chain rule

$$\begin{aligned}
0 &= \frac{d}{dt}\left(\Phi_{s,t}{}^*\Phi^*_{t,s}g\right)(p)\\
&= \left.\frac{\partial}{\partial u}\right|_{(u,v)=(t,t)}\left[\Phi_{s,u}{}^*\Phi^*_{v,s}g(p)\right] + \left.\frac{\partial}{\partial v}\right|_{(u,v)=(t,t)}\left[\left(\Phi_{s,u}{}^*\Phi^*_{v,s}g\right)(p)\right]\\
&= \left.\frac{d}{du}\right|_{u=t}\left[\left(\Phi^*_{s,u}\Phi^*_{t,s}g\right)(p)\right] + \left.\frac{d}{dv}\right|_{v=t}\left[\left(\Phi_{s,t}{}^*\Phi^*_{v,s}g\right)(p)\right]\\
&= \left.\frac{d}{du}\right|_{u=t}\left[\left(\Phi_{s,u}{}^*f\right)(p)\right] + \Phi^*_{s,t}\Phi^*_{t,s}X_t g \quad \text{(using (*))}\\
&= \frac{d}{dt}\left[\left(\Phi^*_{s,t}f\right)(p)\right] + X_t g\\
&= \frac{d}{dt}\left[\left(\Phi^*_{s,t}f\right)(p)\right] + X_t[\Phi^*_{s,t}f]
\end{aligned}$$

This proves the claim. Next, note that by Proposition **??** we have that since $\Phi^X_{s,t} = \left(\Phi^X_{t,s}\right)^{-1}$

$$\left(\Phi^*_{t,s}Y\right)f = \Phi^*_{t,s}\left(Y\left(\Phi^*_{s,t}f\right)\right) \tag{***}$$

for $Y \in \mathfrak{X}(M)$ and smooth f. This last equation still holds for $Y_t = Y(t, .)$ for a time dependent Y.

Next consider a time dependent vector field Y. We wish to compute $\frac{d}{dt}\left(\Phi^*_{t,s} Y_t\right)$ using the chain rule as before. We have

$$\begin{aligned}
\frac{d}{dt}\left(\Phi^*_{t,s} Y_t\right) f &= \left.\frac{d}{du}\right|_{u=t} \left(\Phi^*_{u,s} Y_t\right) f + \left.\frac{d}{dv}\right|_{v=t} \left(\Phi^*_{t,s} Y_v\right) f \\
&= \left.\frac{d}{du}\right|_{u=t} \Phi^*_{u,s}\left(Y_t \Phi^*_{s,u} f\right) + \left.\frac{d}{dv}\right|_{v=t} \left(\Phi^*_{t,s} Y_v\right) f \qquad \text{(using (***))} \\
&= \frac{d}{dt}\left(\Phi^X_{t,s}\right)^* \left(Y_t \Phi^*_{s,t} f\right) + \left(\Phi^*_{t,s} \frac{\partial Y}{\partial t}\right) f + \frac{d}{dt}\left(\Phi^X_{t,s}\right)^* \left(Y_t \Phi^*_{s,t} f\right) + \left(\Phi^*_{t,s} \frac{\partial Y}{\partial t}\right) f \\
&= \left.\frac{d}{du}\right|_{u=t} \Phi^*_{u,s}\left(Y_t \Phi^*_{s,t} f\right) + \left.\frac{d}{dv}\right|_{v=t} \Phi^*_{t,s}\left(Y \Phi^*_{s,v} f\right) + \left(\Phi^*_{t,s} \frac{\partial Y}{\partial t}\right) f \qquad \text{(using (**))} \\
&= \Phi^*_{t,s} X_t\left(Y_t \Phi^*_{s,t} f\right) - \Phi^*_{t,s} Y_t\left(X_t \Phi^*_{s,t} f\right) + \left(\Phi^*_{t,s} \frac{\partial Y}{\partial t}\right) f \\
&= \Phi^*_{t,s}\left([X_t, Y_t] \Phi^*_{s,t} f\right) + \Phi^*_{s,t} \frac{\partial Y}{\partial t} f \\
&= \left(\Phi^*_{t,s}\left[[X_t, Y_t] + \frac{\partial Y}{\partial t}\right]\right) f \quad \text{(using (***) again on } [X_t, Y_t])
\end{aligned}$$

■

Chapter 6

Geometry

The art of doing mathematics consists in finding that special case which contains all the germs of generality.

David Hilbert (1862-1943).

"What is Geometry?". Such is the title of a brief expository article ([**?**]) by Shiing-Shen Chern -one of the giants of 20th century geometry. In the article, Chern does not try to answer the question directly but rather identifies a few of the key ideas which, at one time or another in the history of geometry, seem to have been central to the very meaning of geometry. It is no surprise that the first thing Chern mentions is the axiomatic approach to geometry which was the method of Euclid. Euclid did geometry without the use of coordinates and proceeded in a largely logico-deductive manner supplemented by physical argument. This approach is roughly similar to geometry as taught in (U.S.) middle and high schools and so I will not go into the subject. Suffice it to say that the subject has a very different flavor from modern differential geometry, algebraic geometry and group theoretic geometry[1]. Also, even though it is commonly thought that Euclid though of geometry as a purely abstract discipline, it seems that for Euclid geometry was the study of a physical reality. The idea that geometry had to conform to physical space persisted far after Euclid and this prejudice made the discovery of non-Euclidean geometries a difficult process.

Next in Chern's list of important ideas is Descarte's idea of introducing coordinates into geometry. As we shall see, the use of (or at least existence of) coordinates is one of the central ideas of differential geometry. Descarte's coordinate method put the algebra of real numbers to work in service of geometry. A major effect of this was to make the subject easier (thank you, Descarte!). Coordinates paved the way for the calculus of several variables and then the modern theory of differentiable manifolds which will be a major topic of the sequel.

It was Felix Klein's program (Klein's Erlangen Programm) that made groups

[1] In fact, the author did not enjoy high school geometry at all.

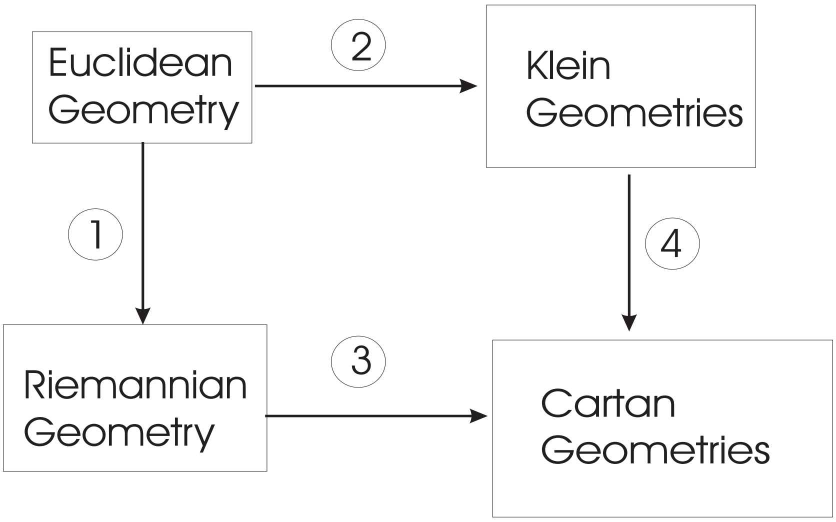
and the ensuing notions of symmetry and invariance central to the very meaning of geometry. This approach was advanced also by Cayley and Cartan. Most of this would fit into what is now referred to as the geometry of homogeneous spaces. Homogenous spaces are not always studied from a purely geometric point of view and, in fact, they make an appearance in many branches of mathematics and physics. Homogeneous spaces also take a central position in the field of Harmonic analysis[2].

These days when one thinks of differential geometry it is often Riemannian geometry that comes to mind. Riemann's approach to geometry differs from that of Klein's and at first it is hard to find common ground outside of the seemingly accidental fact that some homogeneous spaces carry a natural Riemannian structure. The basic object of Riemannian geometry is that of a Riemannian manifold. On a Riemannian manifold length and distance are defined using first the notion of lengths of tangent vectors and paths. A Riemannian manifold may well have a trivial symmetry group (isometry group). This would seem to put group theory in the back seat unless we happen to be studying highly symmetric spaces like "Riemannian homogeneous spaces" or "Riemannian symmetric spaces" such as the sphere or Euclidean space itself. Euclidean geometry is both a Klein geometry and a Riemannian geometry and so it is the basis of two different generalizations shown as (1) and (2) in figure 6.

The notion of a *connection* is an important unifying notion for modern differential geometry. A connection is a device to measure constancy and change and allows us to take derivatives of vector fields and more general fields. In Riemannian geometry, the central bit of extra structure is that of a metric tensor which allows us to measure lengths, areas, volumes and so on. In particular, every Riemannian manifold has a distance function and so it a metric space ("metric" in the sense of general topology). In Riemannian geometry the connection comes from the metric tensor and is called the Levi-Civita connection. Thus distance and length are at the root of how change is reckoned in Riemannian geometry. In anticipation of material we present later, let it be mentioned that, locally, a connection on an $n-$dimensional Riemannian manifold is described by an $\mathfrak{so}(n)-$valued differential $1-$form (differential forms are studied in chapter **??**).

We use the term Klein geometry instead of homogeneous space geometry when we take on the specifically geometric attitude characterized in part by emphasis on the so called Cartan connection which comes from the Maurer-Cartan form on the group itself. It is the generalization of this Lie algebra valued 1-form which gives rise to Cartan Geometry (this is the generalization (4) in the figure). Now Cartan geometry can also be seen as a generalization inspired directly from Riemannian geometry. Riemannian geometry is an example of a Cartan geometry for sure but the way it becomes such is only fully understood from the point of view of the generalization labelled (4) in the figure. From this standpoint the relevant connection is the Cartan version of the Levi-Civita

[2] Lie group representations are certainly important for geometry.



connection which takes values in the Lie algebra of Euc(n) rather than the Lie algebra of $S\mathrm{O}(n)$. This approach is still unfamiliar to many professional differential geometers but as well shall see it is superior in many ways. For a deep understanding of the Cartan viewpoint on differential geometry the reader should look at R.W. Sharp's excellent text [**?**]. The presentation of Cartan geometries found in this book it heavily indebted to Sharp's book.

Klein geometry will be pursued at various point throughout the book but especially in chapters **??**, **??**, and **??**. Our goal in the present section will be to introduce a few of the most important groups and related homogeneous spaces. The main example for any given dimension n, is the $n-$dimensional Euclidean space (together with the group of Euclidean motions).

Most of the basic examples involve groups acting on spaces which topologically equivalent to finite dimensional vector spaces. In the calculus of several variables we think of $\mathbb{R}^3$ as being a model for 3-dimensional physical space. Of course, everyone realizes that $\mathbb{R}^3$ is not quite a faithful model of space for several reasons not the least of which is that, unlike physical space, $\mathbb{R}^3$ is a vector space and so, for instance, has a unique special element (point) called the zero element. Physical space doesn't seem to have such a special point (an origin). A formal mathematical notion more appropriate to the situation which removes this idea of an origin is that of an **affine space** defined below 6.1. As actors in physical space, we implicitly and sometimes explicitly impose coordinate systems onto physical space. Rectangular coordinates are the most familiar type of coordinates that physicists and engineers use. In the physical sciences, coordinates are really implicit in our measurement methods and in the instruments used.

We also *impose* coordinates onto certain objects or onto their surfaces. Particularly simple are flat surfaces like tables and so forth which intuitively are 2-dimensional analogues of the space in which we live. Experience with such objects and with 3-dimensional space itself leads to the idealizations known as the Euclidean plane, 3-dimensional Euclidean space. Our intuition about the Euclidean plane and Euclidean $3-$space is extended to higher dimensions by analogy. Like 3-dimensional space, the Euclidean plane also has no preferred coordinates or implied vector space structure. Euclid's approach to geometry used neither (at least no explicit use). On the other hand, there does seem to be a special family of coordinates on the plane that makes the equations of geometry take on a simple form. These are the rectangular (orthonormal) coordinates mentioned above. In rectangular coordinates the distance between two points is given by the usual Pythagorean prescription involving the sum of squares. What set theoretic model should a mathematician exhibit as the best mathematical model of these Euclidean spaces of intuition? Well, Euclidean space may not be a vector space as such but since we have the notion of translation in space we do have a vector space "acting by translation". Paying attention to certain features of Euclidean space leads to the following notion:

Definition 6.1 *Let* $\mathbf{A}$ *be a set and* V *be a vector space over a field* $\mathbb{F}$. *We say that* $\mathbf{A}$ *is an* ***affine space*** *with* ***difference space*** V *if there is a map* $+ : V \times \mathbf{A} \to \mathbf{A}$ *written* $(v, p) \mapsto v + p$ *such that*
i) $(v + w) + p = v + (w + p)$ *for all* $v, w \in V$ *and* $p \in \mathbf{A}$
ii) $0 + p = p$ *for all* $p \in \mathbf{A}$.
iii) for each fixed $p \in \mathbf{A}$ *the map* $v \mapsto v + p$ *is a bijection.*

If we have an affine space with difference space V then there is a unique map called the **difference map** $- : \mathbf{A} \times \mathbf{A} \to V$ which is defined by

$$q - p = \text{the unique } v \in V \text{ such that } v + p = q.$$

Thus we can form difference of two points in an affine space but we cannot add two points in an affine space or at least if we can then that is an accidental feature and is not part of what it means to be an affine space. We need to know that such things really exist in some sense. What should we point at as an example of an affine space? In a ironic twist the set which is most ready-to-hand for this purpose is $\mathbb{R}^n$ itself. We just allow $\mathbb{R}^n$ to play the role of both the affine space and the difference space. We take advantage of the existence of a predefine vector addition $+$ to provide the translation map. Now $\mathbb{R}^n$ has many features that are not an essential part of its affine space character. When we let $\mathbb{R}^n$ play this role as an affine space, we must be able to turn a sort of blind eye to features which are not essential to the affine space structure such as the underlying vector spaces structure. Actually, it would be a bit more honest to admit that we will in fact use features of $\mathbb{R}^n$ which are accidental to the affine space structure. After all, the first thing we did was to use the vector space addition from $\mathbb{R}^n$ to provide the translation map. This is fine- we only need to keep track of what aspects of our constructions and which of our calculations retain significant for the properly affine aspects of $\mathbb{R}^n$. The introduction of group theory into the picture helps a great deal in sorting things out. More generally any vector space can be turned into an affine space simply by letting the translation operator + required by the definition of an affine space to be the addition that is part of the vector space structure of V. This is just the same as what we did with $\mathbb{R}^n$ and again V itself is its own difference space and the set V is playing two roles. Similar statements apply it the field is $\mathbb{C}$. In fact, algebraic geometers refer to $\mathbb{C}^n$ as complex affine space.

Let $\mathbb{F}$ be one of the fields $\mathbb{R}$ or $\mathbb{C}$. Most of the time $\mathbb{R}$ is the field intended when no specification is made. It is a fairly common practice to introduce multiple notations for the same set. For example, when we think of the vector space $\mathbb{F}^n$ as an affine space we sometimes denote it by $\mathbf{A}^n(\mathbb{F})$ or just $\mathbf{A}^n$ if the underlying field is understood.

Definition 6.2 *If $\mathbf{A}_i$ is an affine space with difference space V_i for $i = 1, 2$ then we say that a map $F : \mathbf{A}_1 \to \mathbf{A}_2$ is an affine transformation if it has the form $x \mapsto F(x_0) + L(x - x_0)$ for some linear transformation $L : V_1 \to V_2$.*

It is easy to see that for a fixed point p in an affine space $\mathbf{A}$ we immediately obtain a bijection $V \to \mathbf{A}$ which is given by $v \mapsto p + v$. The inverse of this bijection is a map $c_p : \mathbf{A} \to V$ which we put to use presently. An affine space always has a globally defined coordinate system. To see this pick a basis for V. This gives a bijection $V \to \mathbb{R}^n$. Now choose a point $p \in \mathbf{A}$ and compose with the canonical bijection $c_p : \mathbf{A} \to V$ just mentioned. This a coordinates system. Any two such coordinate systems are related by an bijective affine transformation (an affine automorphism). These special coordinate systems can be seen as an alternate way of characterizing the affine structure on the set $\mathbf{A}$. In fact, singling out a family of specially related coordinate systems is an often used method saying what we mean when we say that a set has a structure of a given type and this idea will be encounter repeatedly.

If V is a finite dimensional vector space then it has a distinguished topology[3] which is transferred to the affine space using the canonical bijections. Thus we may consider open sets in A and also continuous functions on open sets. Now it is usually convenient to consider the "coordinate representative" of a function rather than the function itself. By this we mean that if $x : \mathbf{A} \to \mathbb{R}^n$ is a affine coordinate map as above, then we may replace a function $f : U \subset \mathbf{A} \to \mathbb{R}$ (or $\mathbb{C}$) by the composition $f \circ x$. The latter is often denoted by $f(x^1, ..., x^n)$. Now this coordinate representative may, or may not, turn out to be a differentiable function but if it is then all other coordinate representatives of f obtained by using other affine coordinate systems will also be differentiable. For this reason, we may say that f itself is (or is not) differentiable. We say that the family of affine coordinate systems provide the affine space with a differentiable structure and this means that every (finite dimensional) affine space is also an example of differentiable manifold (defined later). In fact, one may take the stance that an affine space is the local model for all differentiable manifolds.

[3] If V is infinite dimensional we shall usually assume that V is a topological vector space.

Another structure that will seem a bit trivial in the case of an affine space but that generalizes in a nontrivial way is the idea of a tangent space. We know what it means for a vector to be tangent to a surface at a point on the surface but the abstract idea of a tangent space is also exemplified in the setting of affine spaces. If we have a curve $\gamma : (a, b) \subset \mathbb{R} \to \mathbf{A}$ then the affine space structure allows us to make sense of the difference quotient

$$\dot{\gamma}(t_0) := \lim_{h \to 0} \frac{\gamma(t_0 + h) - \gamma(t_0)}{h} \ , \ t_0 \in (a, b)$$

which defines an element of the difference space V (assuming the limit exists). Intuitively we think of this as the velocity of the curve *based at* $\gamma(t_0)$. We may want to explicitly indicate the base point. If $p \in \mathbf{A}$, then the **tangent space** at p is $\{p\} \times V$ and is often denoted by $T_p\mathbf{A}$. The set of all tangent spaces for points in an open set $U \subset \mathbf{A}$ is called the tangent bundle over U and is another concept that will generalize in a very interesting way. Thus we have not yet arrived at the usual (metric) Euclidean space or Euclidean geometry where distances and angle are among the prominent notions. On the other hand, an affine space supports a geometry called *affine geometry.* The reader who has encountered axiomatic affine geometry will remember that one is mainly concerned about lines and the their mutual points of incidence. In the current analytic approach a line is the set of image points of an affine map $a : \mathbb{R} \to \mathbf{A}$. We will not take up affine geometry proper. Rather we shall be mainly concerned with geometries that result from imposing more structure on the affine space. For example, an affine space is promoted to a metric Euclidean space once we introduce length and angle. Once we introduce length and angle into $\mathbb{R}^n$ it becomes the standard model of n-dimensional Euclidean space and might choose to employ the notation $\mathbf{E}^n$ instead of $\mathbb{R}^n$. The way in which length and angle is introduced into $\mathbb{R}^n$ (or an abstract finite dimensional vector space V) is via an inner product and this is most likely familiar ground for the reader. Nevertheless, we will take closer look at what is involved so that we might place Euclidean geometry into a more general context. In order to elucidate both affine geometry and Euclidean geometry and to explain the sense in which the word geometry is being used, it is necessary to introduce the notion of "group action".

6.1 Group Actions, Symmetry and Invariance

One way to bring out the difference between $\mathbb{R}^n$, $\mathbf{A}^n$ and $\mathbf{E}^n$, which are all the same considered as sets, is by noticing that each has its own natural set of coordinates and coordinate transformations. The natural family of coordinate systems for $\mathbf{A}^n$ are the affine coordinates and are related amongst themselves by affine transformations under which lines go to lines (planes to planes etc.). Declaring that the standard coordinates obtained by recalling that $\mathbf{A}^n = \mathbb{R}^n$ determines all other affine coordinate systems. For $\mathbf{E}^n$ the natural coordinates are the orthonormal rectangular coordinates which are related to each other by affine transformations whose linear parts are orthogonal transformations of the difference space. These are exactly the length preserving transformations or *isometries*. We will make this more explicit below but first let us make some comments about the role of symmetry in geometry.

