Supplement for Manifolds and Differential Geometry

Jeffrey M. Lee

Department of Mathematics and Statistics, Texas Tech University, Lubbock, Texas, 79409

Current address: Department of Mathematics and Statistics, Texas Tech University, Lubbock, Texas, 79409

E-mail address: `jeffrey.lee@ttu.edu`

Contents

-1.1. Main Errata

Chapter 1

(1) Page 3. Near the end of the page "...we have a second derivative $D^2f(a) : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$" should read
"...we have a second derivative $D^2f(a) : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$"

(2) . Page 4. The is a missing exponent of 2 in the partial derivatives in the display at the top of the page. It should (of course) read as

$$u^i = \sum_{j,k} \frac{\partial^2 f^i}{\partial x^j \partial x^k}(a) v^j w^k.$$

(3) Page 19, forth line from the top at the beginning of Notation 1.44:

$$(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1}$$

should be changed to

$$(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \backslash \{0\}$$

(4) Page 23. 6th line from the top. The expression $\mathtt{y} \circ f \circ \mathtt{x}|_{\alpha}^{-1}$ should be changed to $\mathtt{y}_\beta \circ f \circ \mathtt{x}_\alpha|_U^{-1}$

Chapter 2

(1) Page 61. The curve defined in the 6th line from the bottom of the page is missing a t and should read as follows:

$$c_i(t) := \mathtt{x}^{-1}(\mathtt{x}(p) + t\mathbf{e}_i),$$

(2) Page 68. In Definition 2.19: "...so a tangent vector in T_qM" should read "...so a tangent vector in T_qN".

(3) Page 73. At the end of the second display, (M_2, p) should be changed to (M_2, q).

(4) Page 109. In the proof of Theorem 2.113, the first display should read

$$X_1(p), \ldots, X_k(p), \left.\frac{\partial}{\partial y^{k+1}}\right|_p, \ldots, \left.\frac{\partial}{\partial y^n}\right|_p$$

(The initially introduced coordinates are $y^1, ..., y^n$.)

(5) On Page 115, in the second set of displayed equations there is both a spurious q and a spurious p (although, the action of the derivations in question are indeed happening at the respective points: The

display should read as follows:

$$\begin{aligned}(\phi^* df)|_p\, v &= df|_q\,(T_p\phi \cdot v)\\ &= ((T_p\phi \cdot v)\, f)\\ &= v\,(f \circ \phi) = d(\phi^* f)|_p\, v.\end{aligned}$$

(6) Page 124. Problem (18). The second to last line of part (b) we should find $T_0 R^n = \Delta_\infty \cong \left(\mathfrak{m}_\infty / \mathfrak{m}_\infty^2\right)^*$.

Chapter 3

(1) Page 167, Equation 4.3.

Equation 4.3 is missing a square root. It should read

$$L(c) = \int_a^b \left(\sum_{i,j=1}^{n-1} g_{ij}(c(t)) \frac{dc^i}{dt}\frac{dc^j}{dt}\right)^{1/2} dt,$$

where $c^i(t) := u^i \circ c$.

(2) The last displayed set of equations on page 175 is missing two square roots in the integrands and should read as follows:

$$\begin{aligned}L(\gamma) &= \int_a^b \|\dot\gamma(t)\|\, dt = \int_a^b \left\langle \sum \frac{du^i}{dt}\frac{\partial \mathbf{x}}{\partial u^i}, \sum \frac{du^j}{dt}\frac{\partial \mathbf{x}}{\partial u^j} \right\rangle^{1/2} dt\\ &= \int_a^b \left(\sum_{i,j}^{n-1} g_{ij}(u(t)) \frac{du^i}{dt}\frac{du^j}{dt}\right)^{1/2} dt,\end{aligned}$$

Chapter 5

(1) Page 212. In the first displayed equation in the proof of Theorem 5.7.2, the time derivative should be taken at an arbitrary time t and so $\frac{d}{dt}\big|_{t=0} \bar{c}(t)$ should be changed to $\frac{d}{dt}\bar{c}(t)$.

(2) Page 218. The latter part of the third sentence in the proof of Theorem 5.81 on page 218 should read "is a relatively closed set in U".

Chapter 9

(1) Page 395. All integrations in the first set of displayed equations should be over the entire space $\mathbb{R}^n$. (In an earlier version I had put the support in the left half space but after changing my mind I forgot to modify these integrals.)