We assume that the reader is familiar with the notion of a group. Most of the groups that play a role in differential geometry are matrix groups and are the prime examples of so called 'Lie groups' which we study in detail later in the book. For now we shall just introduce the notion of a topological group. All Lie groups are topological groups. Unless otherwise indicated finite groups G will be given the discreet topology where every singleton $\{g\} \subset G$ is both open and closed.

Definition 6.3 *Let G be a group. We say that G is a topological group if G is also a topological space and if the maps $\mu : G \times G \to G$ and $\mathrm{inv} : G \to G$, given by $(g_1, g_2) \to g_1 g_2$ and $g_1 \mapsto g^{-1}$ respectively, are continuous maps.*

If G is a countable or finite set we usually endow G with the discrete topology so that in particular, every point would be an open set. In this case we call G a discrete group.

Even more important for our present discussion is the notion of a group action. Recall that if M is a topological space then so is $G \times M$ with the product topology.

Definition 6.4 *Let G and M be as above. A left (resp. right)* ***group action*** *is a map $\alpha : G \times M \to M$ (resp. $\alpha : M \times G \to M$) such that for every $g \in G$ the* ***partial map*** *$\alpha_g(.) := \alpha(g, .)$ (resp. $\alpha_{,g}(.) := \alpha(., g)$) is continuous and such that the following hold:*

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

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

It is traditional to write $g \cdot x$ or just gx in place of the more pedantic notation $\alpha(g, x)$. Using this notation we have $g_2 \cdot (g_1 \cdot x) = (g_2 g_1) \cdot x$ and $e \cdot x = x$.

We shall restrict our exposition for left actions only since the corresponding notions for a right action are easy to deduce. Furthermore, if we have a right action $x \to \alpha^R(x, g)$ then we can construct an essentially equivalent left action by $\alpha^L : x \to \alpha^L(g, x) := \alpha^R(x, g^{-1})$.

If we have an 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. The set of orbits is denoted by M/G or sometimes $G|M$ if we wish to emphasize that the action is a left action. It is easy to see that two orbits $G \cdot x$ and $G \cdot y$ are either disjoint or identical and so define an equivalence relation. The natural projection onto set of orbits $p : M \to M/G$ is given by

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

If we give M/G the quotient topology then of course p is continuous.

Definition 6.5 *If G acts on M the we call M a G-**space**.*

Exercise 6.6 *Convince yourself that an affine space A with difference space V is a $V-$space where we consider V as an abelian group under addition.*

Definition 6.7 *Let G act on M_1 and M_2. A map $f : M_1 \to M_2$ is called a $G-$map if*

$$f(g \cdot x) = g \cdot f(x)$$

*for all $x \in M$ and all $g \in G$. If f is a bijection then we say that f is a G-automorphism and that M_1 and M_2 are isomorphic as G-spaces. More generally, if G_1 acts on M_1 and G_2 acts on M_2 then a **weak group action morphism** from M_1 to M_2 is a pair of maps (f, ϕ) such that*

(i) $f : M_1 \to M_2$,
(ii) $\phi : G_1 \to G_2$ is a group homomorphism and
*(iii) $f(g \cdot x) = \phi(g) \cdot f(x)$ for all $x \in M$ and all $g \in G$. If f is a bijection and ϕ is an isomorphism then the group actions on M_1 and M_2 are **equivalent**.*

We have refrained from referring to the equivalence in (iii) as a "weak" equivalence since if M_1 and M_2 are G_1 and G_2 spaces respectively which are equivalent in the sense of (iii) then $G_1 \cong G_2$ by definition. Thus if we then identify G_1 and G_2 and call the resulting abstract group G then we recover a $G-$space equivalence between M_1 and M_2.

One main class of examples are those where M is a vector space (say V) and each α_g is a linear isomorphism. In this case, the notion of a left group action is identical to what is called a **group representation** and for each $g \in G$, the map $g \mapsto \alpha_g$ is a group homomorphism[4] with image in the space $G\mathrm{L}(V)$ of linear isomorphism of G to itself. The most common groups that occur in geometry are matrix groups, or rather, groups of automorphisms of a fixed vector space. Thus we are first to consider subgroups of $G\mathrm{L}(V)$. A typical way to single out subgroups of $G\mathrm{L}(V)$ is provided by the introduction of extra structure onto V such as an orientation and/or a special bilinear form such as an real inner product, Hermitian inner product, or a symplectic from to name just a few. We shall introduce the needed structures and the corresponding groups as needed. As a start we ask the reader to recall that an automorphism of a finite dimensional vector space V is *orientation preserving* if its matrix representation with respect to some basis has positive determinant. (A more geometrically pleasing definition of orientation and of determinant etc. will be introduced later). The set of all orientation preserving automorphisms of V is denoted a subgroup of $G\mathrm{L}(V)$ and is denoted $Gl^+(V)$ and referred to as the **proper general linear group**. We are also interested in the group of automorphisms that have determinant equal to 1 which gives us a "*special* linear group". (These will be seen to be volume preserving maps). Of course, once we pick a basis we get an identification of any such linear automorphism

[4] A right group action does not give rise to a homomorphism but an "anti-homomorphism".

group with a group of matrices. For example, we identify $G\mathrm{L}(\mathbb{R}^n)$ with the group $G\mathrm{L}(n,\mathbb{R})$ of nonsingular $n \times n$ matrices. Also, $Gl^+(\mathbb{R}^n) \cong Gl^+(n,\mathbb{R}) = \{A \in G\mathrm{L}(n,\mathbb{R}) : \det A = 1\}$. A fairly clear pattern of notation is emerging and we shall not be so explicit about the meaning of the notation in the sequel.

(oriented) Vector Space	General linear	Proper general linear	Special linear
General V	$G\mathrm{L}(V)$	$Gl^+(V)$	$\mathrm{SL}(V)$
$\mathbb{F}^n$	$G\mathrm{L}(\mathbb{F}^n)$	$Gl^+(\mathbb{F}^n)$	$\mathrm{SL}(\mathbb{F}^n)$
Matrix group	$G\mathrm{L}(n,\mathbb{F})$	$Gl^+(n,\mathbb{F})$	$\mathrm{SL}(n,\mathbb{F})$
Banach space E	$G\mathrm{L}(\mathsf{E})$	?	?

If one has a group action then one has a some sort of geometry. From this vantage point, the geometry is whatever is 'preserved' by the group action.

Definition 6.8 *Let $\mathcal{F}(M)$ be the vector space of all complex valued functions on M. An action of G on M produces an action of G on $\mathcal{F}(M)$ as follows*

$$(g \cdot f)(x) := f(g^{-1}x)$$

This action is called the ***induced action****.*

This induced action preserves the vector space structure on $\mathcal{F}(M)$ and so we have produced an example of a group representation. If there is a function $f \in \mathcal{F}(M)$ such that $f(x) = f(g^{-1}x)$ for all $g \in G$ then f is called an **invariant**. In other words, f is an invariant if it remains fixed under the induced action on $\mathcal{F}(M)$. The existence of an invariant often signals the presence of an underlying geometric notion.

Example 6.9 *Let $G = \mathrm{O}(n,\mathbb{R})$ be the* ***orthogonal group*** *(all invertible matrices Q such that $Q^{-1} = Q^t$), let $M = \mathbb{R}^n$ and let the action be given by matrix multiplication on column vectors in $\mathbb{R}^n$;*

$$(Q, v) \mapsto Qv$$

Since the length of a vector v as defined by $\sqrt{v \cdot v}$ is preserved by the action of $\mathrm{O}(\mathbb{R}^n)$ we pick out this notion of length as a geometric notion. The ***special orthogonal group*** *$S\mathrm{O}(n,\mathbb{R})$ (or just $S\mathrm{O}(n)$) is the subgroup of $\mathrm{O}(n,\mathbb{R})$ consisting of elements with determinant 1. This is also called the rotation group (especially when $n = 3$).*

More abstractly, we have the following

Definition 6.10 *If V be a vector space (over a field $\mathbb{F}$) which is endowed with a distinguished nondegenerate symmetric bilinear form b (a real inner product if $\mathbb{F} = \mathbb{R}$), then we say that V has an orthogonal structure. The set of linear transformations $L : V \to V$ such that $b(Lv, Lw) = b(v, w)$ for all $v, w \in V$ is a group denoted $\mathrm{O}(V, \langle,\rangle)$ or simply $\mathrm{O}(V)$. The elements of $\mathrm{O}(V)$ are called* ***orthogonal transformations.***

A map A between complex vector spaces which is linear over $\mathbb{R}$ and satisfies $A(\bar{v}) = \overline{A(v)}$ is conjugate linear (or antilinear). Recall that for a complex vector space V, a map $h : V \times V \to \mathbb{C}$ that is linear in one variable and conjugate linear in the other is called a **sesquilinear form**. If, further, h is nondegenerate then it is called a **Hermitian form** (or simply a complex inner product).

Definition 6.11 *Let V be a complex vector space which is endowed distinguished nondegenerate Hermitian h, then we say that V has a **unitary structure**. The set of linear transformations $L : V \to V$ such that $h(Lv, Lw) = h(v, w)$ for all $v, w \in V$ is a group denoted $U(V)$. The elements of* $\mathrm{O}(V)$ *are called **unitary transformations.** The standard Hermitian form on $\mathbb{C}^n$ is $(x, y) \mapsto \sum \bar{x}^i y^i$ (or depending on taste $\sum x^i \bar{y}^i$). We have the obvious identification $U(\mathbb{C}^n) = U(n, \mathbb{C})$.*

It is common to assume that $\mathrm{O}(n)$ refers to $\mathrm{O}(n, \mathbb{R})$ while $U(n)$ refers to $U(n, \mathbb{C})$. It is important to notice that with the above definitions $U(\mathbb{C}^n)$ is not the same as $\mathrm{O}(\mathbb{C}^n, \sum x^i y^i)$ since $b(x, y) = \sum x^i y^i$is bilinear rather than sesquilinear.

Example 6.12 *Let G be the **special linear group*** $\mathrm{SL}(n, \mathbb{R}) = \{A \in \mathrm{GL}(n, \mathbb{R}) : \det A = 1\}$. *The length of vectors are not preserved but something else is preserved. If $v^1, ..., v^n \in \mathbb{R}^n$ then the determinant function* $\det(v^1, ..., v^n)$ *is preserved:*

$$\det(v^1, ..., v^n) = \det(Qv^1, ..., Qv^n)$$

Since det *is not really a function on $\mathbb{R}^n$ this invariant doesn't quite fit the definition of an invariant above. On the other hand,* det *is a function of several variables, i.e. a function on $\mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ and so there is the obvious induced action on this space of function. Clearly we shall have to be flexible if we are to capture all the interesting invariance phenomenon. Notice that the* $S\mathrm{O}(n, \mathbb{R}) \subset \mathrm{SL}(n, \mathbb{R})$ *and so the action of* $S\mathrm{O}(n, \mathbb{R})$ *on $\mathbb{R}^n$ is also orientation preserving.*

Orthogonal Structure	Full orthogonal	Special Orthogonal
V + nondegen. sym. form b	$\mathrm{O}(V, b)$	$S\mathrm{O}(V, b)$
$\mathbb{F}^n, \sum x^i y^i$	$\mathrm{O}(\mathbb{F}^n)$	$S\mathrm{O}(\mathbb{F}^n)$
Matrix group	$\mathrm{O}(n, \mathbb{F})$	$S\mathrm{O}(n, \mathbb{F})$
Hilbert space E	$\mathrm{O}(\mathsf{E})$	?

6.2 Some Klein Geometries

Definition 6.13 *A group action is said to be **transitive** if there is only one orbit. A group action is said to be **effective** if $g \cdot x = x$ for all $x \in M$ implies that $g = e$ (the identity element). A group action is said to be **free** if $g \cdot x = x$ for some x implies that $g = e$.*

Klein's view is that a geometry is just a transitive G-space. Whatever properties of figures (or more general objects) that remain unchanged under the action of G are deemed to be geometric properties by definition.

Example 6.14 *We give an example of an* $\mathrm{SL}(n, \mathbb{R})$*-space as follows: Let* M *be the upper half complex plane* $\{z \in \mathbb{C} : \operatorname{Im} z > 0\}$*. Then the action of* $\mathrm{SL}(n, \mathbb{R})$ *on* M *is given by*

$$(A, z) \mapsto \frac{az + b}{cz + d}$$

where $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Exercise 6.15 *Let* $A \in \mathrm{SL}(n, \mathbb{R})$ *as above and* $w = A \cdot z = \frac{az+b}{cz+d}$*. Show that if* $\operatorname{Im} z > 0$ *then* $\operatorname{Im} w > 0$ *so that the action of the last example is well defined.*

6.2.1 Affine Space

Now let us consider the notion of an affine space again. Careful attention to the definitions reveals that an affine space is a set on which a vector space acts (a vector space is also an abelian group under the vector addition $+$). We can capture more of what an affine space $\mathbf{A}^n$ is geometrically by enlarging this action to include the full group of nonsingular affine maps of $\mathbf{A}^n$ onto itself. This group is denoted $\mathrm{Aff}(\mathbf{A}^n)$ and is the set of transformations of $\mathbf{A}^n$ of the form $A : x \mapsto Lx_0 + L(x - x_0)$ for some $L \in \mathrm{GL}(n, \mathbb{R})$ and some $x_0 \in \mathbf{A}^n$. More abstractly, if $\mathbf{A}$ is an affine space with difference space V then we have the group of all transformations of the form $A : x \mapsto Lx_0 + L(x - x_0)$ for some $x_0 \in A$ and $L \in \mathrm{GL}(V)$. This is the **affine group** associated with $\mathbf{A}$ and is denoted by $\mathrm{Aff}(\mathbf{A})$. Using our previous terminology, A is an $\mathrm{Aff}(\mathbf{A})-$space. $\mathbf{A}^n$ $(= \mathbb{R}^n)$ is our standard model of an n-dimensional affine space. It is an $\mathrm{Aff}(\mathbf{A}^n)$-space.

Exercise 6.16 *Show that if* $\mathbf{A}$ *an is a* n*-dimensional affine space with difference space* V *then* Aff*(*$\mathbf{A}$*)*$\cong$ $\mathrm{Aff}(\mathbf{A}^n)$ *and the* Aff*(*$\mathbf{A}$*)-space* $\mathbf{A}$ *is equivalent to the* $\mathrm{Aff}(\mathbf{A}^n)-$*space* $\mathbf{A}^n$*.*

Because of the result of this last exercise it is sufficient mathematically to restrict the study of affine space to the concrete case of $\mathbf{A}^n$ the reason that we refer to $\mathbf{A}^n$ as the standard affine space. We can make things even nicer by introducing a little trick which allows us to present $Aff(\mathbf{A}^n)$ as a matrix group. Let $\mathbf{A}^n$ be identified with the set of column vectors in $\mathbb{R}^{n+1}$ of the form

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

The set of all matrices of the form

$$\begin{bmatrix} 1 & 0 \\ x_0 & L \end{bmatrix} \text{ where } Q \in \mathrm{GL}(n, \mathbb{R}) \text{ and } x_0 \in \mathbb{R}^n$$

is a group. Now

$$\begin{bmatrix} 1 & 0 \\ x_0 & L \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ Lx + x_0 \end{bmatrix}$$

Remember that affine transformations of $\mathbf{A}^n$ $(= \mathbb{R}^n$ as a set) are of the form $x \mapsto Lx + x_0$. In summary, when we identify $\mathbf{A}^n$ with the vectors of the form $\begin{bmatrix} 1 \\ x \end{bmatrix}$ then the group $\mathrm{Aff}(\mathbf{A}^n)$ is the set of $(n+1) \times (n+1)$ matrices of the form indicated and the action of $\mathrm{Aff}(\mathbf{A}^n)$ is given by matrix multiplication.

It is often more appropriate to consider the **proper affine group** $Aff^+(\mathbf{A}, V)$ obtained by adding in the condition that its elements have the form $A : x \mapsto Lx_0 + L(x - x_0)$ with $\det L > 0$.

We have now arrive at another meaning of affine space. Namely, the system $(\mathbf{A}, V, Aff^+(\mathbf{A}), \cdot, +)$. What is new is that we are taking the action "$\cdot$" of $Aff^+(\mathbf{A})$ on as part of the structure so that now A is an $Aff^+(\mathbf{A})-$space.

Exercise 6.17 *Show that the action of V on $\mathbf{A}$ is transitive and free. Show that the action of $Aff^+(\mathbf{A})$ on $\mathbf{A}$ is transitive and effective but not free.*

Affine geometry is the study of properties attached to figures in an affine space that remain, in some appropriate sense, unchanged under the action of the affine group (or the proper affine group Aff^+). For example, coincidence properties of lines and points are preserved by Aff.

6.2.2 Special Affine Geometry

We can introduce a bit more rigidity into affine space by changing the linear part of the group to $\mathrm{SL}(n)$.

Definition 6.18 *The group $SAff(\mathbf{A})$ of affine transformations of an affine space $\mathbf{A}$ that are of the form*

$$A : x \mapsto Lx_0 + L(x - x_0)$$

with $\det L = 1$ will be called the special affine group.

We will restrict ourselves to the standard model of $\mathbf{A}^n$. With this new group the homogeneous space structure is now different and the geometry has changed. For one thing volume now makes sense. The are many similarities between special affine geometry and Euclidean geometry. As we shall see in chapter **??**, in dimension 3, there exist special affine versions of arc length and surface area and mean curvature.

The volume of a parallelepiped is preserved under this action.

6.2.3 Euclidean space

Suppose now that we have an affine space $(A, V, +)$ together with an inner product on the vector space V. Consider the group of affine transformations the form $A : x \mapsto Lx_0 + L(x - x_0)$ where now $L \in \mathrm{O}(V)$. This is called the Euclidean motion group of A. Then $\mathbf{A}$ becomes, in this context, a Euclidean space and we might choose a different symbol, say $\mathbf{E}$ for the space to encode this fact. For example, when $\mathbb{R}^n$ is replaced by $\mathbf{E}^n$ and the Euclidean motion group is denoted $\mathrm{Euc}(\mathbf{E}^n)$.

Remark: Now every tangent space $T_pA = \{p\} \times V$ inherits the inner product from V in the obvious and trivial way. This is our first example of a metric tensor. Having a metric tensor on the more general spaces that we study later on will all us to make sense of lengths of curves, angles between velocity curves, volumes of subsets and much more.

In the concrete case of $\mathbf{E}^n$ (secretly $\mathbb{R}^n$again) we may represent the Euclidean motion group $\mathrm{Euc}(\mathbf{E}^n)$ as a matrix group, denoted $\mathrm{Euc}(n)$, by using the trick we used before. If we want to represent the transformation $x \to Qx + b$ where $x, b \in \mathbb{R}^n$ and $Q \in \mathrm{O}(n, \mathbb{R})$, we can achieve this by letting column vectors

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

represent elements x. Then we take the group of matrices of the form

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

to play the role of $\mathrm{Euc}(n)$. This works because

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

The set of all coordinate systems related to the standard coordinates by an element of $\mathrm{Euc}(n)$ will be called **Euclidean coordinates** (or sometimes orthonormal coordinates). It is these coordinate which we should use to calculate distance.

Basic fact 1: The (square of the) distance between two points in $\mathbf{E}^n$ can be calculated in any of the equivalent coordinate systems by the usual formula. This means that if $P, Q \in \mathbf{E}^n$ and we have some Euclidean coordinates for these two points: $x(P) = (x^1(P), ..., x^n(P))$, $x(Q) = (x^1(Q), ..., x^n(Q))$ then vector difference in these coordinates is $\Delta x = x(P) - x(Q) = (\Delta x^1, ..., \Delta x^n)$. If $y = (y^1, ..., y^n)$ are any other coordinates related to x by $y = Tx$ for some $T \in \mathrm{Euc}(n)$ then we have $\sum (\Delta y^i)^2 = \sum (\Delta x^i)^2$ where $\Delta y = y(P) - y(Q)$ etc. The distance between two points in a Euclidean space is a simple example of a geometric invariant.

Let us now pause to consider again the abstract setting of $G-$spaces. Is there geometry here? We will not attempt to give an ultimate definition of geometry associated to a $G-$space since we don't want to restrict the direction of future generalizations. Instead we offer the following two (somewhat tentative) definitions:

Definition 6.19 *Let M be a $G-$space. A figure in M is a subset of M. Two figures S_1, S_2 in M are said to be* ***geometrically equivalent*** *or* ***congruent*** *if there exists an element $g \in G$ such that $g \cdot S_1 = S_2$.*

Definition 6.20 *Let M be a $G-$space and let $\mathcal{S}$ be some family of figures in M such that $g \cdot S \in \mathcal{S}$ whenever $.S \in \mathcal{S}$. A function $I : \mathcal{S} \to \mathbb{C}$ is called a (complex valued) geometric invariant if $I(g \cdot S) = I(S)$ for all $S \in \mathcal{S}$.*

Example 6.21 *Consider the action of $\mathrm{O}(n)$ on $\mathbf{E}^n$. Let $\mathcal{S}$ be the family of all subsets of $\mathbf{E}^n$ which are the images of C^1 of the form $c : [a, b] \to \mathbf{E}^n$ such that c' is never zero (regular curves). Let $S \in \mathcal{S}$ be such a curve. Taking some regular C^1map c such that S is the image of c we define the length*

$$L(S) = \int_a^b \|c'(t)\| \, dt$$

It is common knowledge that $L(S)$ is independent of the parameterizing map c and so L is a geometric invariant (this is a prototypical example)

Notice that two curves may have the same length without being congruent. The invariant of length by itself does not form a complete set of invariants for the family of regular curves $\mathcal{S}$. The definition we have in mind here is the following:

Definition 6.22 *Let M be a $G-$space and let $\mathcal{S}$ be some family of figures in M. Let $\mathcal{I} = \{I_\alpha\}_{\alpha \in A}$ be a set of geometric invariants for $\mathcal{S}$. We say that $\mathcal{I}$ is a complete set of invariants if for every $S_1, S_2 \in \mathcal{S}$ we have that S_1 is congruent to S_2 if and only if $I_\alpha(S_1) = I_\alpha(S_2)$ for all $\alpha \in A$.*

Example 6.23 *Model Euclidean space by $\mathbb{R}^3$ and consider the family $\mathcal{C}$ of all regular curves $c : [0, L] \to \mathbb{R}^3$ such that $\frac{dc}{dt}$ and $\frac{d^2c}{dt^2}$ never vanish. Each such curve has a parameterization by arc length which we may take advantage of for the purpose of defining a complete set of invariant for such curves. Now let $c : [0, L] \to \mathbb{R}^3$ a regular curve parameterized by arc length. Let*

$$\mathbf{T}(s) := \frac{\frac{dc}{dt}(s)}{\left|\frac{dc}{dt}(s)\right|}$$

define the unit tangent vector field along c. The curvature is an invariant defined on c that we may think of as a function of the arc length parameter. It is defined by $\kappa(s) := \left|\frac{dc}{dt}(s)\right|$. We define two more fields along c. First the normal field is defined by $\mathbf{N}(s) = \frac{dc}{dt}(s)$. Next, define the unit binormal vector field $\mathbf{B}$ by requiring that $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is a positively oriented triple of orthonormal unit vectors. By positively oriented we mean that

$$\det[\mathbf{T}, \mathbf{N}, \mathbf{B}] = 1.$$

We now show that $\frac{d\mathbf{N}}{ds}$ is parallel to $\mathbf{B}$. For this it suffices to show that $\frac{d\mathbf{N}}{ds}$ is normal to both $\mathbf{T}$ and $\mathbf{N}$. First, we have $\mathbf{N}(s) \cdot \mathbf{T}(s) = 0$. If this equation *is differentiated we obtain* $2\frac{d\mathbf{N}}{ds}\cdot\mathbf{T}(s)=0$*. On the other hand we also have* $1=\mathbf{N}(s)\cdot\mathbf{N}(s)$ *which differentiates to give* $2\frac{d\mathbf{N}}{ds}\cdot\mathbf{N}(s)=0$*. From this we see that there must be a function* $\tau=\tau(s)$ *such that* $\frac{d\mathbf{N}}{ds}:=\tau\mathbf{B}$*. This is a function of arc length but should really be thought of as a function on the curve. This invariant is called the* ***torsion****. Now we have a matrix defined so that*

$$\frac{d}{ds}[\mathbf{T},\mathbf{N},\mathbf{B}]=[\mathbf{T},\mathbf{N},\mathbf{B}]\begin{bmatrix}0&\kappa&0\\-\kappa&0&\tau\\0&\tau&0\end{bmatrix}$$

or in other words

$$\begin{array}{rcccc}\frac{d\mathbf{T}}{ds}&=&&\kappa\mathbf{N}&\\ \frac{d\mathbf{N}}{ds}&=&-\kappa\mathbf{T}&&\tau\mathbf{B}\\ \frac{d\mathbf{B}}{ds}&=&&\tau\mathbf{N}&\end{array}$$

Since $F=[\mathbf{T},\mathbf{N},\mathbf{B}]$ *is by definition an orthogonal matrix we have* $F(s)F^t(s)=I$*. It is also clear that there is some matrix function* $A(s)$ *such that* $F'=F(s)A(s)$*. Also, Differentiating we have* $\frac{dF}{ds}(s)F^t(s)+F(s)\frac{dF}{ds}^t(s)=0$ *and so*

$$\begin{aligned}FAF^t+FA^tF^t&=0\\A+A^t&=0\end{aligned}$$

since F is invertible. Thus $A(s)$ is antisymmetric. But we already have established that $\frac{d\mathbf{T}}{ds}=\kappa\mathbf{N}$ *and* $\frac{d\mathbf{B}}{ds}=\tau\mathbf{N}$ *and so the result follows. It can be shown that the functions κ and τ completely determine a sufficiently regular curve up to reparameterization and rigid motions of space. The three vectors form a vector field along the curve c. At each point* $p=\mathbf{c}(s)$ *along the curve c the provide and oriented orthonormal basis (or frame) for vectors based at p. This basis is called the Frenet frame for the curve. Also, $\kappa(s)$ and $\tau(s)$ are called the (unsigned) curvature and torsion of the curve at* $\mathbf{c}(s)$*. While, κ is never negative by definition we may well have that $\tau(s)$ is negative. The curvature is, roughly speaking, the reciprocal of the radius of the circle which is tangent to* $\mathbf{c}$ *at* $\mathbf{c}(s)$ *and best approximates the curve at that point. On the other hand, τ measures the twisting of the plane spanned by* $\mathbf{T}$ *and* $\mathbf{N}$ *as we move along the curve. If* $\gamma:I\to\mathbb{R}^3$ *is an arbitrary speed curve then we define* $\kappa_\gamma(t):=\kappa\circ h^{-1}$ *where* $h:I'\to I$ *gives a unit speed reparameterization* $\mathbf{c}=\gamma\circ h:I'\to\mathbb{R}^n$*. Define the torsion function τ_γ for γ by* $\tau\circ h^{-1}$*. Similarly we have*

$$\begin{aligned}\mathbf{T}_\gamma(t)&:=\mathbf{T}\circ h^{-1}(t)\\\mathbf{N}_\gamma(t)&:=\mathbf{N}\circ h^{-1}(t)\\\mathbf{B}_\gamma(t)&:=B\circ h^{-1}(t)\end{aligned}$$

Exercise 6.24 *If* $\mathbf{c}:I\to\mathbb{R}^3$ *is a unit speed reparameterization of* $\gamma:I\to\mathbb{R}^3$ *according to* $\gamma(t)=\mathbf{c}\circ h$ *then show that*

1. $\mathbf{T}_\gamma(t)=\gamma'/\left\|\gamma'\right\|$

2. $\mathbf{N}_\gamma(t) = \mathbf{B}_\gamma(t) \times \mathbf{T}_\gamma(t)$

3. $\mathbf{B}_\gamma(t) = \frac{\gamma' \times \gamma''}{\|\gamma' \times \gamma''\|}$

4. $\kappa_\gamma = \frac{\|\gamma' \times \gamma''\|}{\|\gamma''\|^3}$

5. $\tau_\gamma = \frac{(\gamma' \times \gamma'') \cdot \gamma'''}{\|\gamma' \times \gamma''\|^2}$

Exercise 6.25 *Show that* $\gamma'' = \frac{dv}{dt}\mathbf{T}_\gamma + v^2 \kappa_\gamma \mathbf{N}_\gamma$ *where* $v = \|\gamma'\|$.

Example 6.26 *For a curve confined to a plane we haven't got the opportunity to define* $\mathbf{B}$ *or* τ*. However, we can obtain a more refined notion of curvature. We now consider the special case of curves in* $\mathbb{R}^2$*. Here it is possible to define a signed curvature which will be positive when the curve is turning counterclockwise. Let* $J : \mathbb{R}^2 \to \mathbb{R}^2$ *be given by* $J(a,b) := (-b,a)$*. The signed curvature* $\kappa_\gamma^\pm$ *of* γ *is given by*

$$\kappa_\gamma^\pm(t) := \frac{\gamma''(t) \cdot J\gamma'(t)}{\|\gamma'(t)\|^3}.$$

If γ *is a parameterized curve in* $\mathbb{R}^2$ *then* $\kappa_\gamma \equiv 0$ *then* γ *(parameterizes) a straight line. If* $\kappa_\gamma \equiv k_0 > 0$ *(a constant) then* γ *parameterizes a portion of a circle of radius* $1/k_0$*. The unit tangent is* $\mathbf{T} = \frac{\gamma'}{\|\gamma'\|}$*. We shall redefine the normal* $\mathbf{N}$ *to a curve to be such that* $\mathbf{T}, \mathbf{N}$ *is consistent with the orientation given by the standard basis of* $\mathbb{R}^2$*.*

Exercise 6.27 *If* $\mathbf{c} : I \to \mathbb{R}^2$ *is a unit speed curve then*

1. $\frac{d\mathbf{T}}{ds}(s) = \kappa_{\mathbf{c}}(s)\mathbf{N}(s)$

2. $\mathbf{c}''(s) = \kappa_{\mathbf{c}}(s)\left(J\mathbf{T}(s)\right)$

Example 6.28 *Consider the action of* $Aff^+(2)$ *on the affine plane* $\mathbf{A}^2$*. Let* $\mathcal{S}$ *be the family of all subsets of* $\mathbf{A}^2$ *which are zero sets of quadratic polynomials of the form*

$$ax^2 + bxy + cy^2 + dx + ey + f.$$

with the nondegeneracy condition $4ac - b^2 \neq 0$*. If* S *is simultaneously the zero set of nondegenerate quadratics* $p_1(x,y)$ *and* $p_2(x,y)$ *then* $p_1(x,y) = p_2(x,y)$*. Furthermore, if* $g \cdot S_1 = S_2$ *where* S_1 *is the zero set of* p_1 *then* S_2 *is the zero set of the nondegenerate quadratic polynomial* $p_2 := p_1 \circ g^{-1}$*. A little thought shows that we may as well replace* $\mathcal{S}$ *by the set of set of nondegenerate quadratic polynomials and consider the induced action on this set:* $(g,p) \mapsto p \circ g^{-1}$*. We may now obtain an invariant: Let* $S = p^{-1}(0)$ *and let* $I(S) =: I(p) =: sgn(4ac - b^2)$ *where* $p(x,y) = ax^2 + bxy + cy^2 + dx + ey + f$*. In this case,* $\mathcal{S}$ *is divided into exactly two equivalence classes*

Notice that in example 6.21 above the set of figures considered was obtained as the set of images for some convenient family of maps *into* the set M. This raises the possibility of a slight change of viewpoint: maybe we should be studying the maps themselves. Given a G-set M, we could consider a family of maps $\mathcal{C}$ from some fixed topological space T (or family of topological spaces like intervals in R say) into M which is invariant in the sense that if $c \in \mathcal{C}$ then the map $g \cdot c$ defined by $g \cdot c : t \mapsto g \cdot c(t)$ is also in $\mathcal{C}$. Then two elements c_1 and c_2 would be "congruent" if and only if $g \cdot c_1 = c_2$ for some $g \in G$. Now suppose that we find a function $I : \mathcal{C} \to \mathbb{C}$ which is an invariant in the sense that $I(g \cdot c) = I(c)$ for all $c \in \mathcal{C}$ and all $g \in G$. We do not necessarily obtain an invariant for the set of images of maps from $\mathcal{C}$. For example, if we consider the family of regular curves $c : [0, 1] \to \mathbf{E}^2$ and let $G = \mathrm{O}(2)$ with the action introduced earlier, then the *energy functional* defined by

$$E(c) := \int_0^1 \frac{1}{2} \left\| c'(t) \right\|^2 dt$$

is an invariant for the induced action on this set of curves but even if c_1 and c_2 have the same image it does not follow that $E(c_1) = E(c_2)$. Thus E is a geometric invariant of the curve but not of the set which is its image. In the elementary geometric theory of curves in Euclidean spaces one certainly wants to understand curves as sets. One starts out by studying the maps $c : I \to \mathbf{E}^n$ first. In order to get to the geometry of the subsets which are images of regular curves one must consider how quantities defined depend on the choice of parameterization. Alternatively, a standard parameterization (parameterization by arc length) is always possible and this allows one to provide geometric invariants of the image sets (see Appendix **??**) .

Similarly, example 6.22 above invites us to think about maps *from* M into some topological space T (like $\mathbb{R}$ for example). We should pick a family of maps $\mathcal{F}$ such that if $f \in \mathcal{F}$ then $g \cdot f$ is also in $\mathcal{F}$ where $g \cdot f : x \mapsto f(g^{-1} \cdot x)$. Thus we end up with G acting on the set $\mathcal{F}$. This is an induced action. We have chosen to use g^{-1} in the definition so that we obtain a left action on $\mathcal{F}$ rather than a right action. In any case, we could then consider f_1 and f_2 to be congruent if $g \cdot f_1 = f_2$ for some $g \in G$.

There is much more to the study of figure in Euclidean space than we have indicated here. We prefer to postpone introduction of these concepts until after we have a good background in manifold theory and then introduce the more general topic of Riemannian geometry. Under graduate level differential geometry courses usually consist mostly of the study of curves and surfaces in $3-$dimensional Euclidean space and the reader who has been exposed to this will already have an idea of what I am talking about. A quick review of curves and surfaces is provided in appendix **??**. The study of Riemannian manifolds and submanifolds that we take up in chapters **??** and **??**.

We shall continue to look at simple homogeneous spaces for inspiration but now that we are adding in the notion of time we might try thinking in a more dynamic way. Also, since the situation has become decidedly more physical it would pay to start considering the possibility that the question of what counts as geometric might be replaced by the question of what counts as physical. We must eventually also ask what other group theoretic principals (if any) are need to understand the idea of invariants of motion such as conservation laws.

6.2.4 Galilean Spacetime

Spacetime is the set of all events in a (for now 4 dimensional) space. At first it might seem that time should be included by simply taking the Cartesian product of space $\mathbf{E}^3$ with a copy of $\mathbb{R}$ that models time: Spacetime=Space×$\mathbb{R}$. Of course, this leads directly to $\mathbb{R}^4$, or more appropriately to $\mathbf{E}^3 \times \mathbb{R}$. Topologically this is right but the way we have written it implies an inappropriate and unphysical decomposition of time and space. If we only look at (affine) self transformations of $\mathbf{E}^3 \times \mathbb{R}$ that preserve the decomposition then we are looking at what is sometimes called Aristotelean spacetime (inappropriately insulting Aristotel). The problem is that we would not be taking into account the relativity of motion. If two spaceships pass each other moving a constant relative velocity then who is to say who is moving and who is still (or if both are moving). The decomposition $\mathbf{A}^4 = \mathbb{R}^4 = \mathbf{E}^3 \times \mathbb{R}$ suggests that a body is at rest if and only if it has a career (worldline) of the form $p \times \mathbb{R}$; always stuck at p in other words. But relativity of constant motion implies that no such assertion can be truly objective relying as it does on the assumption that one coordinate system is absolutely "at rest". Coordinate transformations between two sets of coordinates on $4-$dimensional spacetime which are moving relative to one another at constant nonzero velocity should mix time and space in some way. There are many ways of doing this and we must make a choice. The first concept of spacetime that we can take seriously is Galilean spacetime. Here we lose absolute motion (thankfully) but retain an absolute notion of time. In Galilean spacetime, it is entirely appropriate to ask whether two events are simultaneous or not. Now the idea that simultaneity is a well define is very intuitive but we shall shortly introduce Minkowski spacetime as more physically realistic and then discover that simultaneity will become a merely relative concept! The appropriate group of coordinate changes for Galilean spacetime is the **Galilean group** and denoted by Gal. This is the group of transformations of $\mathbb{R}^4$ (thought of as an affine space) generated by the follow three types of transformations:

1. **Spatial rotations**: These are of the form

$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & R & \\ 0 & & & \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}$$

where $R \in \mathrm{O}(3)$,

2. **Translations of the origin**

$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_0 \\ x_0 \\ y_0 \\ z_0 \end{bmatrix}$$

   for some $(t_0, x_0, y_0, z_0) \in \mathbb{R}^4$.

3. **Uniform motions**. These are transformations of the form

$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} t \\ x + v_1 t \\ y + v_2 t \\ z + v_3 t \end{bmatrix}$$

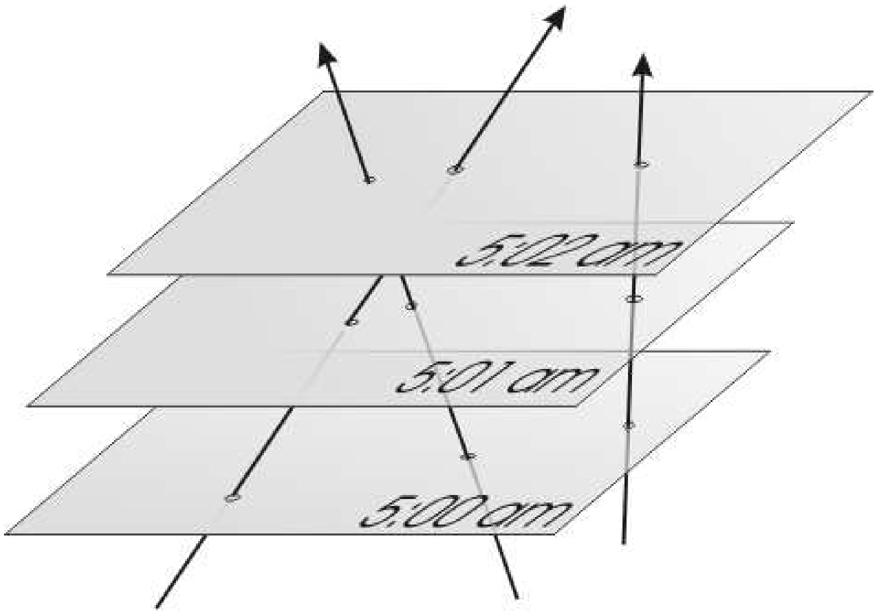
   for some (velocity) $v = (v_1, v_2, v_3)$. The picture here is that there are two observers each making measurements with respect to their own rectangular spatial coordinate systems which are moving relative to each other with a constant velocity v. Each observer is able to access some universally available clock or synchronized system of clock which will give time measurements that are unambiguous except for choice of the "zero time" (and choice of units which we will assume fixed).

The set of all coordinates related to the standard coordinates of $\mathbb{R}^4$ by an element of Gal will be referred to as Galilean inertial coordinates. When the set $\mathbb{R}^4$ is provided with the action by the Galilean group we refer to it a Galilean spacetime. We will not give a special notation for the space. As a matter of notation it is common to denoted (t, x, y, z) by (x^0, x^1, x^2, x^3). The spatial part of an inertial coordinate system is sometimes denoted by $\mathbf{r} := (x, y, z) = (x^1, x^2, x^3)$.