(2) Page 395. In "Case 2" right *before* the line the begins "If $j = 1$", the variable u^1 ranges from $-\infty$ to 0 and so "$= \int_{\mathbb{R}^{n-1}}$" should be replaced by "$= \int_{\mathbb{R}^{n-1}_{u^1 \leq 0}}$".

(3) Page 395. In "Case 2" right *after* the line the begins "If $j = 1$", the displayed equation should read

$$\int_{\mathbb{R}^n_{u^1 \leq 0}} d\omega_1 = \int_{\mathbb{R}^{n-1}} \left(\int_{-\infty}^{0} \frac{\partial f}{\partial u^1} du^1 \right) du^2 \cdots du^n$$

(4) Page 396. In the first line of the second set of displayed equations a U_α need to be changed to M. Also, we use the easily verified fact that $\int_M \omega = \int_U \omega$ when the support of ω is contained in an open set U. (We apply this to the $\rho_\alpha \omega$ in the proof.) The display should read:

$$\begin{aligned} \int_M d\omega &= \int_M \sum_\alpha d(\rho_\alpha \omega) = \sum_\alpha \int_{U_\alpha} d(\rho_\alpha \omega) \\ &= \sum_\alpha \int_{\mathtt{x}_\alpha(U_\alpha)} (\mathtt{x}_\alpha^{-1})^* d(\rho_\alpha \omega) = \sum_\alpha \int_{\mathtt{x}_\alpha(U_\alpha)} d((\mathtt{x}_\alpha^{-1})^* \rho_\alpha \omega) \\ &= \sum_\alpha \int_{\partial\{\mathtt{x}_\alpha(U_\alpha)\}} ((\mathtt{x}_\alpha^{-1})^* \rho_\alpha \omega) = \sum_\alpha \int_{\partial U_\alpha} \rho_\alpha \omega = \int_{\partial M} \omega, \end{aligned}$$

(5) Page 396-397. Starting at the bottom of 396 and continuing onto 397, all references to the interval $(-\pi, \pi)$ should obviously be changed to $(0, \pi)$. The integrals at the top of page 397 should be changed accordingly.

(6) Page 399. The last line before the final display should read "where $(\rho X^k)_i := d(\rho X^k)(E_i)$ for all k, i."

Chapter 10

(1) On page 444 immediately before Definition 10.3, the following should be inserted: Recall that if a sequence of module homomorphisms, $\cdots \to A_{k-1} \xrightarrow{f_k} A^k \xrightarrow{f_{k+1}} A_{k+1} \to \cdots$ has the property that $\mathrm{Ker}(f_{k+1}) = \mathrm{Im}(f_k)$, then we say that the sequence is exact at A_k and the sequence is said to be **exact** if it is exact at A_k for all k (or all k such that the kernel an image exist).

(2) Page 444. In the middle of the page, the sentence that defines a chain map should state that the map is linear. Indeed, in this section it is likely that all maps between modules or vector spaces are linear by default.

(3) Page 444. The last line before definition 10.5 should read "$x \in Z^k(A)$". The following should be added: Note that if $x - x' = dy$ then $f(x) - f(x') = f(dy) = d(f(y)$ so that $f(x) \sim f(x')$.

(4) Page 451. In the second and third lines of the first display of "Case I", the factor $(-1)^k$ should be omitted (in both lines). This factor does appear in the forth line.

(5) Page 451. In the second to last display there is a spurious "*". Replace $(f \circ s_a \circ \pi^*)\,\pi^*\alpha$ by $(f \circ s_a \circ \pi)\,\pi^*\alpha$.

Appendicies

(1) Page 639, Definition A.4,
"A morphism that is both a monomorphism and an epimorphism is called an isomorphism."
should be changed to
"A morphism that is both a monomorphism and an epimorphism is called a **bimorphism**. A morphism is called an **isomorphism** if it has both a right and left inverse."

(2) Page 648. In the constant rank theorem a q needs to be changed to a 0. We should read "...there are local diffeomorphisms $g_1 : (\mathbb{R}^n, p) \to (\mathbb{R}^n, 0)$".

Chapter 0

Background Material

0.1. Sets and Maps

Any reader who already knows the meaning of such words as *domain*, *codomain*, *surjective*, *injective*, *Cartesian product,* and *equivalence relation,* should just skip this section to avoid boredom.

According to G. Cantor, one of the founders of set theory, a **set** is a collection of objects, real or abstract, taken as a whole or group. The objects must be clear and distinct to the intellect. If one considers the set of all ducks on a specific person's farm then this seems clear enough as long as the notion of a duck is well defined. But is it? When one thinks about the set of all ducks that ever lived, then things get fuzzy (think about evolutionary theory). Clearly there are philosophical issues with how we define and think about various objects, things, types, kinds, ideas and collections. In mathematics, we deal with seemingly more precise, albeit abstract, notions and objects such as integers, letters in some alphabet or a set of such things (or a set of sets of such things!). We will not enter at all into questions about the clarity or ontological status of such mathematical objects nor shall we discuss "proper classes". Suffice it to say that there are monsters in those seas.

Each object or individual in a set is called an **element** or a **member** of the set. The number 2 is an element of the set of all integers-we say it **belongs** to the set. Sets are equal if they have the same members. We often use curly brackets to specify a set. The letter a is a member of the set $\{a, B, x, 6\}$. We use the symbol $\in$ to denote membership; $a \in A$ means "a is a member of the set A". We read it in various ways such as "a is a element of A" or "a in A". Thus, $x \in \{a, B, x, 6\}$ and if we denote the set of natural numbers by $\mathbb{N}$ then $42 \in \mathbb{N}$ and $3 \in \mathbb{N}$ etc. We include in our naive theory of sets a unique set called the **empty set** which has no members at all. It is denoted by $\varnothing$ or by $\{\}$. Besides just listing the members of a set we make use of notation such as $\{x \in \mathbb{N} : x \text{ is prime}\}$ which means the set of all natural numbers with the property of being a prime numbers. Another example; $\{a : a \in \mathbb{N} \text{ and } a < 7\}$. This is the same set as $\{1, 2, 3, 4, 5, 6\}$.

If every member of a set A is also a member of a set B then we say that A is a **subset** of B and write this as $A \subset B$ or $B \supset A$. Thus if $a \in A \Longrightarrow b \in A$. (The symbol "$\Longrightarrow$" means "implies"). The empty set is a subset of any set we may always write $\varnothing \subset A$ for any A. Also, $A \subset A$ for any set A. It is easy to see that if $A \subset B$ and $B \subset A$ then A and B have the same members–they are the same set differently named. In this case we write $A = B$. If $A \subset B$ but $A \neq B$ then we say that A is a **proper subset** of B and write $A \subsetneqq B$.

The **union** of two sets A and B is defined to be the set of all elements that are members of either A or B (or both). The union is denoted $A \cup B$. Thus $x \in A \cup B$ if and only if $x \in A$ or $x \in B$. Notice that here by "$x \in A$ or $x \in B$" we mean to allow the possibility that x might be a member of both A and B. The **intersection** of A and B the set of elements belonging to both A and B. We denote this by $A \cap B$. For example, $\{6, \pi, -1, 42, \{-1\}\} \cap \mathbb{N} = \{6, 42\}$ (Implicitly we assume that no two elements of a set given as a list are equal unless they are denoted by the same symbol. For example, in $\{6, a, B, x, 42\}$, context implies that $a \neq 42$.) If $A \cap B = \varnothing$ then A and B have no elements in common and we call them **disjoint**.

There are the following algebraic laws: If A, B, and C are sets then

(1) $A \cap B = B \cap A$ and $A \cup B = B \cup A$ (Commutativity),

(2) $(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$ (Associativity),

(3) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivity). It follows that we also have

(4) $A \cup (A \cap B) = A$ and $A \cap (A \cup B) = A$

We can take $A \cup B \cup C$ to mean $(A \cup B) \cup C$. In fact, if $A_1, \cdots, A_n$ are sets than we can inductively make sense out of the union $A_1 \cup A_2 \cup \cdots \cup A_n$ and intersection $A_1 \cap A_2 \cap \cdots \cap A_n$.