Basic fact 2: The "spatial separation" between any two events E_1 and E_2 in Galilean spacetime is calculated as follows: Pick some Galilean inertial coordinates $x = (x^0, x^1, x^2, x^3)$ and let $\Delta\mathbf{r} := \mathbf{r}(E_2) - \mathbf{r}(E_1)$ then the (square of the) spatial separation is calculated as $s = |\Delta\mathbf{r}| = \sum_{i=1}^{3}(x^i(E_2) - x^i(E_1))^2$. The result **definitely does** depend on the choice of Galilean inertial coordinates. Spacial separation is a relative concept in Galilean spacetime. On the other hand, the **temporal separation** $|\Delta t| = |t(E_2) - t(E_1)|$ **does not depend of the choice of coordinates**. Thus, it makes sense in this world to ask whether two events occurred at the same time or not.

6.2.5 Minkowski Spacetime

As we have seen, the vector space $\mathbb{R}^4$ may be provided with a special scalar product given by $\langle x, y \rangle := x^0 y^0 - \sum_{i=1}^{3} x^i y^i$ called the Lorentz scalar product (in the setting of Geometry this is usual called a Lorentz metric). If one considers the physics that this space models then we should actual have $\langle x, y \rangle := c^2 x^0 y^0 -$



$\sum_{i=1}^{3} x^i y^i$ where the constant c is the speed of light in whatever length and time units one is using. On the other hand, we can follow the standard trick of using units of length and time such that in these units the speed of light is equal to 1. This scalar product space is sometimes denoted by $\mathbb{R}^{1,3}$. More abstractly, a Lorentz vector space $V^{1,3}$ is a 4−dimensional vector space with scalar product $\langle .,. \rangle$ which is isometric to $\mathbb{R}^{1,3}$. An orthonormal basis for a Lorentz space is by definition a basis (e_0, e_1, e_2, e_3) such that the matrix which represents the scalar product with respect to basis is

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

Once we pick such a basis for a Lorentz space we get an isomorphism with $\mathbb{R}^{1,3}$. Physically and geometrically, the standard basis of $\mathbb{R}^{1,3}$ is just one among many orthonormal bases so if one is being pedantic, the abstract space $V^{1,3}$ would be more appropriate. The group associated with a Lorentz scalar product space $V^{1,3}$ is the Lorentz group $L = \mathrm{O}(V^{1,3})$ which is the group of linear isometries of $V^{1,3}$. Thus $g \in \mathrm{O}(V^{1,3}, \langle .,. \rangle)$ if and only if

$$\langle gv, gw \rangle = \langle v, w \rangle$$

for all $v, w \in V^{1,3}$.

Now the origin is a preferred point for a vector space but is not a physical reality and so we want to introduce the appropriate metric affine space.

Definition 6.29 ***Minkowski space*** *is the metric affine space M^{1+3} (unique up to isometry) with difference space given by a Lorentz scalar product space $V^{1,3}$ (Lorentz space).*

Minkowski space is sometimes referred to as Lorentzian affine space.

The group of coordinate transformations appropriate to M^{1+3} is described the group of affine transformations of M^{1+3} whose linear part is an element of $\mathrm{O}(V^{1,3}, \langle .,. \rangle)$. This is called the Poincaré group P. If we pick an origin $p \in M^{1+3}$ an orthonormal basis for $V^{1,3}$ then we may identify M^{1+3} with $\mathbb{R}^{1,3}$ (as an affine space[5]). Having made this arbitrary choice the Lorentz group is identified with the group of matrices $\mathrm{O}(1,3)$ which is defined as the set of all 4×4 matrices Λ such that

$$\Lambda^t \eta \Lambda = \eta.$$

and a general Poincaré transformation is of the form $x \mapsto \Lambda x + x_0$ for $x_0 \in \mathbb{R}^{1,3}$ and $\Lambda \in \mathrm{O}(1,3)$. We also have an alternative realization of M^{1+3} as the set of all column vectors of the form

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \in \mathbb{R}^5$$

where $x \in \mathbb{R}^{1,3}$. Then the a Poincaré transformation is given by a matrix of the form the group of matrices of the form

$$\begin{bmatrix} 1 & 0 \\ b & Q \end{bmatrix} \text{ for } Q \in \mathrm{O}(1,3).$$

Basic fact 3. The spacetime interval between two events E_1 and E_2 in Minkowski spacetime is may be calculated in any (Lorentz) inertial coordinates by $\Delta\tau := -(\Delta x^0)^2 + \sum_{i=1}^4 (\Delta x^i)^2$ where $\Delta x = x(E_2) - x(E_1)$. The result is independent of the choice of coordinates. Spacetime separation in M^4 is "absolute". On the other hand, in Minkowski spacetime spatial separation and temporal separation are both relative concepts and only make sense within a particular coordinate system. It turns out that real spacetime is best modeled by Minkowski spacetime (at least locally and in the absence of strong gravitational fields). This has some counter intuitive implications. For example, it does not make any sense to declare that some supernova exploded into existence at the precise time of my birth. There is simply no fact of the matter. It is similar to declaring that the star Alpha Centaury is "above" the sun. Now if one limits oneself to coordinate systems that have a small relative motion with respect to each other then we may speak of events occurring at the same time (approximately). If one is speaking in terms of precise time then even the uniform motion of a train relative to an observer on the ground destroys our ability to declare that two events happened at the same time. If the fellow on the train uses the best system of measurement (best inertial coordinate system)

[5] The reader will be relieved to know that we shall eventually stop needling the reader with the pedantic distinction between $\mathbb{R}^n$ and $\mathbb{A}^n$, $\mathbb{E}^n$ and so on.

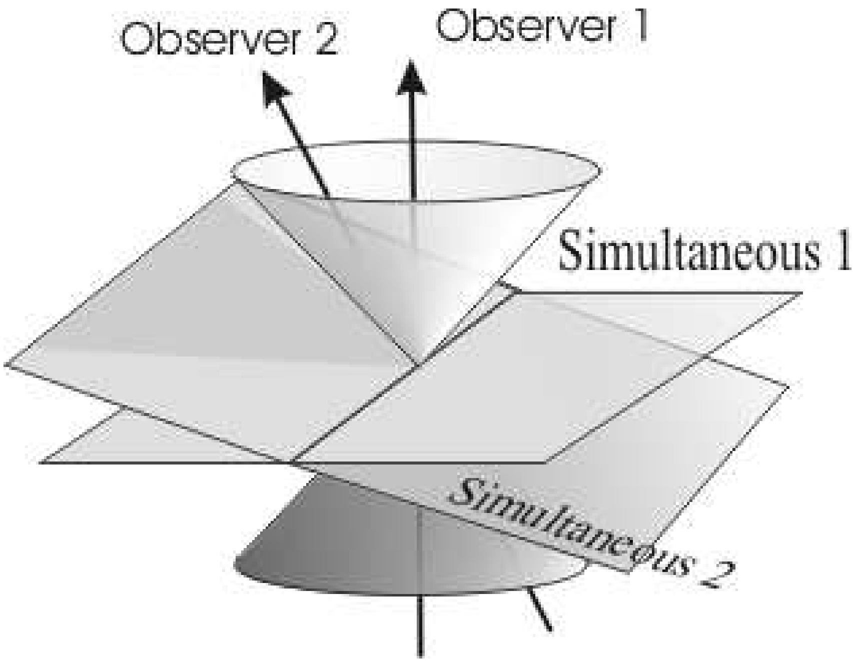


Figure 6.1: Relativity of Simultaneity

available to him and his sister on the ground does the same then it is possible that they may not agree as to whether or not two firecrackers, one on the train and one on the ground, exploded at the same time or not. It is also true that the question of whether the firecrackers exploded when they were 5 feet apart or not becomes relativized. The sense in which the spaces we have introduced here are "flat" will be made clear when we study curvature.

In the context of special relativity, there are several common notations used for vectors and points. For example, if (x^0, x^1, x^2, x^3) is a Lorentz inertial coordinate system then we may also write (t, x, y, z) or $(t, \mathbf{r})$ where $\mathbf{r}$ is called the spatial position vector. In special relativity it is best to avoid curvilinear coordinates because the simple from that the equations of physics take on when expressed in the rectangular inertial coordinates is ruined in a noninertial coordinate systems. This is implicit in all that we do in Minkowski space. Now while the set of points of M^4 has lost its vector space structure so that we no longer consider it legitimate to add points, we still have the ability to take the difference of two points and the result will be an element of the scalar product space $V^{1,3}$. If one takes the difference between p and q in M^4 we have an element $v \in V$ and if we want to consider this vector as a tangent vector based at p then we can write it as (p, v) or as v_p. To get the expression for the inner product of the tangent vectors $(p, v) = \overrightarrow{pq_1}$ and $(p, w) = \overrightarrow{pq_2}$ in coordinates (x^μ), let $v^\mu := x^\mu(q_1) - x^\mu(p)$ and $w^\mu := x^\mu(q_2) - x^\mu(p)$ and then calculate:

$\langle v_p, w_p\rangle = \eta_{\mu\nu}v^\mu w^\nu$. Each tangent space is naturally an inner product. Of course, associated with each inertial coordinate system (x^μ) there is a set of 4−vectors fields which are the coordinate vector fields denoted by $\partial_0, \partial_1, \partial_2$,and ∂_3. A contravariant vector or vector field $v = v^\mu\partial_\mu$ on M^4 has a twin in covariant form[6]. This is the covector (field) $v^\flat = v_\mu dx^\mu$ where $v_\mu := \eta_{\mu\nu}v^\nu$. Similarly if $\alpha = \alpha_\mu dx^\mu$ is a covector field then there is an associated vector field given $\alpha^\# = \alpha^\mu\partial_\mu$ where $\alpha^\mu := \alpha_\nu\eta^{\mu\nu}$ and the matrix $(\eta^{\mu\nu})$ is the inverse of $(\eta_{\mu\nu})$ which in this context might seem silly since $(\eta_{\mu\nu})$ is its own inverse. The point is that this anticipates a more general situation and also maintains a consistent use of index position. The Lorentz inner product is defined for any pair of vectors based at the same point in spacetime and is usually called the Lorentz metric-a special case of a semi-Riemannian metric which is the topic of a later chapter. If v^μ and w^μ are the components of two vectors in the current Lorentz frame then $\langle v, w\rangle = \eta_{\mu\nu}v^\mu w^\nu$.

Definition 6.30 *A 4-vector v is called* ***space-like*** *if and only if $\langle v, v\rangle < 0$,* ***time-like*** *if and only if $\langle v, v\rangle > 0$ and* ***light-like*** *if and only if $\langle v, v\rangle = 0$. 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 6.31 (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 $\langle v, v\rangle < 0$.*

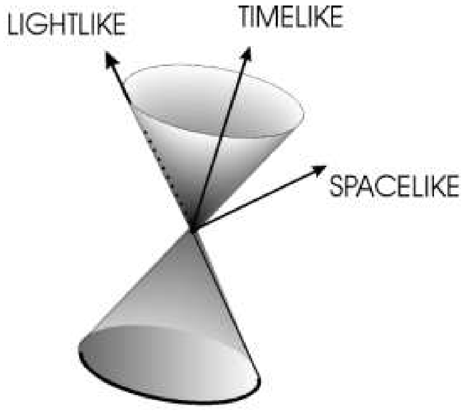
Definition 6.32 *A vector v based at a point in M^4 such that $\langle\partial_0, v\rangle > 0$ will be called* ***future pointing*** *and the set of all such forms the interior of the "future" light-cone.*

Definition 6.33 *A Lorentz transformation that sends future pointing timelike vector to future pointing timelike vectors is called an orthochronous Lorentz transformation.*

Now an important point is that there are two different ways that physics gives rise to a vector and covector field. The first is exemplified by the case of a single particle of mass m in a state of motion described in our coordinate system by a curve $\gamma : t \to (t, x(t), y(t), z(t))$ such that $\dot\gamma$ is a timelike vector for all parameter values t. The 3-velocity is a concept that is coordinate dependent and is given by $\mathbf{v} = (\frac{dx}{dt}(t), \frac{dy}{dt}(t), \frac{dz}{dt}(t))$. In this case, the associated 4−velocity u is a vector field along the curve γ an defined to be the unit vector (that is $\langle u, u\rangle = -1$) which is in the direction of $\dot\gamma$. The the contravariant 4−momentum is the 4-vector field along γ given by and $p = mu$. The other common situation is where matter is best modeled like a fluid or gas.

Now if γ is the curve which gives the career of a particle then the spacetime interval between two events in spacetime given by $\gamma(t)$ and $\gamma(t+\epsilon)$ is $\langle\gamma(t+$

[6] This is a special case of an operation that one can use to get 1-forms from vector fields and visa versa on any semi-Riemannian manifold and we will explain this in due course.



$\epsilon) - \gamma(t), \gamma(t+\epsilon) - \gamma(t)\rangle$ and should be a timelike vector. If ϵ is small enough then this should be approximately equal to the time elapsed on an ideal clock traveling with the particle. Think of zooming in on the particle and discovering that it is actually a spaceship containing a scientist and his equipment (including say a very accurate atomic clock).The actual time that will be observed by the scientist while traveling from significantly separated points along her career, say $\gamma(t_1)$ and $\gamma(t)$ with $t > t_1$ will be given by

$$\tau(t) = \int_{t_1}^{t} |\langle \dot{\gamma}, \dot{\gamma} \rangle| \; dt$$

Standard arguments with change of variables show that the result is independent of a reparameterization. It is also independent of the choice of Lorentz coordinates. We have skipped the physics that motivates the interpretation of the above integral as an elapsed time but we can render it plausible by observing that if γ is a straight line $t \mapsto (t, x_1 + v_1 t, x_2 + v_2 t, x_3 + v_3 t)$ which represents a uniform motion at constant speed then by a change to a new Lorentz coordinate system and a change of parameter the path will be described simply by $t \rightarrow (t, 0, 0, 0)$. The scientist is at rest in her own Lorentz frame. Now the integral reads $\tau(t) = \int_0^t 1 \, dt = t$. For any timelike curve γ the quantity $\tau(t) = \int_{t_1}^{t} |\langle \dot{\gamma}, \dot{\gamma} \rangle|^{1/2} \, dt$ is called the **proper time** of the curve from $\gamma(t_1)$ to $\gamma(t)$.

The famous twins paradox is not really a true paradox since an adequate explanation is available and the counter-intuitive aspects of the story are actually physically correct. In short the story goes as follows. Joe's twin Bill leaves in a spaceship and travels at say 98% of the speed of light to a distant star and then returns to earth after 100 years of earth time. Let use make two simplifying assumptions which will not change the validity of the discussion. The first is that the earth, contrary to fact, is at rest in some inertial coordinate system (replace the earth by a space station in deep space if you like). The second assumption is that Joe's twin Bill travels at constant velocity on the forward and return trip. This entails an unrealistic instant deceleration and acceleration at the star; the turn around but the essential result is the same if we make this part more realistic. Measuring time in the units where $c = 1$, the first half of Bill's trip is given $(t, .98t)$ and second half is given by $(t, -.98t)$. Of course, this entails that in the earth frame the distance to the star is $.98\frac{lightyears}{year} \times 100$ years $= 98$ light-years. Using a coordinate system fixed relative to the earth we calculate the proper time experienced by Bill:

$$\begin{aligned}\int_0^{100} |\langle\dot\gamma,\dot\gamma\rangle|\,dt &= \int_0^{50}\sqrt{|-1+(.98)^2|}dt + \int_0^{50}\sqrt{|-1+(-.98)^2|}dt \\ &= 2\times 9.\,949\,9 = 19.\,900\end{aligned}$$

Bill is 19.9 years older when we returns to earth. On the other hand, Joe's has aged 100 years! One may wonder if there is not a problem here since one might argue that from Joe's point of view it was the earth that travelled away from him and then returned. This is the paradox but it is quickly resolved once it is pointed out that the is not symmetry between Joe's and Bill's situation. In order to return to earth Bill had to turn around which entail an acceleration and more importantly prevented a Bill from being stationary with respect to any single Lorentz frame. Joe, on the other hand, has been at rest in a single Lorentz frame the whole time. The age difference effect is real and a scaled down version of such an experiment involving atomic clocks put in relative motion has been carried out and the effect measured.

6.2.6 Hyperbolic Geometry

We have looked at affine spaces, Euclidean spaces, Minkowski space and Galilean spacetime. Each of these has an associated group and in each case straight lines are maps to straight lines. In a more general context the analogue of straight lines are called geodesics a topic we will eventually take up in depth. The notion of distance makes sense in a Euclidean space essentially because each tangent space is an inner product space. Now each tangent space of $\mathbf{E}^n$ is of the form $\{p\} \times \mathbb{R}^n$ for some point p. This is the set of tangent vectors at p. If we choose an orthonormal basis for $\mathbb{R}^n$, say $e_1, ..., e_n$ then we get a corresponding orthonormal basis in each tangent space which we denote by $e_{1,p}, ..., e_{n,p}$ and where $e_{i,p} := (p, e_i)$ for $i = 1, 2, ..., n$. With respect to this basis in the tangent space the matrix which represents the inner product is the identity matrix $I = (\delta_{ij})$. This is true uniformly for each choice of p. One possible generalization is to let the matrix vary with p. This idea eventually leads to Riemannian geometry. We can give an important example right now. We have already seen that the subgroup $\mathrm{SL}(2, \mathbb{R})$ of the group $\mathrm{SL}(2, \mathbb{C})$ also acts on the complex plane and in fact fixes the upper half plane $\mathbb{C}^+$. In each tangent space of the upper half plane we may put an inner product as follows. If $v_p = (p, v)$ and $w_p = (p, w)$ are in the tangent space of p then the inner product is

$$\langle v_p, w_p \rangle := \frac{v^1 w^1 + v^2 w^2}{y^2}$$

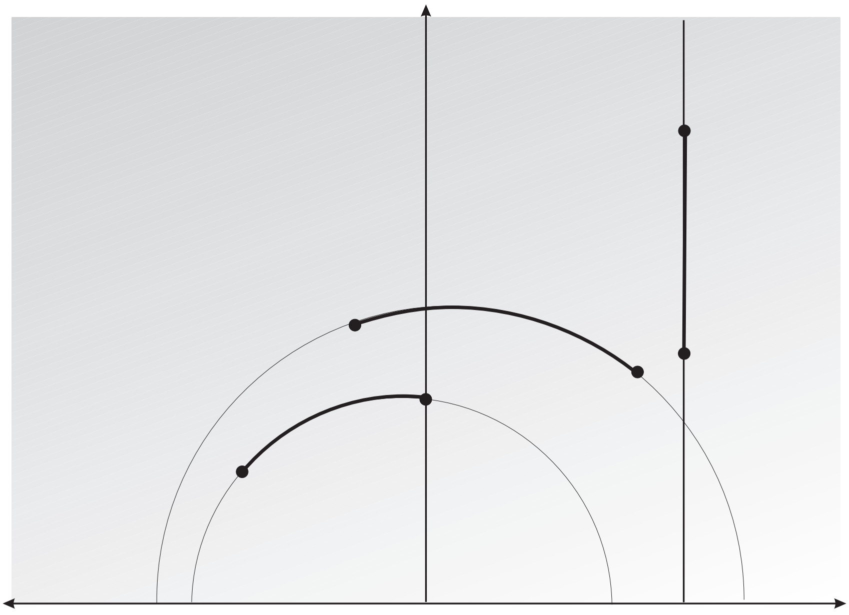
where $p = (x, y) \sim x + iy$ and $v = (v^1, v^2)$, $w = (w^1, w^2)$. In this context, the assignment of an inner product the tangent space at each point is called a metric. We are identifying $\mathbb{R}^2$ with $\mathbb{C}$. Now the length of a curve $\gamma : t \mapsto (x(t), y(t))$ defined on $[a, b]$ and with image in the upper half-plane is given by

$$\int_a^b \|\dot{\gamma}(t)\|\, dt = \int_a^b \left(\frac{\dot{x}(t)^2 + \dot{y}(t)^2}{y(t)^2} \right)^{1/2} dt.$$

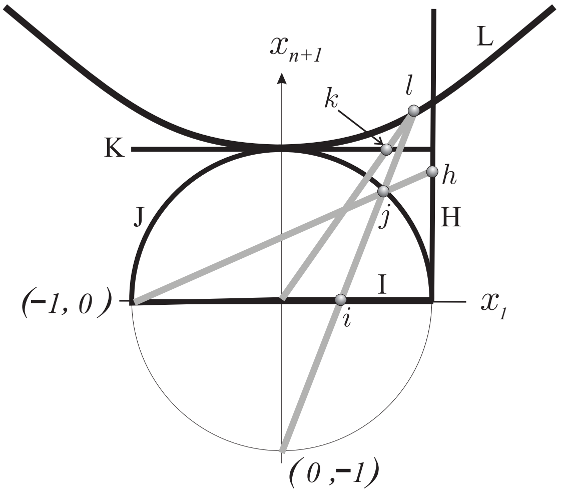
We may define arc length starting from some point $p = \gamma(c)$ along a curve $\gamma : [a, b] \to \mathbb{C}^+$ in the same way as was done in calculus of several variables:

$$s := l(t) = \int_a^t \left(\frac{\dot{x}(\tau)^2 + \dot{y}(\tau)^2}{y(\tau)^2} \right)^{1/2} d\tau$$

The function $l(t)$ is invertible and it is then possible to reparameterize the curve by arc length $\widetilde{\gamma}(s) := \gamma(l^{-1}(s))$.The distance between any two points in the upper half-plane is the length of the shortest curve that connects the two points. Of course, one must show that there is always a shortest curve. It turns out that the shortest curves, the geodesics, are curved segments lying on circles which meet the real axis normally, or are vertical line segments.



The upper half plane with this notion of distance is called the Poincaré upper half plane and is a realization of an abstract geometric space called the hyperbolic plane. The geodesics are the “straight lines” for this geometry.



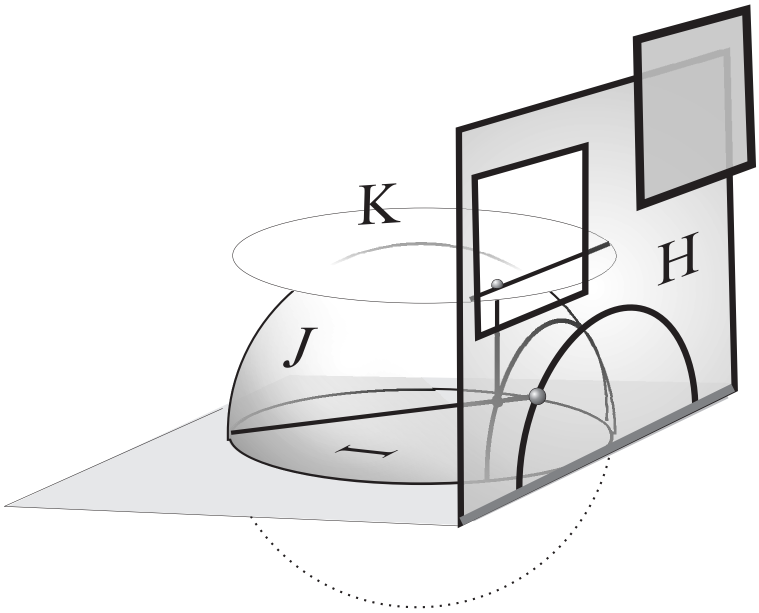
6.2.7 Models of Hyperbolic Space

hjj (do this section-follow'flavors of geometry)

$$\begin{aligned} H &= \{(1, x_2, ..., x_{n+1}) : x_{n+1} > 0\} \\ I &= \{(x_1, ..., x_n, 0) : x_1^2 + \cdots + x_n^2 < 1\} \\ J &= \{(x_1, ..., x_{n+1}) : x_1^2 + \cdots + x_{n+1}^2 = 1 \text{ and } x_{n+1} > 0\} \\ K &= \{(x_1, ..., x_n, 1) : x_1^2 + \cdots + x_n^2 < 1\} \\ L &= \{(x_1, ..., x_n, x_{n+1}) : x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1 \text{ and } x_{n+1} > 0\} \end{aligned}$$

$$\begin{aligned} &\alpha : J \to H;\ \ (x_1, ..., x_{n+1}) \mapsto (1, 2x_2/(x_1 + 1), ..., 2x_{n+1}/(x_1 + 1)) \\ &\beta : J \to I;\ \ (x_1, ..., x_{n+1}) \mapsto (x_1/(x_{n+1} + 1), ..., x_n/(x_{n+1} + 1), 0) \\ &\gamma : K \to J;\ \ (x_1, ..., x_n, 1) \mapsto (x_1, ..., x_n, \sqrt{1 - x_1^2 - \cdots - x_n^2}) \\ &\delta : L \to J;\ \ (x_1, ..., x_{n+1}) \mapsto (x_1/x_{n+1}, ..., x_n/x_{n+1}, 1/x_{n+1}) \end{aligned}$$

The map a is stereographic projection with focal point at $(-1, 0, ..., 0)$ and maps j to h in the diagrams. The map β is a stereographic projection with focal point at $(0, ..., 0, -1)$ and maps j to i in the diagrams. The map γ is vertical orthogonal projection and maps k to j in the diagrams. The map δ is stereographic projection with focal point at $(0, ..., 0, -1)$ as before but this time



projecting onto the hyperboloid L.

$$\begin{aligned}
ds_H^2 &= \frac{dx_2^2 + \cdots + dx_{n+1}^2}{x_{n+1}^2};\\
ds_I^2 &= 4\frac{dx_1^2 + \cdots + dx_n^2}{\left(1 - x_1^2 - \cdots - x_n^2\right)^2};\\
ds_J^2 &= \frac{dx_1^2 + \cdots + dx_n^2}{x_{n+1}^2}\\
ds_K^2 &= \frac{dx_1^2 + \cdots + dx_n^2}{1 - x_1^2 - \cdots - x_n^2} + \frac{\left(x_1 dx_1 + \cdots + x_n dx_n\right)^2}{\left(1 - x_1^2 - \cdots - x_n^2\right)^2}\\
dx_L^2 &= dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2
\end{aligned}$$

To get a hint of the kind of things to come, notice that we can have two geodesics which start at nearby points and start of in the same direction and yet the distance between corresponding points increases. In some sense, the geometry acts like a force which (in this case) repels nearby geodesics. The specific invariant responsible is the curvature. Curvature is a notion that has a generalization to a much larger context and we will see that the identification of curvature with a force field is something that happens in both Einstein's general theory of relativity and also in gauge theoretic particle physics. One of the goals of this book is to explore this connection between force and geometry.

Another simple example of curvature acting as a force is the following. Imaging that the planet was completely spherical with a smooth surface. Now imagine the people a few miles apart but both on the equator. Give each person a pair of roller skates and once they have them on give them a simultaneous push toward the north. Notice that they start out with parallel motion. Since they will eventually meet at the north pole the distance between them must be shrinking. What is pulling them together. Regardless of whether one wants to call the cause of their coming together a force or not, it is clear that it is the curvature of the earth that is responsible. Readers familiar with the basic idea behind General Relativity will know that according to that theory, the "force" of gravity is due to the curved shape of 4-dimensional spacetime. The origin of the curvature is said to be due to the distribution of mass and energy in space.

6.2.8 The Möbius Group

Let $\mathbb{C}^{+\infty}$ denote the set $\mathbb{C} \cup \{\infty\}$. The topology we provide $\mathbb{C}^{+\infty}$ is generated by the open subsets of $\mathbb{C}$ together with sets of the form $O \cup \{\infty\}$ where O is the compliment of a compact subset of $\mathbb{C}$. This is the $1-$**point compactification** of $\mathbb{C}$. Topologically, $\mathbb{C}^{+\infty}$ is just the sphere S^2.

Now consider the group $\mathrm{SL}(2,\mathbb{C})$ consisting of all invertible 2×2 complex matrices with determinant 1. We have an action of $\mathrm{SL}(2,\mathbb{C})$ on $\mathbb{C}^{+\infty}$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}.$$

For any fixed $A \in \mathrm{SL}(2,\mathbb{C})$ them map $z \mapsto A \cdot z$ is a homeomorphism (and much more as we shall eventually see). Notice that this is not the standard action of $\mathrm{SL}(2,\mathbb{C})$ on $\mathbb{C}^2$ by multiplication $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \mapsto A \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ but there is a relationship between the two actions. Namely, let

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

and define $z = z_1/z_2$ and $w = w_1/w_1$. Then $w = A \cdot z = \frac{az+b}{cz+d}$. The two component vectors $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ are sometimes called spinors in physics.

Notice is that for any $A \in \mathrm{SL}(2,\mathbb{C})$ the homeomorphisms $z \mapsto A \cdot z$ and $z \mapsto (-A) \cdot z$ are actually equal. Thus group theoretically, the set of all distinct transformations obtained in this way is really the quotient group $\mathrm{SL}(2,\mathbb{C})/\{I,-I\}$ and this is called the **Möbius group** or group of Möbius transformations.

There is quite a bit of geometry hiding in the group $\mathrm{SL}(2,\mathbb{C})$ and we will eventually discover a relationship between $\mathrm{SL}(2,\mathbb{C})$ and the Lorentz group.

When we add the notion of time into the picture we are studying spaces of "events" rather than "literal" geometric points. On the other hand, the spaces of evens might be considered to have a geometry of sorts and so in that sense the events are indeed points. An approach similar to how we handle Euclidean space will allow us to let spacetime be modeled by a Cartesian space $\mathbb{R}^4$; we find a family of coordinates related to the standard coordinates by the action of a group. Of course, in actual physics, the usual case is where space is 3-dimensional and spacetime is $4-$dimensional so lets restrict attention this case. But what is the right group? What is the right geometry? We now give two answers to this question. The first one corresponds to intuition quite well and is implicit in the way physics was done before the advent of special relativity. The idea is that there is a global measure of time that applies equally well to all points of $3-$dimensional space and it is unique up to an affine change of parameter $t \mapsto t' = at + b$. The affine change of parameter corresponds to a change in units.

Chapter 7

Singular Distributions

Lemma 7.1 *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 **??** the order that we compose the flows does not change the value of $f(t_1, ..., t_n)$. Thus

$$\begin{aligned}
&\frac{\partial}{\partial t_i} f(t_1, ..., t_n) \\
&= \frac{\partial}{\partial t_i} Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x) \\
&= \frac{\partial}{\partial t_i} Fl_{t_i}^{X_i} \circ Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x) \text{ (put the } i\text{-th flow first)} \\
&X_i(Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x)).
\end{aligned}$$

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 7.2 *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 7.3 *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 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 7.4 *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 ι^*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 $X(\iota(s)) \in T_s\iota(T_sS)$ 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 ι^*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 **??** 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 that pass through a fixed x in M to obtain a unique maximal integral submanifold through x. We have prove the following result:

Proposition 7.5 *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 7.6 *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 7.7 *Show that $\mathcal{S}(\mathcal{X})$ is the smallest stable family of local vector fields containing $\mathcal{X}$.*

Definition 7.8 *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 7.9 *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 7.10 *Convince yourself that $\mathrm{aut}_{loc}(\Delta)$ is stable.*

For the next theorem recall the definition of $\mathfrak{X}_\Delta$.

Theorem 7.11 *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 **??**

$$\mathrm{Fl}_{-t}^X \circ \iota = \iota \circ \mathrm{Fl}_{-t}^{\iota^* X}$$

where $\iota : \mathcal{L}_x \hookrightarrow M$ is inclusion. Thus

$$\begin{aligned} T_x(\mathrm{Fl}^X_{-t})(\Delta_x) &= T(\mathrm{Fl}^X_{-t}) \cdot T_x\iota \cdot (T_x\mathcal{L}_x) \\ &= T(\iota \circ \mathrm{Fl}^{\iota^* X}_{-t}) \cdot (T_x\mathcal{L}_x) \\ &= T\iota T_x(\mathrm{Fl}^{\iota^* X}_{-t}) \cdot (T_x\mathcal{L}_x) \\ &= T\iota T_{\mathrm{Fl}^{\iota^* X}_{-t}(x)}\mathcal{L}_x = \Delta_{\mathrm{Fl}^{\iota^* X}_{-t}(x)}. \end{aligned}$$

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 $((\mathrm{Fl}^X_t)^* Y)(x) \in \Delta$ so $(\mathrm{Fl}^X_t)^* Y)$ 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 $(\mathrm{Fl}^X_t)^* Y \in \mathfrak{X}_\Delta$ 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

$$T(\mathrm{Fl}^X_t)\Delta_x = \Delta_{\mathrm{Fl}^X_t(x)}$$

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

$$f(t^1, ..., t^n) := (\mathrm{Fl}^{X_1}_{t^1} \mathrm{Fl}^{X_2}_{t^2} \circ \cdots \circ \mathrm{Fl}^{X_k}_{t^k})(x)$$

which is defined (and smooth) near $0 \in \mathbb{R}^k$. As in lemma 7.1 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_xS is spanned by $\frac{\partial f}{\partial t^j}(0)$ for $j = 1, 2, ..., k$ and

$$\begin{aligned} \frac{\partial f}{\partial t^j}(0) &= \left.\frac{\partial}{\partial t^j}\right|_0 (\mathrm{Fl}^{X_1}_{t^1} \mathrm{Fl}^{X_2}_{t^2} \circ \cdots \circ \mathrm{Fl}^{X_k}_{t^k})(x) \\ &= T(\mathrm{Fl}^{X_1}_{t^1} \mathrm{Fl}^{X_2}_{t^2} \circ \cdots \circ \mathrm{Fl}^{X_{j-1}}_{t^{j-1}}) X_j((\mathrm{Fl}^{X_j}_{t^j} \mathrm{Fl}^{X_{j+1}}_{t^{j+1}} \circ \cdots \circ \mathrm{Fl}^{X_k}_{t^k})(x)) \\ &= ((\mathrm{Fl}^{X_1}_{-t^1})^* (\mathrm{Fl}^{X_2}_{-t^2})^* \circ \cdots \circ (\mathrm{Fl}^{X_{j-1}}_{-t^{j-1}})^* X_j)(f(t^1, ..., t^n)). \end{aligned}$$

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 8

Infinite Dimnsional Manifolds

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

8.1 Topological Manifolds

A **refinement** of an open cover $\{U_\beta\}_{\beta \in B}$ of a topological space X is another open cover $\{V_i\}_{i \in I}$ 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 may be described as a cover $\{V_i\}_{i \in I}$ together with a set map $i \mapsto \beta(i)$ of the index sets $I \to B$ such that $V_i \subset U_{\beta(i)} M.$ for all i. We say that a cover $\{V_i\}_{i \in I}$ is a **locally finite** cover if in every point of X has a neighborhood that intersects only a finite number of the sets from the cover.

We recall a couple of concepts from point set topology: 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. A **base** (or basis) for the topology of a topological space X is a collection $\mathfrak{B}$ of open 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 8.1 *A* ***topological manifold*** *of dimension n is a second countable Hausdorff topological space, say M, such that every point $p \in M$ is contained in some open set U_p that is the domain of a homeomorphism $\phi : U_p \to V$ onto an open subset of $\mathbb{R}^n$.*

Thus we say that topological manifold M is "locally Euclidean".

If we had allowed n to vary with the point 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, on a connected space, this in fact not possible. It is a consequence of a 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 easy to prove if the manifold has a differentiable structure (defined below) but more difficult in general. We shall simply record Brower's theorem:

Theorem 8.2 (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 M 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 that is the domain of a homeomorphism $\psi : U \to V$ onto an open subset V of some Euclidean half space $\mathbb{R}^n_- =: \{\vec{x} : x^1 \leq 0\}$[1]. A point that is mapped to the hypersurface $\partial\mathbb{R}^n_- = \mathbb{R}^n_0 =: \{\vec{x} : x^1 = 0\}$ under one of these homeomorphism is called a boundary point. As a corollary to Brower's 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 $\mathrm{int}(M) := M - \partial M$.

Topological manifolds, as we have defined them, are paracompact and also "normal" which means that given any pair of disjoint closed sets $F_1, F_2 \subset M$ there are open sets U_1 and U_2 containing F_1 and F_2 respectively such that U_1 and U_2 are also disjoint. The property of being paracompact may be combined with normality to show that topological manifolds support C^0-partitions of unity: Given any cover of M by open sets $\{U_\alpha\}$ there is a family of continuous functions $\{\beta_i\}$ called a C^0-partition of unity whose domains form a cover of M such that

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

(ii) each $p \in M$ has a neighborhood that intersects 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 8.3 *For differentiable manifolds we will be much more interested in the existence of $C^\infty-$partitions of unity. Finite dimensional differentiable manifolds always support smooth partitions of unity. This has been called smooth paracompactness.*

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

8.2 Charts, Atlases and Smooth Structures

Definition 8.4 *Let M be a Hausdorff topological space. A* ***chart*** *on M is a homeomorphism of an open subset $U \subset M$ onto an open subset of a finite dimensional normed space E. We say that the chart takes values in E. A chart $\mathtt{x} : U \to \mathtt{x}(U) \subset \mathsf{E}$ is traditionally indicated by the pair $(U, \mathtt{x})$ and the pair itself is also called a chart. The space E is called the model space.*

Convention(questionthisman): From now on in this book all topological spaces will be assumed to be Hausdorff unless otherwise stated.

Definition 8.5 *Let $\mathcal{A} = \{(U_\alpha, \mathtt{x}_\alpha)\}_{\alpha \in A}$ be a collection of charts on a topological space M each member of which takes values in a fixed model space E. We call $\mathcal{A}$ an* ***atlas of class*** *C^r $(0 \leq r \leq \infty)$ if the following conditions are satisfied:*
i) $\cup_{\alpha \in A} U_\alpha = M$
ii) Whenever $U_\alpha \cap U_\beta$ is not empty then the map

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

is a C^r diffeomorphism.

An **atlas of class** C^r is also called a C^r **atlas**. It is our convention that every chart in an specific atlas takes values in a single model space E. If we need to specify the model space we call the atlas an **E-valued atlas**.

Definition 8.6 *Two C^r-atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ on M are equivalent if $\mathcal{A}_1 \cup \mathcal{A}_2$ is also a C^r-atlas. A C^r differentiable structure on M is an equivalence class of C^r-atlases.*

The union of the C^r atlases in an equivalence class is the unique maximal C^r atlas in the class. The set of equivalence classes of C^r atlases (differentiable structures) is in 1-1 corresponence with the set of maximal C^r atlases. Thus an alternative way to define a differentiable structure is as a maximal atlas. Every atlas is contained in a unique maximal atlas and so as soon as we have an atlas we we have a determined differentiable structure. One also calls a C^r differentiable structure a "smooth structure" especially in the case $r = \infty$.

Some authors start out assuming only that M is a set and that charts are bijections rather than homeomorphisms. The definition of an atlas then has an extra condition. Namely, that sets of the form $\mathtt{x}_\alpha (U_\alpha \cap U_\beta)$ are always open (automatically true in our case). Once this extra condition is added it is an exercise in point set topology to show that there is a unique topology induced on the set M which makes the sets U_α open and all the charts maps homeomorphisms.

Now we could define a C^r manifold as a topological space together with C^r structure but is more common to include some assumptions about the topology of M in the definition. The problem is that there is no universal agreement as to how many and what type of topological assumptions are appropriate.

Definition 8.7 *A **differentiable manifold of class** C^r is a space M together with a specified C^r differentiable structure on M. The **dimension** of M is by definition the dimension of the model space E.*

The model space is traditional taken to be one of the coordinate spaces $\mathbb{R}^n$ and in fact by choosing a basis for E we can easily compose every chart with the resulting isomorphism with $\mathbb{R}^n$ (where $\dim \mathsf{E}$) and in so doing obtain an $\mathbb{R}^n$. For this reason we may assume whenever convenient that $\mathsf{E} = \mathbb{R}^n$. If we were to take this definition seriously, then a differentiable manifold would be a pair $(M, \mathcal{A})$ where $\mathcal{A}$ is a maximal atlas. However, we follow the tradition of refering to M itself as the differentiable manifold. To say that a space M is a differentiable manifold is just to say that it supports a C^r differentiable structure and that a such a differentiable structure has been specified. The very important point is that a single topological space my support two radically different differentiable structures. More on this below.