If A and B are sets then $A \backslash B$ denotes that set of all elements of A that are not also elements of B. The notation $A - B$ is also used and is called the difference. When considering subsets of a fix set X determined by the context of the discussion we can denote $X \backslash A$ as A^c and refer to A^c as the complement of A (in X). For example, if A is a subset of the set of real numbers (denoted $\mathbb{R}$) then $A^c = \mathbb{R}\backslash A$. It is easy to see that if $A \subset B$ then $B^c \subset A^c$. We have **de Morgan's laws**:

(1) $(A \cup B)^c = A^c \cap B^c$

(2) $(A \cap B)^c = A^c \cup B^c$

A **family of sets** is just a set whose members are sets. For example, for a fixed set X we have the family $\mathcal{P}(X)$ of all subsets of X. The family $\mathcal{P}(X)$ is called the **power set** of X. Note that $\varnothing \in \mathcal{P}(X)$. If $\mathcal{F}$ is some family of sets then we can consider the union or intersection of all the sets in $\mathcal{F}$.

$$\textstyle\bigcup_{A\in\mathcal{F}} A := \{x : x \in A \text{ for some } A \in \mathcal{F}\}$$
$$\textstyle\bigcap_{A\in\mathcal{F}} A := \{x : x \in A \text{ for every } A \in \mathcal{F}\}$$

It is often convenient to consider **indexed families** of sets. For example, if A_1, A_2, and A_3 are sets then we have a family $\{A_1, A_2, A_3\}$ indexed by the index set $\{1, 2, 3\}$. The family of sets may have an infinite number of members and we can use an infinite index set. If I is an index set (possibly uncountably infinite) then a family $\mathcal{F}$ indexed by I has as members sets which are each denoted A_i for some $i \in I$ and some fixed letter A. For notation we use

$$\mathcal{F} = \{A_i : i \in I\} \text{ or } \{A_i\}_{i\in I}.$$

We can then write unions and intersections as $\cup_{i\in I} A_i$ and $\cap_{i\in I} A_i$. De Morgan's laws become

$$\left(\textstyle\bigcup_{i\in I} A_i\right)^c = \textstyle\bigcap_{i\in I} A_i^c,$$
$$\left(\textstyle\bigcap_{i\in I} A_i\right)^c = \textstyle\bigcup_{i\in I} A_i^c.$$

A partition of a set X is a family of subsets of X, say $\{A_i\}_{i\in I}$ such that $A_i \cap A_j = \varnothing$ when ever $i \neq j$ and such that $X = \bigcup_{i\in I} A_i$.

An **ordered pair** with first element a and second element b is denoted (a, b). To make this notion more precise one could take (a, b) to be the set $\{a, \{a, b\}\}$. Since the notion of an ordered pair is so intuitive it is seldom necessary to resort to thinking explicitly of $\{a, \{a, b\}\}$. Given two sets A and B, the Cartesian product of A and B is the set of order pairs (a, b) where $a \in A$ and $b \in B$. We denote the Cartesian product of A and B by $A\times B$. We can also consider ordered n-tuples such as $(a_1, ..., a_n)$. For a list of sets $A_1, A_2, \ldots, A_k$ we define the n-fold Cartesian product:

$$A_1 \times A_2 \times \cdots \times A_n := \{(a_1, ..., a_n) : a_i \in A_i\}.$$

If $A = A_1 = A_2 = \cdots = A_n$ then we denote $A_1 \times A_2 \times \cdots \times A_n$ by A^n. For any set X the **diagonal** subset Δ_X of $X \times X$ is defined by $\Delta_X := \{(a, b) \in X \times X : a = b\} = \{(a, a) : a \in X\}$.

Example 0.1. If $\mathbb{R}$ denotes the set of real numbers then, for a given positive integer n, $\mathbb{R}^n$ denotes the set of n-tuples of real numbers. This set has a lot of structure as we shall see.

One of the most important notions in all of mathematics is the notion of a "relation". A **relation** from a set A to a set B is simply a subset of $A \times B$. A relation from A to A is just called a relation *on* A. If R is a relation from A to B then we often write $a\,R\,b$ instead of $(a, b) \in R$. We single out two important types of relations:

Definition 0.2. An **equivalence relation** on a set X is a relation on X, usual denoted by the symbol $\sim$, such that (i) $x \sim x$ for all $x \in X$, (ii) $x \sim y$ if and only if $y \sim x$, (iii) if $x \sim y$ and $y \sim z$ then $x \sim z$. For each $a \in X$ the set of all x such that $x \sim a$ is called the equivalence class of a (often denoted by $[a]$). The set of all equivalence classes form a partition of X. The set of equivalence classes if often denoted by $X/\sim$.

Conversely, it is easy to see that if $\{A_i\}_{i \in I}$ is a partition of X then we my defined a corresponding equivalence relation by declaring $x \sim y$ if and only if x and y belong to the same A_i.

For example, ordinary equality is an equivalence relation on the set of natural numbers. Let $\mathbb{Z}$ denote the set of integers. Then equality modulo a fixed integer p defines and equivalence relation on $\mathbb{Z}$ where $n \sim m$ iff[1] $n - m = kp$ for some $k \in \mathbb{Z}$. In this case the set of equivalence classes is denoted $\mathbb{Z}_p$ or $\mathbb{Z}/p\mathbb{Z}$.

Definition 0.3. A **partial ordering** on a set X (assumed nonempty) is a relation denoted by, say $\preceq$, that satisfies (i) $x \preceq x$ for all $x \in X$, (ii) if $x \preceq y$ and $y \preceq x$ then $x = y$, and (iii) if $x \preceq y$ and $y \preceq z$ then $x \preceq z$. We say that X is **partially ordered** by $\preceq$. If a partial ordering also has the property that for every $x, y \in X$ we have either $x \preceq y$ or $y \preceq x$ then we call the relation a **total ordering** or (linear ordering). In this case, we say that X is totally ordered by $\preceq$.

Example 0.4. The set of real numbers is totally ordered by the familiar notion of less than; $\leq$.

Example 0.5. The power set $\mathcal{P}(X)$ is partially ordered by set inclusion $\subset$ (also denoted $\subseteq$).

If X is partially ordered by $\preceq$ then an element x is called a **maximal element** if $x \preceq y$ implies $x = y$. A **minimal element** is defined similarly.

[1] "iff" means "if and only if".

Maximal elements might not be unique or even exist at all. If the relation is a total ordering then a maximal (or minimal) element, if it exists, is unique.

A rule that assigns to each element of a set A an element of a set B is called a map, mapping, or function from A to B. If the map is denoted by f, then $f(a)$ denotes the element of B that f assigns to a. If is referred to as the image of the element a. As the notation implies, it is not allowed that f assign to an element $a \in A$ two *different* elements of B. The set A is called the **domain** of f and B is called the **codomain** of f. The domain an codomain must be specified and are part of the definition. The prescription $f(x) = x^2$ does not specify a map or function until we specify the domain. If we have a function f from A to B we indicate this by writing $f : A \to B$ or $A \xrightarrow{f} B$. If a is mapped to $f(a)$ indicate this also by $a \mapsto f(a)$. This notation can be used to define a function; we might say something like "consider the map $f : \mathbb{Z} \to \mathbb{N}$ defined by $n \mapsto n^2 + 1$." The symbol "$\mapsto$" is read as "maps to" or "is mapped to".

It is desirable to have a definition of map that appeals directly to the notion of a set. To do this we may define a function f from A to B to be a relation $f \subset A \times B$ from A to B such that if $(a, b_1) \in f$ and $(a, b_2) \in f$ then $b_1 = b_2$.

Example 0.6. The relation $R \subset \mathbb{R} \times \mathbb{R}$ defined by $R := \{(a, b) \in \mathbb{R} \times \mathbb{R} : a^2 + b^2 = 1\}$ is not a map. However, the relation $\{(a, b) \in \mathbb{R} \times \mathbb{R} : a^2 - b = 1\}$ is a map; namely the map $\mathbb{R} \to \mathbb{R}$ defined, as a rule, by $a \mapsto a^2 - 1$ for all $a \in \mathbb{R}$.

Definition 0.7. (1) A map $f : A \to B$ is said to be **surjective** (or "onto") if for every $b \in B$ there is at least one ma$\in A$ such that $f(a) = b$. Such a map is called a **surjection** and we say that f maps A **onto** B.
(2) A map $f : A \to B$ is said to be **injective** (or "one to one") if whenever $f(a_1) = f(a_2)$ then $a_1 = a_2$. We call such a map an **injection**.

Example 0.8. The map $f : \mathbb{R} \to \mathbb{R}^3$ given by $f : t \mapsto (\cos t, \sin t, t)$ is injective but not surjective. Let $S^2 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ denote the unit circle in $\mathbb{R}^2$. The map $f : \mathbb{R} \to S^2$ given by $t \mapsto (\cos t, \sin t)$ is surjective but not injective. Note that specifying the codomain is essential; the map $f : \mathbb{R} \to \mathbb{R}^2$ also given by the formula $t \mapsto (\cos t, \sin t)$ is *not* surjective.