Recall that any atlas determines a unique C^r-differentiable structure on M since it determines 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 other words, for a given manifold, we just need some atlas which we may enlarge with compatible charts as needed. For example, the space $\mathbb{R}^n$ is itself a C^∞ manifold (and hence a C^r-manifold for any $r \geq 0$) since we can take for an atlas for $\mathbb{R}^n$ the single chart $(\mathrm{id}, \mathbb{R}^n)$ where $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is just the identity map $\mathrm{id}(x) = x$. 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 the technical idea of a maximal atlas usually fades into the background. For example, once we have the atlas $\{(\mathrm{id}, \mathbb{R}^2)\}$ 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 latter atlas and so we could throw this chart into the atlas and "fatten it up" a bit. Obviously, there are many more charts that could be thrown into the mix if needed because in this case any local diffeomorphism $U \subset \mathbb{R}^2 \to \mathbb{R}^2$ would be compatible with the "identity" chart $(\mathrm{id}, \mathbb{R}^2)$ 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 $(\mathrm{id}, \mathbb{R}^1)$ 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 $\{(\mathrm{id}, \mathbb{R}^1)\}$ and so they determine different maximal atlases. The problem is that the inverse of $x \mapsto x^3$ is not differentiable (in the usual sense) at the origin. Now we have two different differentiable structures on the line $\mathbb{R}^1$. Actually, the two atlases are equivalent in another sense that we will make precise below (they are diffeomorphic). On the other hand, it is a deep result proved fairly recently that there exist infinitely many truly different non-diffeomorphic differentiable structures on $\mathbb{R}^4$. The reader ought to be wondering what is so special about dimension four. The next example generalizes these observations a bit:

Example 8.8 *For each positive integer n, the space $\mathbb{R}^n$ is a differentiable manifold in a simple way. Namely, there is a single chart that forms an atlas*[2] *which consists of the identity map $\mathbb{R}^n \to \mathbb{R}^n$. The resulting differentiable structure is called the standard differentiable structure on $\mathbb{R}^n$. 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 but not compatible with standard structure. Thus we seem to have two different differentiable structures and hence two different differentiable 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 8.34 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 $\mathrm{x} : \mathrm{v} \mapsto (f_1^*\mathrm{v}, ..., f_n^*\mathrm{v}) \in \mathbb{R}^n$. On the other hand V may as well be modeled (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 8.9 *The sphere $S^2 \subset \mathbb{R}^3$. Choose a pair of antipodal points such as north and south poles where $z = 1$ and -1 respectively. 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 can provide the $n-$sphere $S^n \subset \mathbb{R}^{n+1}$ with a differentiable structure using two charts (U^+, ψ^+) and (U^-, ψ^-). Here,*

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

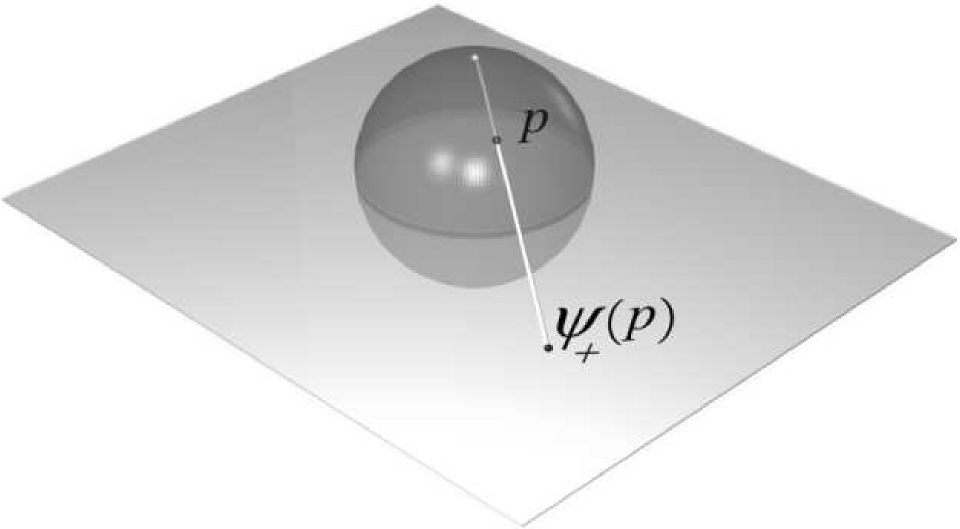
and ψ^+ (resp. ψ_-) is stereographic projection from the north pole $(0, 0....0, 1)$ (resp. south pole $(0, 0, ..., 0, -1)$). Explicitly we have

$$\psi^+(\vec{x}) = \frac{1}{(1 - x_{n+1})}(x_1, ..., x_n) \in \mathbb{R}^n$$
$$\psi_-(\vec{x}) = \frac{1}{(1 + x_{n+1})}(x_1, ..., x_n) \in \mathbb{R}^n$$

Exercise 8.10 *Compute $\psi_+ \circ \psi_-^{-1}$ and $\psi_-^{-1} \circ \psi_+$.*

Example 8.11 *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 $\ell \in P_2(\mathbb{R})$*

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



*not contained in the x, y plane. Every line $\ell \in U_z$ 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 $\ell \mapsto (x(\ell), y(\ell))$. 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). These charts cover $P_2(\mathbb{R})$ and so we have an atlas. 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*

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

Example 8.12 *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 shall refer to such an open subset $U \subset M$ with this differentiable structure as an **open submanifold** of M.*

Example 8.13 *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 that 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 topology is the relative topology of S in $\mathbb{R}^{n+1}$.

Example 8.14 *The set of all $m \times n$ matrices $\mathbb{M}_{m\times n}$ (also written $\mathbb{R}^m_n$) is an $mn-$manifold modeled on $\mathbb{R}^{mn}$. We only need one chart again since it is clear that $\mathbb{M}_{m\times n}$ is in natural one to one 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 $GL(n, \mathbb{R})$ 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 of dimensions n_1 and n_2 respectively, we can form the topological Cartesian product $M_1 \times M_2$. We may give $M_1 \times M_2$ a differentiable structure 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 from $\mathcal{A}_{M_1}$ 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. With this product differentiable structure, $M_1 \times M_2$ is called a **product manifold**.

Example 8.15 *The circle is clearly a $C^\infty-$manifold and hence so is the product $T = S^1 \times S^1$ which is a torus.*

Example 8.16 *For any manifold M we can construct the "cylinder" $M \times I$ where I is some open interval in $\mathbb{R}$.*

Exercise 8.17 *$P(\mathbb{R}^2)$ is called the projective plane. Let V_z be the x, y plane and let V_x be the y, z plane. Let $U_{\mathrm{V}_x}, \mathtt{x}_x$ and $U_{\mathrm{V}_z}, \mathtt{x}_z$ be the corresponding coordinate charts. Find $\mathtt{x}_z \circ \mathtt{x}_x^{-1}$.*

8.3 Pseudo-Groups and Model Spaces*

Without much work we can generalize our definitions in such a way as to provide, as special cases, the definitions of some common notions such as that of *complex manifold* and *manifold with boundary.* If fact, we will also take this opportunity to include infinite dimensional manifolds. An infinite dimensional manifold is modeled on an infinite dimensional Banach space.. It is quite important for our purposes to realize that the spaces (so far just $\mathbb{R}^n$) that 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 modeled on $\mathbb{R}^n$ where on the latter space we single out the local C^r-diffeomorphisms between open sets. But we will also study complex manifolds, foliated manifolds, manifolds with boundary, Hilbert manifolds and so on. Thus we need appropriate sets or spaces but also, significantly, we need a distinguished family of maps on the space. In this context the following notion becomes useful:

Definition 8.18 *A **pseudogroup of transformations**, say $\mathcal{G}$, of a topological space X is a family $\{\Phi_\gamma\}_{\gamma\in\Gamma}$ of homeomorphisms with domain U_γ and range V_γ, both open subsets of X, that satisfies the following properties:*

1) $\mathrm{id}_X \in \mathcal{G}$.

2) For all $\Phi_\gamma \in \mathcal{G}$ and open $U \subset U_\gamma$ the restrictions $\Phi_\gamma|_U$ are in $\mathcal{G}$.

3) $f_\gamma \in \mathcal{G}$ implies $f_\gamma^{-1} \in \mathcal{G}$.

4) The composition of elements of $\mathcal{G}$ are elements of $\mathcal{G}$ whenever the composition is defined with nonempty domain.

5) For any subfamily $\{\Phi_\gamma\}_{\gamma\in G_1} \subset \mathcal{G}$ 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 $\mathcal{G}$ if it is a homeomorphism.

Definition 8.19 *A **sub-pseudogroup** $\mathcal{S}$ of a pseudogroup is a subset of $\mathcal{G}$ that is also a pseudogroup (and so closed under composition and inverses).*

We will be mainly interested in what we shall refer to as C^r-pseudogroups and the spaces that support them. Our main example will be the set $\mathcal{G}^r_{\mathbb{R}^n}$ of all C^r maps between open subsets of $\mathbb{R}^n$. More generally, for a Banach space B we have the C^r-pseudogroup $\mathcal{G}^r_{\mathsf{B}}$ consisting of all C^r maps between open subsets of a Banach space B. Since this is our prototype the reader should spend some time thinking about this example.

Definition 8.20 *A C^r- pseudogroup of transformations of a subset M of Banach space B is a defined to be pseudogroup which results from restricting a sub-pseudogroup of $\mathcal{G}^r_{\mathsf{B}}$ to the subset M. The set M with the relative topology and this C^r- pseudogroup is called a **model space** .*

Example 8.21 *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 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 referred to in the definition is just $\mathbb{C}^n$ itself.*

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 8.22 *Let $\lambda \in \mathsf{M}^*$ be a continuous linear functional on a Banach space M. In the case of $\mathbb{R}^n$ it will be enough to consider projection onto the first coordinate x^1. Now let $\mathsf{M}_\lambda^+ = \{x \in \mathsf{M}\colon \lambda(x) \geq 0\}$ and $\mathsf{M}_\lambda^- = \{x \in \mathsf{M}\colon \lambda(x) \leq 0\}$ and also let $\partial\mathsf{M}_\lambda^+ = \partial\mathsf{M}_\lambda^- = \{x \in \mathsf{M}\colon \lambda(x) = 0\}$ (the kernel of λ). Clearly M_λ^+ and M_λ^- are homeomorphic and $\partial\mathsf{M}_\lambda^-$ is a closed subspace.* [3] *The spaces M_λ^+ and M_λ^- are referred to as* ***half spaces****.*

Example 8.23 *Let $\mathcal{G}^r_{\mathsf{M}_\lambda^-}$ be the restriction to M_λ^- of the set of C^r-diffeomorphisms ϕ from open subsets of M to open subsets of M that have the following property:*

**) If the domain U of $\phi \in \mathcal{G}^r_{\mathsf{M}}$ has nonempty intersection with $\mathsf{M}_0 := \{x \in \mathsf{M}\colon \lambda(x) = 0\}$ then $\phi|_{\mathsf{M}_\lambda^- \cap U}\left(\mathsf{M}_\lambda^- \cap U\right) \subset \mathsf{M}_\lambda^- \cap U$ and $\phi|_{\mathsf{M}_0 \cap U}(\mathsf{M}_0 \cap U) \subset \mathsf{M}_0 \cap U$.*

Notation 8.24 *It will be convenient to 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 modeled 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 modeled on a space of the form $\mathsf{M} \times \mathsf{E}$. This will be explained in detail when we study vector bundles.*

Let us now '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 of a fixed model space M.

Definition 8.25 *Let $\mathcal{G}$ be a C^r-pseudogroup of transformations on a model space M. A $\mathcal{G}$-atlas for a topological space M is a family of charts $\{\mathtt{x}_\alpha, U_\alpha\}_{\alpha \in A}$ (where A is just an indexing set) that 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 $\mathcal{G}$.

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 definitions, the model space M is a subset of a Banach space. If M is the whole Banach space (the most common situation) and if $\mathcal{G} = \mathcal{G}^r_{\mathsf{M}}$ (the whole pseudogroup of local C^rdiffeomorphisms) then, generalizing our former definition a bit, we call the atlas a C^r atlas.

Exercise 8.26 *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 $\mathbb{R}^n$.*

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

In practice, a $\mathcal{G}-$manifold is just a space M (soon to be a topological manifold) together with a $\mathcal{G}-$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 $\mathcal{G}^r$. 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 C^r 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 $\mathcal{G}^r-$**structure** on M.

As before, 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 and every member of the equivalence class determines the maximal atlas –just toss in every possible compatible chart and we end up with the maximal atlas again. Thus the current definition of differentiable structure is equivalent to our previous definition.

Just as before, 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 demands we have ended with the same notion of differentiable manifold. To see how this approach would go the reader should consult the excellent book [?].

Now from the vantage point of this general notion of model space we get a slick definition of manifold with boundary.

Definition 8.27 *A set M is called a C^r-**differentiable manifold with boundary** (or just C^r manifold with boundary) if it comes equipped with a $\mathcal{G}^r_{\mathsf{M}^-_\lambda}-$structure. M is given the topology induced by the maximal atlas for the given $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ structure.*

Whenever we speak of a C^r-manifold with boundary we will have a fixed $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ structure and therefore a maximal $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ atlas $\mathcal{A}_M$ in mind. A chart from $\mathcal{A}_M$ will be called an **admissible chart**.

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 usually assume) then the manifold is said to be a **pure manifold** and is said to be **modeled on** M.

Remark 8.28 *In the case of* $\mathsf{M} = \mathbb{R}^n$ *the chart maps* $\mathtt{x}$ *are maps into* $\mathbb{R}^n$ *and so projecting to each factor we have that* $\mathtt{x}$ *is comprised of* n *functions* x^i *and we write* $\mathtt{x} = (x^1, ..., x^n)$*. Because of this we sometimes talk about "x-coordinates versus y-coordinates" and so forth. Also, we use several rather self explanatory expressions such as "**coordinates", "coordinate charts**", "**coordinate systems"** and so on and these are all used to refer roughly to same thing as "chart" as we have defined the term. A chart* $\mathtt{x}, U$ *on* M *is said to be* ***centered at*** p *if* $\mathtt{x}(p) = 0 \in \mathsf{M}$.

Example 8.29 *Each Banach space* M *is a differentiable manifold in a trivial way. Namely, there is a single chart that forms an atlas*[4] *which is just the identity map* $\mathsf{M} \to \mathsf{M}$*. In particular* $\mathbb{R}^n$ *with the usual coordinates is a smooth manifold.*

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 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 modeled 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 $M_1 \times M_2$. The definition of a product proceeds just as in the finite dimensional case: 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

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

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.

It should be clear from the context that M_1 and M_2 are modeled 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.

8.4 Smooth Maps and Diffeomorphisms

Functions.

A function defined on a manifold or on some open subset is differentiable by definition if it appears differentiable in every coordinate system that intersects

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

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 8.30 *Let M be a C^r-manifold modeled on the Banach space M (usually $\mathsf{M} = \mathbb{R}^n$ for some n). 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 if and only if for every admissible chart $U, \mathtt{x}$ with $U \cap O \neq \emptyset$ the function*

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

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 we have

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

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 8.31 *We have seen that when we compose various maps as above the domain of the result will in general be an open set that 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 8.32 *Let M and N be C^r manifolds with corresponding maximal atlases $\mathcal{A}_M$ and $\mathcal{A}_N$ and modeled on M and F respectively. A 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*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: \mathsf{M} \to \mathsf{F}$$

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

Exercise 8.33 *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 modeled on M and F respectively. A (pointed) map $f : (M, p) \to (N, q)$ is said to be r **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

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: (\mathsf{M}, \mathtt{x}(p)) \to (\mathsf{F}, \mathtt{y}(q))$$

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

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

Definition 8.34 *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 C^r diffeomorphisms of a manifold M onto itself is denoted* $\operatorname{Diff}^r(M)$. *In case $r = \infty$ we simply write* $\operatorname{Diff}(M)$.

We will use the convention that $\operatorname{Diff}^0(M)$ denotes the group of homeomorphisms of M onto itself.

Example 8.35 *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 8.36 *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 8.37 *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 if and only if there is a C^r diffeomorphism $f : M \to N$.*

Recall the we have pointed out that we can put more than one differentiable structure on $\mathbb{R}$ by using the function $x^{1/3}$ as a chart. This generalizes in the obvious way: The map $\varepsilon : (x^1, x^2, ..., x^n) \mapsto ((x^1)^{1/3}, x^2, ..., x^n)$ is a chart for $\mathbb{R}^n$ but not compatible with the standard (identity) 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. They are different. But they are equivalent in another sense. Namely, they are diffeomorphic via the map ε. So 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 pairwise non-diffeomorphic structures (a very deep and difficult result) but $\mathbb{R}^k$ for $k \geq 5$ has only one diffeomorphism class.

Definition 8.38 *A map $f : M \to N$ is called a local diffeomorphism if and only if 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.*

Example 8.39 *The map $\pi : S^2 \to P(\mathbb{R}^2)$ 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 8.40 *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 $\mathbb{R}^2 \to \mathbb{R}^2$. The fact that we automatically get a diffeomorphism here follows from a moderately hard theorem proved later in the book.

Example 8.41 *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 that are defined on subsets $S \subset M$ that are not necessarily open. We shall call such a map f smooth (resp. C^r) if there is an open set O containing S and map $\widetilde{f}$ that is smooth (resp. C^r) on O and such that $\left.\widetilde{f}\right|_S = f$. 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 8.42 *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 $0 \le i \le k-1$.

8.5 Local expressions

Many authors seem to be over zealous and overly pedantic when it comes to the notation used for calculations in a coordinate chart. We will then make some simplifying conventions that are exactly what every student at the level of advanced calculus is already using anyway. Consider an arbitrary pair of charts $\mathtt{x}$ and $\mathtt{y}$ 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

$$y^i(p) = y^i(x^1(p), ..., x^n(p)) \tag{8.1}$$

which makes sense but in the literature we also see

$$y^i = y^i(x^1, ..., x^n). \tag{8.2}$$

In this last expression one might wonder if the x^i are functions or numbers. But this ambiguity is sort of purposeful for if 8.1 is true for all $p \in U \cap V$

then 8.2 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 ambiguous notational is harder to pull of in the case of infinite dimensional manifolds. We will instead write two different expressions in which the lettering and fonts are intended to be at least reminiscent of the classical notation:

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

and

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

In the first case, $\mathtt{x}$ and $\mathtt{y}$ are functions on $U \cap V$ while in the second case, x and y are elements of $\mathtt{x}(U \cap V)$ and $\mathtt{y}(U \cap V)$ respectively[5]. 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:

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

and

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

which should be compared with the classical expressions

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

and

$$w^i = \frac{\partial y^i}{\partial x^k} v^k.$$

8.6 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 $C^\infty-$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 $\mathbb{R}^n \times \mathbb{R}^n$

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

[5] Notice the font differences.

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 8.43 (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 $(p, v, \alpha) \sim (q, w, \beta)$ if and only if $p = q$ and*

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

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. The tangent vectors at a point p are 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 and is called the tangent space at p. 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 $\mathtt{y}, V$ be two charts containing p in their 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 writing

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

Recall, that a manifold with boundary is modeled on a half space $\mathsf{M}_\lambda^- := \{x \in \mathsf{M} : \lambda(x) \leq 0\}$ where M is some Banach space and λ is a continuous linear functional on M. The usual case is where $\mathsf{M} = \mathbb{R}^n$ for some n and with our conventions λ is then taken to be the first coordinate function x^1. If M is a manifold with boundary, the tangent bundle TM is defined as before. For instance, even if $p \in \partial M$ the fiber T_pM may still be thought of as consisting of equivalence classes where $(p, v, \alpha) \sim (p, w, \beta)$ if and only if $D(\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}_\alpha(p)} \cdot v = w$. Notice that for a given chart $\mathtt{x}_\alpha$, the vectors v in the various (p, v, α) still run throughout M and so T_pM still has tangent vectors "pointing in all directions".