The are some special maps to consider. For any set A we have the identity map $\mathrm{id}_A : A \to A$ such $\mathrm{id}_A(a) = a$ for all $a \in A$. If $A \subset X$ then the map $\imath_{A,X}$ from A to X given by $\imath_{A,X}(a) = a$ is called the **inclusion map** from A *into* X. Notice that id_A and $\imath_{A,X}$ are different maps since they have different codomains. We shall usually abbreviate $\imath_{A,X}$ to $\imath$ and suggestively write $\imath : A \hookrightarrow X$ instead of $\imath : A \to X$.

When we have a map $f : A \to B$ there is are tow induced maps on power sets. For $S \subset A$ let use define $f(S) := \{b \in B : b = f(s) \text{ for some } s \in S\}$. Then reusing the symbol f we obtain a map $f : \mathcal{P}(A) \to \mathcal{P}(B)$ given by $S \mapsto f(S)$. We call $f(S)$ the image of S under f. We have the following properties

(1) If $S_1, S_2 \subset A$ then $f(S_1 \cup S_2) = f(S_1) \cup f(S_2)$,

(2) If $S_1, S_2 \subset A$ then $f(S_1 \cap S_2) \subset f(S_1) \cap f(S_2)$.

A second map is induced on power sets $f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A)$. Notice the order reversal. Here the definition of $f^{-1}(E)$ for $E \subset B$ is $f^{-1}(E) := \{a \in A : f(a) \in B\}$. This time the properties are even nicer:

(1) If $E_1, E_2 \subset B$ then $f^{-1}(E_1 \cup E_2) = f^{-1}(E_1) \cup f^{-1}(E_2)$

(2) If $E_1, E_2 \subset B$ then $f^{-1}(E_1 \cap E_2) = f^{-1}(E_1) \cap f^{-1}(E_2)$ (equality!)

(3) $f^{-1}(E^c) = \left(f^{-1}(E)\right)^c$

In fact, these ideas work in more generality. For example, if $\{E_i\}_{i\in I} \subset \mathcal{P}(B)$ then $f^{-1}\left(\bigcup_{i\in I} E_i\right) = \bigcup_{i\in I} f^{-1}(E_i)$ and similarly for intersection.

Now suppose that we have a map $f : A \to B$ and another map $g : B \to C$ then we can form the **composite map** $g \circ f : A \to C$ (or **composition**) by $(g \circ f)(a) = g(f(a))$ for $a \in A$. Notice that we have been carefully assumed that the codomain of f is also the domain of g. But in many areas of mathematics this can be relaxed. For example, if we have $f : A \to B$ and $g : X \to C$ where $B \subset X$ then we can define $g \circ f$ by the same prescription as before. In some fields (such as algebraic topology) this is dangerous and one would have to use an inclusion map $g \circ \imath_{B,X} \circ f$. In some cases we can be even more relaxed and compose maps $f : A \to B$ and $g : X \to C$ by letting the domain of $g \circ f$ be $\{a \in A : f(a) \in X\}$ assuming this set is not empty.

If $f : A \to B$ is a given map and $S \subset A$ then we obtain a map $f|_S : S \to B$ by the simple rule $f|_S(a) = f(a)$ for $a \in S$. The map $f|_S$ is called the **restriction** of f to S. It is quite possible that $f|_S$ is injective even if f is not (think about $f(x) = x^2$ restricted to $\{x > 0\}$).

Definition 0.9. A map $f : A \to B$ is said to be **bijective** and referred to as a **bijection** if it is *both* a surjection and an injection. In this case we say that f is a **one to one correspondence** between A and B. If f is a bijection then we have the **inverse map** $f^{-1} : B \to A$ defined by stipulating that $f^{-1}(b)$ is equal to the unique ma$\in A$ such that $f(a) = b$. In this case we have $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$.

It is sometimes convenient and harmless to relax our thinking about f^{-1}. If $f : A \to B$ is injective but not surjective then there is a bijective map in the offing. This map has codomain $f(A)$ and otherwise agrees with f. It is just f considered as a map onto $f(A)$. What shall we call this map? If we are careful to explain ourselves we can use the same letter f and then we have an inverse map $f^{-1} : f(A) \to A$. We say that the injective map f is a *bijection onto its image* $f(A)$.

If a set contains only a finite number of elements then that number is called the **cardinality** of the set. If the set is infinite we must be more clever. If there exists some injective map from A to B we say that the cardinality of A is less than the cardinality of B and write $|A| \leq |B|$. If there exists some surjective map from A to B we say that the cardinality of A is greater than the cardinality of B and write $|A| \geq |B|$. If there exists a bijection then we say the sets have the same cardinality and write $|A| = |B|$. The Schröder-Bernstein theorem states that if $|A| \leq |B|$ and $|A| \geq |B|$ then $|A| = |B|$. This is not obvious. If there exist an injection but no bijection then we write $|A| < |B|$.

Definition 0.10. If there exists a bijection from A to the set of natural numbers $\mathbb{N}$ then we say that A a **countably infinite** set. If A is either a finite set or countably infinite set we say it is a **countable** set. If $|\mathbb{N}| < |A|$ then we say that A is **uncountable**.

Example 0.11. The set of rational numbers, $\mathbb{Q}$ is well known to be countably infinite while the set of real numbers is uncountable (Cantor's diagonal argument).

0.2. Linear Algebra (under construction)

We will defined vector space below but since every vector space has a so called base field, we should first list the axioms of a field. Map from $A \times A$ to another set (often A) is sometimes called an operation. When the operation is denoted by a symbol such as $+$ then we write $a + b$ in place of $+(a, b)$. The use of this symbol means we will refer to the operation as addition. Sometimes the symbol for the operation is something like $\cdot$ or $\odot$ in which case we would write $a \cdot b$ and call it multiplication. In this case we will actually denote the operation simply by juxtaposition just as we do for multiplication of real or complex numbers. Truth be told, the main examples of fields are the real numbers or complex numbers with the familiar operations of addition and multiplication. The reader should keep these in mind when reading the following definition but also remember that the ideas are purely abstract and many examples exist where the meaning of $+$ and $\cdot$ is something unfamiliar.

Definition 0.12. A set $\mathbb{F}$ together with an operation $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ and an operation $\cdot$ also mapping $\mathbb{F} \times \mathbb{F} \to \mathbb{F}$ is called a field is the following axioms are satisfied.

F1. $r + s = s + r$. for all $r, s \in \mathbb{F}$

F2. $(r + s) + t = r + (s + t)$. for all $r, s, t \in \mathbb{F}$

F3. There exist a unique element $0 \in \mathbb{F}$ such that $r + 0 = r$ for all $r \in \mathbb{F}$.

F4. For every $r \in \mathbb{F}$ there exists a unique element $-r$ such that $r + (-r) = 0$.

F5. $r \cdot s = s \cdot r$ for all $r, s \in \mathbb{F}$.

F6. $(r \cdot s) \cdot t = r \cdot (s \cdot t)$

F7. There exist a unique element $1 \in \mathbb{F}$, such that $1 \neq 0$ and such that $1 \cdot r = r$ for all $r \in \mathbb{F}$.

F8. For every $r \in \mathbb{F}$ with $r \neq 0$ there is a unique element r^{-1} such that $r \cdot r^{-1} = 1$.

F9. $r \cdot (s + t) = r \cdot s + r \cdot t$ for all $r, s, t \in \mathbb{F}$.

Most of the time we will write rs in place of $r \cdot s$. We may use some other symbols such as $\oplus$ and $\odot$ in order to distinguish from more familiar operations. Also, we see that a field is really a triple $(\mathbb{F}, +, \cdot)$ but we follow the common practice of referring to the field by referring to the set so we will denote the field by $\mathbb{F}$ and speak of elements of the field and so on. The field of real numbers is denoted $\mathbb{R}$ and the field of complex numbers is denoted $\mathbb{C}$. We use $\mathbb{F}$ to denote some unspecified field but we will almost always have $\mathbb{R}$ or $\mathbb{C}$ in mind. Another common field is the the set $\mathbb{Q}$ of rational numbers with the usual operations.

Example 0.13. Let p be some prime number. Consider the set $\mathbb{F}_p = \{0, 1, 2, ..., p - 1\}$. Define an addition $\oplus$ on $\mathbb{F}_p$ by the following rule. If $x, y \in \mathbb{F}_p$ then let $x \oplus y$ be the unique $z \in \mathbb{F}_p$ such that $x + y = z + kp$ for some integer k. Now define a multiplication by $x \odot y = w$ where w is the unique element of $\mathbb{F}_p$ such that $xy = w + kp$ for some integer k. Using the fact that p is prime it is possible to show that $(\mathbb{F}_p, \oplus, \odot)$ is a (finite) field.