On the other hand, if $p \in \partial M$ then for any half-space chart $\mathtt{x}_\alpha : U_\alpha \to \mathsf{M}_\lambda^-$ with p in its domain, $T\mathtt{x}_\alpha^{-1}(\partial \mathsf{M}_\lambda^-)$ is a subspace of T_pM. This is the subspace of vectors tangent to the boundary and is identified with the tangent space to ∂M (also a manifold as well shall see) $T_p\partial M$.

Exercise 8.44 *Show that this subspace does not depend on the choice of* $\mathtt{x}_\alpha$.

Definition 8.45 *Let M be a manifold with boundary and suppose that $p \in \partial M$. A tangent vector $v = [(p, v, \alpha)] \in T_pM$ is said to be outward pointing if $\lambda(v) > 0$ and inward pointing if $\lambda(v) < 0$. Here $\alpha \in A$ indexes charts as before; $\mathtt{x}_\alpha : U_\alpha \to \mathsf{M}_\lambda^-$.*

Exercise 8.46 *Show that the above definition is independent of the choice of the half-space chart $\mathtt{x}_\alpha : U_\alpha \to \mathsf{M}_\lambda^-$.*

Definition 8.47 (Tangent vectors via curves) *Let p be a point in a C^r manifold with $r > 1$. Suppose that we have C^r curves c_1 and c_2 mapping into the 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 tangent vectors at $p \in M$. The **tangent bundle** TM is the disjoint union of all the tangent spaces for all points in M.*

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

Remark 8.48 (Notation) *Let $X_p \in T_pM$ for p in the domain of an admissible chart $(U_\alpha, \mathtt{x}_\alpha)$. Under our first definition, 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 8.47.*

For the next definition of tangent vector we need to think about the set of real valued functions defined near a some fixed point p. 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 (just pick representatives for the germs of f and g, take restrictions of these to a common domain, multiply and then take the germ of the result). This makes $\mathcal{F}_p$ a ring (and an algebra over the field $\mathbb{R}$). Furthermore, if f is a representative for the equivalence class $\breve{f} \in \mathcal{F}_p$ then we can unambiguously define the value of $\breve{f}$ at p by $\breve{f}(p) = f(p)$. Thus we have an **evaluation map** $ev_p : F_p \to \mathbb{R}$. 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 could put the word "germ" in parentheses.

Remark 8.49 *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 8.50 *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*

$$df(p) \cdot X_p = \left.\frac{d}{dt}\right|_{t=0} f \circ c \in \mathbb{R}.$$

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 $\left.df\right|_p$. More generally, if $f :: M \to \mathsf{E}$ for some Banach space E then $df(p) : T_pM \to \mathsf{E}$ is defined by the same formula.

Remark 8.51 (Very useful notation) *This use of the "differential" notation for maps into vector spaces is useful for coordinates expressions. Let $p \in U$ where x, U is a chart and consider again a tangent vector X_p at p. Then the local representative of X_p (or principal part of X_p) in this chart is exactly $d\mathsf{x}(X_p)$.*

Definition 8.52 *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 8.53 *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 8.54 *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 8.55 *$\mathcal{D}_{X_p}$ is a derivation of the algebra $\mathcal{F}_p$. That is $\mathcal{D}_{X_p}$ is $\mathbb{R}-$linear and 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 $\left.\frac{\partial}{\partial x^i}\right|_{x_0} : f \mapsto \frac{\partial f}{\partial x^i}(x_0)$. 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 $x_0 \in \mathbb{R}^n$. 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 we can and will abbreviate $\mathcal{D}_{X_p} f = X_p f$ and actually *define* tangent vectors to be derivations. For this we need a couple of lemmas:

Lemma 8.56 *If c is (the germ of) a constant function then $\mathcal{D}c = 0$.*

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

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

and so $\mathcal{D}1 = 0$. ■

Lemma 8.57 *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 8.58 *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

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

and use 8.56 to get the formula of the proposition. The rest is easy but it is important to note that $a^i(x)$ is in the domain of $\mathcal{D}$ and so must be C^∞ near x_0. In fact, a careful examination of the situation reveals that we need to be in the C^∞ category for this to work and so on a C^r manifold for $r < \infty$, the space of all derivations of germs of C^r functions is not the same as the tangent space. ■

An important point is that the above construction carries over via charts to manifolds. 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 an isomorphism between $\mathsf{Der}(\mathcal{F}_p)$ and $\mathsf{Der}(\mathcal{F}_{\mathtt{x}(p)})$ by the following simple rule:

$$\mathcal{D}_{\mathtt{x}(p)} \mapsto \mathcal{D}_p$$
$$\text{where } \mathcal{D}_p f = \mathcal{D}_{\mathtt{x}(p)}(f \circ \mathtt{x}^{-1}).$$

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 finite dimensional C^∞ manifolds:

Definition 8.59 (Tangent vectors as derivations) *Let M be a C^∞ manifold of dimension $n < \infty$. Consider the set of all (germs of) C^∞ functions $\mathcal{F}_p$ at $p \in M$. A **tangent vector** at p is a linear map $X_p : \mathcal{F}_p \to \mathbb{R}$ that is also a derivation in the sense that for $f, g \in \mathcal{F}_p$*

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

Once again, the tangent space at p is the set of all tangent vectors at p and the tangent bundle is defined as a disjoint union of tangent spaces 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 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 8.47 or in definition 8.43. At least we have not shown that this cannot happen.

8.7 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. We may also obtain a tangent vector in the sense of definition8.43 by letting X_p be associated to the triple (p, v, α) where $v^i := \left.\frac{d}{dt}\right|_{t=0} x^i_\alpha \circ c$ for a chart $\mathtt{x}_\alpha, U_\alpha$ with $p \in U_\alpha$.

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 a la definition 8.43, is represented by the triple (p, v, α) where $v = (v^1, ...v^n)$ is given by
$$v^i = X_p x^i \text{ (acting as a derivation)}$$

3. Suppose that, a la definition 8.43, a vector X_p at $p \in M$ is represented by (p, v, α) where $v \in \mathsf{M}$ and α names the chart $\mathtt{x}_\alpha, U_\alpha$. We obtain a derivation by defining
$$X_p f = D(f \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}_\alpha(p)} \cdot v$$
In case the manifold is modeled on $\mathbb{R}^n$ we have the more traditional notation
$$X_p f = \sum v^i \left.\frac{\partial}{\partial x^i}\right|_p f.$$
for $v = (v^1, ...v^n)$.

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

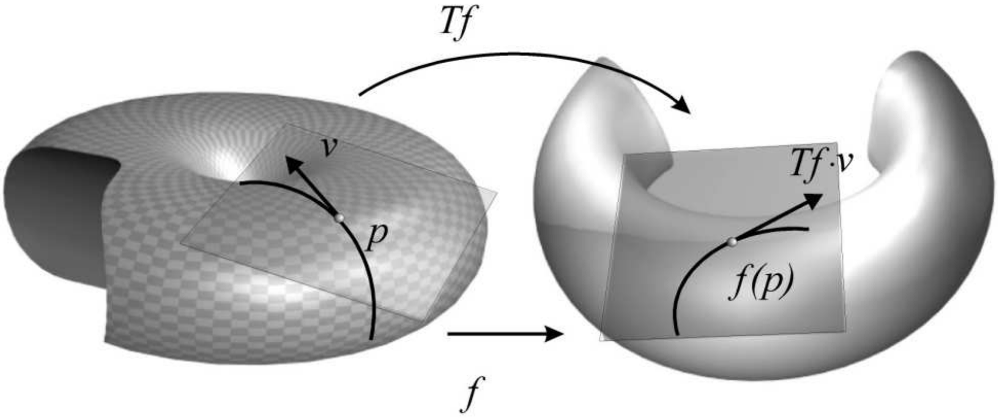
Definition 8.60 *For a chart $\mathtt{x} = (x^1, ..., x^n)$ with domain U containing a point p we define a tangent vector $\left.\frac{\partial}{\partial x^i}\right|_p \in T_pM$ by*
$$\left.\frac{\partial}{\partial x^i}\right|_p f = D_i(f \circ \mathtt{x}^{-1})(\mathtt{x}(p))$$
Alternatively, we may 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 $(p, \mathsf{e}_i, \alpha)$ where 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 8.61 *For a finite dimensional $C^\infty-$manifold M and $p \in M$, let $\left(\mathtt{x}_\alpha = (x^1, ..., x^n), U_\alpha\right)$ be a chart whose domain contains p. Show that the vectors $\left.\frac{\partial}{\partial x^1}\right|_p, ..., \left.\frac{\partial}{\partial x^n}\right|_p$ (using our last definition of tangent vector) are a basis for the tangent space T_pM.*

8.8 The Tangent Map

The first definition given below of the tangent map of a smooth 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 $T_pf : T_pM \to T_{f(p)}N$



Definition 8.62 *If we have a smooth function between manifolds*

$$f : M \to N$$

*and we consider a point $p \in M$ and its image $q = f(p) \in N$. Choose any chart $(\mathtt{x}, U)$ containing p and a chart $(\mathtt{y}, V)$ containing $q = f(p)$ so that for any $v \in T_pM$ we have the representative $d\mathtt{x}(v)$ with respect to $\mathtt{x}, U$. Then the **tangent map** $T_pf : T_pM \to T_{f(p)}N$ is defined by letting the representative of $T_pf \cdot v$ in the chart $(\mathtt{y}, V)$ be given by*

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

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

Since we have several definitions of tangent vector we expect to see several equivalent definitions of the tangent map. Here is another:

Definition 8.63 (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*

$$T_pf : T_pM \to T_qN$$

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

$$T_pf \cdot v = [f \circ c] \in T_qN$$

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

Another alternative definition of tangent map that works for finite dimensional smooth manifolds is given in terms of derivations:

Definition 8.64 (Tangent Map III) *Let M be a smooth n-manifold. View tangent vectors as derivations as explained above. Then continuing our set up above and letting g be a smooth germ at $q = f(p) \in N$ we define $T_pf \cdot v$ as a derivation by*

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

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_qM. This map is yet another version of the ***tangent map*** *T_pf.*

It is easy to check that for a smooth $f : M \to \mathsf{E}$ the differential $df(p) : T_pM \to \mathsf{E}$ is the composition of the tangent map T_pf and the canonical map $T_y\mathsf{E} \to \mathsf{E}$ where $y = f(p)$. Diagrammatically we have

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

8.9 The Tangent and Cotangent Bundles

8.9.1 Tangent Bundle

We have defined the **tangent bundle** of a manifold as the disjoint union of the tangent spaces $TM = \bigsqcup_{p \in M} T_pM$. We also gave similar definition of cotangent bundle. We show in proposition 8.67 below that TM is itself a differentiable manifold but first we record the following two definitions.

Definition 8.65 *Given 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*** *or sometimes the* ***tangent map****.*

Definition 8.66 *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 set TM together with the map $\tau_M : TM \to M$ is an example of a vector bundle defined in the sequel.)*

Proposition 8.67 *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 lift Tf is smooth and the following diagram commutes.*

$$\begin{tikzcd} TM \arrow[r, "Tf"] \arrow[d, "\tau_M"'] & TN \arrow[d, "\tau_N"] \\ M \arrow[r, "f"] & N \end{tikzcd}$$

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

$$T\mathtt{x} : TU \to T\mathtt{x}(TU) \cong \mathtt{x}(U) \times \mathsf{M}$$
$$T\mathtt{x} : \xi \mapsto (\mathtt{x} \circ \tau_M(\xi), v)$$

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

$$\begin{tikzcd} TU \arrow[r, "(\mathtt{x}{,}d\mathtt{x})"] \arrow[d] & \mathtt{x}(U) \times \mathsf{M} \arrow[d] \\ U \arrow[r, "\mathtt{x}"] & \mathtt{x}(U) \end{tikzcd}$$

For a finite dimensional manifold and a chart $\left(\mathtt{x} = (x^1, ..., x^n), U\right)$ 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 8.68 *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 $(U, T\mathtt{x})$ and $(V, T\mathtt{y})$ 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 $T\mathtt{y} \circ T\mathtt{x}^{-1}$:$(x, v) \mapsto (y, w)$ where

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

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

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

and in finite dimensions the classical notation

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

This classical notation may not be logically precise but it is easy to read and understand. In any case one could perhaps write

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

Exercise 8.69 *If M is actually equal to an open subset U of a Banach space M then we defined TU to be $U \times \mathsf{M}$. How should this be reconciled with the definitions of this chapter? Show that once this reconciliation is in force the tangent bundle chart map $T\mathtt{x}$ really is the tangent map of the coordinate map $\mathtt{x}$.*

8.9.2 The Cotangent Bundle

For each $p \in M$, T_pM has a dual space T_p^*M called the cotangent space at p.

Definition 8.70 *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 $\pi_M : \bigsqcup_{p \in M} T_p^*M \to M$ to be the obvious projection taking elements in each space T_p^*M to the corresponding point p.

Let $\{U_\alpha, \mathtt{x}_\alpha\}_{\alpha\in A}$ be an atlas of admissible charts on M. Now endow T^*M with the smooth structure given by the charts

$$T^*\mathtt{x}_\alpha : T^*U_\alpha = \pi_M^{-1}(U_\alpha) \to T^*\mathtt{x}_\alpha(T^*U_\alpha) \cong \mathtt{x}_\alpha(U_\alpha) \times \mathsf{M}^*$$

where the map $T^*\mathtt{x}_\alpha$ restricts to $T_p^*M \subset T^*U_\alpha$ as the **contragradient** of $T\mathtt{x}_\alpha$:

$$T^*\mathtt{x}_\alpha := (T\mathtt{x}_\alpha^{-1})^* : T_p^*M \to (\{\mathtt{x}_\alpha(p)\} \times \mathsf{M})^* = \mathsf{M}^*$$

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 dual to $\frac{\partial}{\partial x^1}\big|_p, ..., \frac{\partial}{\partial x^n}\big|_p$. If $\theta \in T^*U_\alpha$ then we can write

$$\theta = \sum \xi_i(\theta)\; dx^i\big|_{\pi_M(\theta)}$$

for some numbers $\xi_i(\theta)$ depending on θ. In fact, we have

$$\theta(\frac{\partial}{\partial x^i}\bigg|_{\pi(\theta)}) = \sum \xi_j(\theta) dx^j(\frac{\partial}{\partial x^i}\bigg|_{\pi(\theta)}) = \sum_j \xi_j(\theta)\delta_i^j = \xi_i(\theta).$$

Thus we see that

$$\xi_i(\theta) = \theta(\frac{\partial}{\partial x^i}\bigg|_{\pi(\theta)}).$$

So if $U_\alpha, \mathtt{x}_\alpha = (x^1, ..., x^n)$ is a chart on an n-manifold M, then the natural chart $(TU_\alpha, T^*\mathtt{x}_\alpha)$ defined above is given by

$$\theta \mapsto (x^1 \circ \pi_M(\theta), ..., x^n \circ \pi_M(\theta), \xi_1(\theta), ..., \xi_n(\theta))$$

and abbreviated to $(x^1, ..., x^n, \xi_1, ..., \xi_n)$.

Suppose that $(x^1, ..., x^n, \xi_1, ..., \xi_n)$ and $(\overline{x}^1, ..., \overline{x}^n, \overline{\xi}_1, ..., \overline{\xi}_n)$ are two such charts constructed in this way from two charts on $(U_\alpha, \mathtt{x}_\alpha)$ and $(U_\beta, \mathtt{x}_\beta)$ respectively with $U_\alpha \cap U_\beta \neq \emptyset$. We are writing $\mathtt{x}_\alpha = (x^1, ..., x^n)$ and $\mathtt{x}_\beta = (\overline{x}^1, ..., \overline{x}^n)$ for notational convenience. Then $T^*U_\alpha \cap T^*U_\beta \neq \emptyset$ and on the overlap we have

$$T^*\mathtt{x}_\beta \circ T^*\mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U_\beta) \times \mathsf{M}^* \to \mathtt{x}_\beta(U_\alpha \cap U_\beta) \times \mathsf{M}^*.$$

Notation 8.71 *With $\mathtt{x}_{\beta\alpha} = \mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ the contragradient of $D\mathtt{x}_{\beta\alpha}$ at $x \in \mathtt{x}_\alpha(U_\alpha \cap U_\beta)$ is a map $D^*\mathtt{x}_{\beta\alpha}(x) : \mathsf{M}^* \to \mathsf{M}^*$ defined by*

$$D^*\mathtt{x}_{\beta\alpha}(x) \cdot v = \left(D\mathtt{x}_{\beta\alpha}^{-1}(\mathtt{x}_{\beta\alpha}(x))\right)^t \cdot v$$

With this notation we can write coordinate change maps as $(\mathtt{x}_{\beta\alpha}, D^*\mathtt{x}_{\beta\alpha})$ or to be exact

$$\left(T^*\mathtt{x}_\beta \circ T^*\mathtt{x}_\alpha^{-1}\right)(x, v) = \left(\mathtt{x}_{\beta\alpha}(x),\, D^*\mathtt{x}_{\beta\alpha}(x) \cdot v\right).$$

In case $\mathsf{M} = \mathbb{R}^n$ we write $pr_i \circ \mathtt{x}_{\beta\alpha} = \mathtt{x}^i_{\beta\alpha}$ and then

$$\overline{x}^i \circ \pi(\theta) = \mathtt{x}^i_{\beta\alpha}(x^1 \circ \pi(\theta), ..., x^n \circ \pi(\theta))$$
$$\bar{\xi}_i(\theta) = \sum_{k=1}^{n} (D\mathtt{x}^{-1}_{\beta\alpha})^k_i (x^1 \circ \pi(\theta), ..., x^n \circ \pi(\theta)) \cdot \xi_k(\theta)$$

This pedantic mess should be abbreviated. So let us write $x^i \circ \pi$ as just x^i and suppress the argument θ. Thus we may write

$$y = \mathtt{y} \circ \mathtt{x}^{-1}(x)$$
$$\bar{\xi} = \left.\frac{\partial \mathtt{x}}{\partial \mathtt{y}}\right|_{\mathtt{x}(p)} \xi$$

which may be further written in a classical style:

$$y^i = y^i(x^1, ..., x^n)$$
$$\bar{\xi}_i = \xi_k \frac{\partial x^k}{\partial y^i}.$$

8.10 Important Special Situations.

Let V be either a finite dimensional vector space or a Banach space. If a manifold under consideration is an open subset U of a the 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 $\{x\} \times \mathsf{V}$ (by we now have several different but equivalent definitions of $T_x\mathsf{V}$). Then the identifying map is just $v \mapsto (x, v)$. 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 $v \in \mathsf{V}$ and define a curve $c_v : \mathbb{R} \to U \subset \mathsf{V}$ by $c_v(t) = x + tv$. Then $T_0c_v \cdot 1 = \dot{c}_v(0) \in T_xU$. The map $v \mapsto \dot{c}_v(0)$ is then our identifying map. Depending on how one defines the tangent bundle TU, the equality $TU = U \times \mathsf{V}$ is either a definition or a natural identification. The fact that there are these various identifications and that some things have several "equivalent" definitions is somewhat of a nuisance and can be confusing 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 $f : U \subset \mathsf{V} \to \mathsf{W}$ and if $v_p = (p, v) \in T_pU$ then $T_pf \cdot v_p = T_pf \cdot (p, v) = (p, Df|_p \cdot v)$. In other words the tangent map is $(p, v) \mapsto (p, Df|_p \cdot v)$ and so one might as well just think about the ordinary derivative $Df|_p$. In fact, in the case of a vector space some authors actually identify T_pf with $Df|\ _p$ as they also identify T_pU with V. There is usually no harm in this and it actually streamlines the calculations a bit.

The identifications of $T_x\mathsf{V}$ with V and TU with $U\times\mathsf{V}$ will be called canonical identifications. Whenever we come across a sufficiently natural isomorphism, then that isomorphism could be used to identify the two spaces. We will see cases where there are several different natural isomorphisms which compete for use as identifications. This arises when a space has more than one structure.

An example of this situation is the case of a manifold of matrices such as $GL(n,\mathbb{R})$. Here $GL(n,\mathbb{R})$ 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 $T_A GL(n,\mathbb{R})$. Another interesting fact is that many important maps such as $c_Q : A \mapsto Q^t AQ$ are actually linear so with the identifications $T_A G\mathrm{L}(n,\mathbb{R}) = \mathbb{M}_{n\times n}(\mathbb{R})$ we have

$$T_A c_Q \text{ `` = '' } \left.Dc_Q\right|_A = c_Q : \mathbb{M}_{n\times n}(\mathbb{R}) \to \mathbb{M}_{n\times n}(\mathbb{R}).$$

Whenever we come across a sufficiently natural isomorphism, then that isomorphism could be used to identify the two spaces.

Definition 8.72 (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 another 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 $c(t) = (c_1(t), c_2(t))$ and $c(0) = (x,y)$ then the map $[c] \mapsto ([c_1],[c_2])$ is a natural isomorphism which we use to simply identify $[c] \in T_{(x,y)}(M_1 \times M_2)$ 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 $\iota^y : x \mapsto (x,y)$.

$$\begin{tikzcd}
(M_1,x) \arrow[r, shift right, "\iota^y"'] & (M_1\times M_2,(x,y)) \arrow[l, shift right, "pr_1"'] \arrow[r, shift left, "pr_2"] & (M_2,x) \arrow[l, shift left, "\iota_x"]
\end{tikzcd}$$

$$\begin{tikzcd}
T_x M_1 \arrow[r, shift right, "T_x\iota^y"'] & T_{(x,y)}M_1\times M_2 \arrow[l, shift right, "T_{(x,y)}pr_1"'] \arrow[r, shift left, "T_{(x,y)}pr_2"] & T_x M_2 \arrow[l, shift left, "T_y\iota_x"]
\end{tikzcd}$$

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 temporarily denote the isomorphic images of T_xM_1 and T_yM_2 in $T_{(x,y)}(M_1 \times M_2)$ under these two maps by the symbols $(T_xM)_1$ and $(T_yM)_2$. We then have the internal direct sum decomposition $(T_xM)_1 \oplus (T_yM)_2 = T_{(x,y)}(M_1 \times M_2)$ and the isomorphism

$$T\iota^y \times T\iota_x : T_xM_1 \times T_yM_2 \to (T_xM)_1 \oplus (T_yM)_2 = T_{(x,y)}(M_1 \times M_2).$$

The inverse of this isomorphism is

$$T_{(x,y)}pr_1 \times T_{(x,y)}pr_2 : T_{(x,y)}(M_1 \times M_2) \to T_xM_1 \times T_yM_2$$

which is then taken as an identification and, in fact, this is none other than the map $[c] \mapsto ([c_1],[c_2])$. Let us say a bit about the naturalness of the identification of $[c] \in T_{(x,y)}(M_1 \times M_2)$ with $([c_1],[c_2]) \in T_xM_1 \times T_yM_2$. In the smooth category there is a product operation. The essential point is that for any two manifolds M_1 and M_2 the manifold $M_1 \times M_2$ together with the two projection maps serves as the product in the technical sense that for any smooth maps $f : N \longrightarrow M_1$ and $g : N \longrightarrow M_2$ we always have the unique map $f \times g$ which makes the following diagram commute:

$$\begin{tikzcd}
& N \arrow[dl, "f"'] \arrow[d, "f\times g"] \arrow[dr, "g"] & \\
M_1 & M_1 \times M_2 \arrow[l, "pr_1"'] \arrow[r, "pr_2"] & M_2
\end{tikzcd}$$

Now for a point $x \in N$ write $p = f(x)$ and $p = g(x)$. On the tangent level we have

$$\begin{tikzcd}
& T_xN \arrow[dl, "T_xf"'] \arrow[d, "T_x(f\times g)"] \arrow[dr, "T_xg"] & \\
T_pM_1 & T_{(p,q)}(M_1 \times M_2) \arrow[l, "T_{(p,q)}pr_1"'] \arrow[r, "T_{(p,q)}pr_2"] & T_qM_2
\end{tikzcd}$$

which is a diagram in the vector space category. In the category of vector spaces the product of T_pM_1 and T_pM_2 is $T_pM_1 \times T_pM_2$ (outer direct sum) together with the projections onto the two factors. It is then quite reassuring to notice that under the identification introduced above this diagram corresponds to

$$\begin{tikzcd}
& T_xN \arrow[dl, "T_xf"'] \arrow[d, "T_xf\times T_xg"] \arrow[dr, "T_xg"] & \\
T_pM_1 & T_pM_1 \times T_qM_2 \arrow[l, "Pr_1"'] \arrow[r, "Pr_2"] & T_qM_2
\end{tikzcd}$$

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

Proving this last lemma is much easier and more instructive than reading the proof so we leave it to the reader in good conscience.

The following diagram commutes:

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

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

8.11 Vector fields and Differential 1-forms

Definition 8.74 *A smooth **vector field** is a smooth map $X : M \to TM$ such that $X(p) \in T_pM$ for all $p \in M$. In other words, a vector field on M is a smooth section of the tangent bundle $\tau_M : TM \to M$. We often write $X(p) = X_p$.*

The map X being smooth is equivalent to the requirement that the function $Xf : M \to \mathbb{R}$ given by $p \mapsto X_p f$ 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 $x \mapsto (x, X_U(x))$ where the **local representative** (or **principal part**) X_U is given by projecting $T\mathtt{x} \circ X \circ \mathtt{x}^{-1}$ onto the second factor in $T\mathsf{E} = \mathsf{E} \times \mathsf{E}$:

$$\begin{aligned}
x &\mapsto \mathtt{x}^{-1}(x) = p \mapsto X(p) \mapsto T\mathtt{x} \cdot X(p) \\
&= (\mathtt{x}(p), X_U(\mathtt{x}(p))) = (x, X_U(x)) \mapsto X_U(x)
\end{aligned}$$

In finite dimensions one can write $X_U(x) = (v_1(x), ..., v_n(x))$.

Notation 8.75 *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.*

A smooth (resp C^r) section of the cotangent bundle is called a smooth (resp C^r) **covector field** or also a smooth (resp C^r) **1-form** . The set of all C^r 1-forms is denoted by $\mathfrak{X}^{r*}(M)$ and the smooth 1-forms are denoted by $\mathfrak{X}^*(M)$. For any open set $U \subset M$, the set $C^\infty(U)$ of smooth functions defined 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 $\mathcal{C}^\infty(U)$.

Remark 8.76 *The assignment $U \mapsto \mathfrak{X}_M(U)$ is a presheaf of modules over $\mathcal{C}^\infty$ and $\mathfrak{X}^*(M)$ is a module over the ring of functions $C^\infty(M)$ and a similar statement holds for the C^r case. (Jump forward to sections* **??** *and* **??** *for definitions.)*

Definition 8.77 *Let $f : M \to \mathbb{R}$ be a C^r 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 8.50. is a 1-form called the differential of f.*

If $U, \mathtt{x}$ is a chart on M then we also have the following familiar looking formula in the finite dimensional case

$$df = \sum \frac{\partial f}{\partial x^i} dx^i$$

which is interpreted to mean that at each $p \in U_\alpha$ we have

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

In general, if we have a chart $U, \mathtt{x}$ on a possibly infinite dimensional manifold then we may write

$$df = \frac{\partial f}{\partial \mathtt{x}} d\mathtt{x}$$

We have really seen this before. All that has happened new is that p is allowed to vary so we have a field.

There is a slightly different way to view a 1-form that is often useful. Namely, we may think of $\alpha \in \mathfrak{X}^*(M)$ as a map $TM \to \mathbb{R}$ given simply by $\alpha(v) = \alpha(p)(v) = \alpha_p(v)$ whenever $v \in T_pM \subset TM$.

If $\phi : M \to N$ is a C^∞ map and $f : N \to \mathbb{R}$ a C^∞ function we define the **pullback** of f by ϕ as

$$\phi^* f = f \circ \phi$$

and the **pullback** of a 1-form $\alpha \in \mathfrak{X}^*(N)$ by $\phi^*\alpha = \alpha \circ T\phi$. To get a clearer picture of what is going on we could view things at a point and then we have $\phi^*\alpha|_p \cdot v = \alpha|_{\phi(p)} \cdot (T_p\phi \cdot v)$.

Next we describe the local expression for the pull-back of a 1-form. Let $U, \mathtt{x}$ be a chart on M and $V, \mathtt{y}$ be a coordinate chart on N with $\phi(U) \subset V$. A typical 1-form has a local expression on V of the form $\alpha = \sum a_i dy^i$ for $a_i \in C^\infty(V)$. The local expression for $\phi^*\alpha$ on U is $\phi^*\alpha = \sum a_i \circ \phi \, d\left(y^i \circ \phi\right) = \sum a_i \circ \phi \frac{\partial(y^i \circ \phi)}{\partial x^i} dx^i$.

The pull-back of a function or 1-form is defined whether $\phi : M \to N$ 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 8.78 *Let $\phi : M \to N$ be a C^∞ diffeomorphism with $r \geq 1$. The* ***push-forward*** *of a function $f \in C^\infty(M)$ is denoted $\phi_* f$ and defined by $\phi_* f(p) := f(\phi^{-1}(p))$. We can also define the push-forward of a 1-form as $\phi_*\alpha = \alpha \circ T\phi^{-1}$.*

Exercise 8.79 *Find the local expression for $\phi_* f$. Explain why we need ϕ to be a diffeomorphism.*

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 8.80 *The differential is natural with respect to pullback. In other words, if $\phi : N \to M$ is a C^∞ map and $f : M \to \mathbb{R}$ a C^∞ function with $r \geq 1$ then $d(\phi^* f) = \phi^* df$. Consequently, the 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 (i.e. identity map id_M restricted to U) then $f|_U = \iota^* f$ and $df|_U = \iota^* df$ so the statement about restrictions is just a special case. ■

Definition 8.81 *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[6] over the ring of smooth functions $\mathcal{C}^\infty(U)$.

Definition 8.82 *To a vector field X on U we associate the map $\mathcal{L}_X : C^\infty(U) \to \mathfrak{X}_M(U)$ 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 Leibniz 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 $\mathcal{L}_X\ f$.

Remark 8.83 *We often leave out parentheses and just write $Xf(p)$ instead of the more careful $(Xf)(p)$.*

In summary, we have the **derivation law** (Leibniz rule) for vector fields:

$$X(fg) = fXg + gXf.$$

or in other notation $\mathcal{L}_X\ (fg) = f\mathcal{L}_X\ g + g\mathcal{L}_X\ f$.

[6] See section **??**.

8.12 Moving frames

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 $\frac{\partial}{\partial x^i}$ form a basis at each tangent space at point in U. The set of fields $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$ is called a **holonomic moving frame** over U or, more commonly, a **coordinate frame**. 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 $dx^i : p \mapsto dx^i\big|_p$ defines a field of covectors such that $dx^1\big|_p, ..., dx^n|_p$ forms a basis of T_p^*M for each $p \in U$. These covector fields form what is called a **holonomic**[7] **coframe** or **coframe field** over U. In fact, the functions X^i are given by $X^i = dx^i(X) : p \mapsto dx^i\big|_p (X_p)$ and so we write

$$X|_U = \sum dx^i(X) \frac{\partial}{\partial x^i}.$$

Notation 8.84 *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 we can also get a family of vector (resp. covector) fields that are linearly independent at each point in their mutual domain and yet are not necessarily of the form $\frac{\partial}{\partial x^i}$ (resp. dx^i) for any coordinate chart:

Definition 8.85 *Let $E_1, E_2, ..., E_n$ be smooth vector fields defined on some open subset U of a smooth n-manifold M. If $E_1(p), E_2(p), ..., E_n(p)$ form a basis for T_pM for each $p \in U$ then we say that $E_1, E_2, ..., E_n$ is a (non-holonomic) **moving frame** or a **frame field** over U.*

If $E_1, E_2, ..., E_n$ is moving frame over $U \subset M$ and X is a vector field defined on U then we may write

$$X = \sum X^i E_i \text{ on } U$$

[7] The word holonomic comes from mechanics and just means that the frame field derives from a chart. A related fact is that $[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = 0$.

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) moving coframe field $\theta^1, ..., \theta^n$ and then $X^i = \theta^i(X)$ so

$$X = \sum \theta^i(X)E_i \text{ on } U.$$

Let us now consider an important special situation. If $M \times N$ is a product manifold and $(U, \mathtt{x})$ is a chart on M and $(V, \mathtt{y})$ is a chart on N then we have a chart $(U \times V, \mathtt{x} \times \mathtt{y})$ on $M \times N$ where the individual coordinate functions are $x^1 \circ pr_1, ..., x^m \circ pr_1$ $y^1 \circ pr_2, ..., y^n \circ pr_2$ which we temporarily denote by $\widetilde{x}^1, ..., \widetilde{x}^m, \widetilde{y}^1, ..., \widetilde{y}^n$. Now what is the relation between the coordinate frame fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ and $\frac{\partial}{\partial \widetilde{x}^i}, \frac{\partial}{\partial \widetilde{y}^i}$? Well, the latter set of $n+m$ vector fields is certainly a linearly independent set at each point $(p, q) \in U \times V$. The crucial relations are $\frac{\partial}{\partial \widetilde{x}^i} f = \frac{\partial}{\partial x^i} (f \circ pr_1)$ and $\frac{\partial}{\partial \widetilde{y}^i} = \frac{\partial}{\partial y^i} (f \circ pr_2)$.

Exercise 8.86 *Show that* $\frac{\partial}{\partial \widetilde{x}^i}{}_{(p,q)} = Tpr_1 \frac{\partial}{\partial x^i}{}_p$ *for all* q *and that* $\frac{\partial}{\partial \widetilde{y}^i} = Tpr_2 \frac{\partial}{\partial y^i}{}_q$.

Remark 8.87 *In some circumstances it is safe to abuse notation and denote* $x^i \circ pr_1$ *by* x^i *and* $y^i \circ pr_2$ *by* y^i. *Of course we then are denoting* $\frac{\partial}{\partial \widetilde{x}^i}$ *by* $\frac{\partial}{\partial x^i}$ *and so on.*

8.13 Partitions of Unity

A smooth 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 the so called "bump functions" and "cut-off 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 possibly only tool for doing this is the partition of unity. For differential geometry, what we actually need it is a *smooth* partition of unity. Banach manifolds do not always support smooth partitions of unity. We will state and prove the existence of partitions of unity for finite dimensional manifolds and refer the reader to [**?**] for the situation in Banach spaces. A bump function is basically a smooth function with support inside some prescribed open set O and that is nonzero on some prescribed set inside O. Notice that this would not in general be possible for a complex analytic function. A slightly stronger requirement is the existence of cut-off functions. A cut-off function is a bump function that is equal to unity in a prescribed region around a given point:

Definition 8.88 *A* ***spherical cut-off function*** *of class* C^r *for the nested pair of balls* $B(r, p) \subset B(R, p)$ $(R > r)$*is a* C^r *function* $\beta : M \to \mathbb{R}$ *such that* $\beta|_{\overline{B(r,p)}} \equiv 1$ *and* $\beta|_{M \backslash B(R,p)} \equiv 0$.

More generally we have the following

Definition 8.89 *Let K be a closed subset of M contained in an open subset $U \subset M$. A* ***cut-off function*** *of class C^r for a nested pair $K \subset U$ is a C^r function $\beta : M \to \mathbb{R}$ such that $\beta|_K \equiv 1$ and $\beta|_{M \backslash U} \equiv 0$.*

Definition 8.90 *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 $\overline{V} \subset U$. A manifold M is said to* ***admit spherical cut-off functions*** *if given any point $p \in M$ and any open ball U of p, there is another open ball V of p such that $\overline{V} \subset U$ and a cut-off function $\beta_{\overline{V},U}$ for the nested pair $\overline{V} \subset U$*

For most purposes, the existence of spherical cut off functions will be sufficient and in the infinite dimensional case it might be all we can get.

Definition 8.91 *Let E be a Banach space and suppose that the norm on E is smooth (resp.C^r) on the open set $\mathsf{E}\backslash\{0\}$. The we say that E is a* ***smooth*** *(resp. C^r)* ***Banach space****.*

Proposition 8.92 *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) cut-off function β for the pair $B_r \subset B_R$. Thus β is defined on all of E , 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 $\beta(x) = g(2 - |x|)$. Check that this does the job using the fact that $x \mapsto |x|$ is assumed to be smooth (resp. C^r). ■

Corollary 8.93 *If a manifold M is modeled on a smooth (resp. C^r) Banach space M (for example, if M is a finite dimensional smooth 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 $x_0 = \psi(p) \in \mathsf{M}$ for some chart ψ, U. Then the local representative $\bar{\alpha}_{x_0} = (\psi^{-1})^* \alpha_p$ can be considered a linear function on M since we have the canonical identification $\mathsf{M} \cong \{x_0\} \times \mathsf{M} = \mathsf{T}_{x_0}\mathsf{M}$. Thus we can define

$$\varphi(x) = \begin{cases} \beta(x)\bar{\alpha}_{x_0}(x) & \text{for} \quad x \in B_R(x_0) \\ 0 & \text{otherwise} \end{cases}$$

and now making sure that R is small enough that $B_R(x_0) \subset \psi(U)$ 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 $df(p) = \alpha_p$ ■

It is usually taken for granted that derivations on smooth functions are vector fields and that all C^∞ vector fields arise in this way. In fact, this not true in general. It is true however, for finite dimensional manifold. More generally, we have the following result:

Proposition 8.94 *The map from $\mathfrak{X}(M)$ to the vector space of derivations* $\mathsf{Der}(M)$ *given by $X \mapsto \mathcal{L}_X$ is a linear monomorphism if M is modeled on a C^∞ Banach space.*

Proof. The fact that the map is linear 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 8.93 $\alpha_p(X_p) = 0$ 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 $X = 0$. ■

Another very useful result is the following:

Theorem 8.95 *Let $L : \mathfrak{X}(M) \to \mathcal{C}^\infty(M)$ be a $\mathcal{C}^\infty(M)-$linear function on vector fields. If M admits (spherical?)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 p and with closure inside U. Then letting β be a function that is identically 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 8.96 *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 $X \in \mathfrak{X}(M)$.*

Definition 8.97 *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* $\mathrm{supp}(f)$.

For finite dimensional manifolds we have the following stronger result.

Lemma 8.98 (Existence of cut-off 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 $\mathbb{R}^n$ that is identically equal to 1 on K, has compact support in U and $0 \leq \beta \leq 1$.*

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

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

This is the circular cut-off that always exists for smooth Banach spaces.

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 now with various choices of radii and such that $K \subset \cup K_i$. The U_i are 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 cut-off function;

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

■

Definition 8.99 *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 be assumed to be 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 8.100 *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 8.101 *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 $U_1, U_2, ...$that 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 $\underline{V_n}$ 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 that X is the countable union of these open sets $\{O_i\}$ and each $\underline{O_{i-1}}$ 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 that 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 8.102 *A C^r partition of unity on a C^r manifold M is a collection $\{V_i, \rho_i\}$ where*
(i) $\{V_i\}$ is a locally finite cover of M;
(ii) each ρ_i is a C^r function with $\rho_i \geq 0$ and compact support contained in V_i;
(iii) 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 domains $\{U_\alpha\}$ of some atlas $\mathcal{A} = \{U_\alpha, \mathtt{x}_\alpha\}$ for 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.*

Smooth (C^r, $r > 0$) partitions of unity do not necessarily exist on a Banach space and less so for manifolds modeled on such Banach spaces. On the other hand, some Banach spaces do admit partitions of unity. It is a fact that all separable Hilbert spaces admit partitions of unity. For more information see [**?**]. We will content ourselves with showing that all finite dimensional manifolds admit smooth partitions of unity.

Notice that in theorem 8.101 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 $U_{\beta(i)}$. This is actually true for any paracompact space but we will not prove it here. Now for the existence of a smooth partition of unity we basically need the paracompactness but since we haven't proved the above statement about compact closures (shrink?ing lemma) we state the theorem in terms of second countability:

Theorem 8.103 *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 8.98 above there is a smooth cut-off function β_i with $\mathrm{supp}(\beta_i) = \overline{W}_i$. 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 functions that form the partition of unity:

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

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

8.14