Consider the following sets; $\mathbb{R}^n$, the set $C([0, 1])$ of continuous functions defined on the closed interval $[0, 1]$, the set of $n \times m$ real matrices, the set of directed line segments emanating from a fixed origin in Euclidean space. What do all of these sets have in common? Well, one thing is that in each case there is a natural way to add elements of the set and was way to "scale" or multiply elements by real numbers. Each is natural a real vector space. Similarly, $\mathbb{C}^n$ and complex $n \times m$ matrices and many other examples are complex vectors spaces. For any field $\mathbb{F}$ we define an $\mathbb{F}$-vector space or a vector space over $\mathbb{F}$. The field in question is called the field of scalars for the vector space.

Definition 0.14. A vector space over a field $\mathbb{F}$ is a set V together with an addition operation $V \times V \to V$ written $(v, w) \mapsto v + w$ and an operation of scaling $\mathbb{F} \times V \to V$ written simply $(r, v) \mapsto rv$ such that the following axioms hold:

V1. $v + w = w + v$ for all $v, w \in V$.

V2. $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$.

V3. The exist a unique member of V denoted $\mathbf{0}$ such that $v + \mathbf{0} = v$ for all $v \in V$.

V4. For each $v \in V$ there is a unique element $-v$ such that $v + (-v) = \mathbf{0}$.

V5. $(rs)\, v = r\,(sv)$ for all $r, s \in \mathbb{F}$ and all $v \in V$.

V6. $r\,(v + w) = rv + rw$ for all $v, w \in V$ and $r \in \mathbb{F}$.

V7. $(r + s)\, v = rv + sv$ for all $r, s \in \mathbb{F}$ and all $v \in V$.

V8. $1v = v$ for all $v \in V$.

Notice that in axiom V7 the $+$ on the left is the addition in $\mathbb{F}$ while that on the right is addition in V. These are separate operations. Also, in axiom V4 the $-v$ is not meant to denote $-1v$. However, it can be shown from the axioms that it is indeed true that $-1v = -v$.

Examples $\mathbb{F}^n$.

Subspaces

Linear combinations, Span, span(A) for $A \subset V$.

Linear independence

Basis, dimension

Change of basis

Two bases $\mathcal{E} = (e_1, \dots, e_n)$ and $\tilde{\mathcal{E}} = (\tilde{e}_1, \dots, \tilde{e}_n)$ are related by a non-sigular matrix C according to $e_i = \sum_{j=1}^n C_i^j \tilde{e}_j$. In this case we have $v = \sum_{i=1}^n v^i e_i = \sum_{j=1}^n \tilde{v}^j \tilde{e}_j$. On the other hand,

$$\begin{aligned} \sum_{j=1}^n \tilde{v}^j \tilde{e}_j = v = \sum_{i=1}^n v^i e_i = \sum_{i=1}^n v^i \left(\sum_{j=1}^n C_i^j \tilde{e}_j \right) \\ = \sum_{j=1}^n \left(\sum_{i=1}^n C_i^j v^i \right) \tilde{e}_j \end{aligned}$$

from which we deduce that

$$\tilde{v}^j = \sum_{i=1}^n C_i^j v^i.$$

$$L : V \to W$$

$$\begin{aligned}
Le_j &= \sum_{i=1}^{m} A_j^i f_i \\
Lv &= L\sum_{j=1}^{n} v^j e_j = \sum_{j=1}^{n} v^j Le_j = \sum_{j=1}^{n} v^j \left(\sum_{i=1}^{m} A_j^i f_i\right) \\
&= \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_j^i v^j\right) f_i
\end{aligned}$$

From which we conclude that if $Lv = w = \sum w^i f_i$ then

$$w^i = \sum_{j=1}^{n} A_j^i v^j$$

Consider another basis $\tilde{f}_i$ and suppose $f_i = \sum_{j=1}^{m} D_i^j \tilde{f}_j$. Of course, $\tilde{w}^i = \sum_{j=1}^{n} \tilde{A}_j^i \tilde{v}^j$ where $L\tilde{e}_j = \sum_{i=1}^{m} \tilde{A}_j^i \tilde{f}_i$ and $w = \sum \tilde{w}^i \tilde{f}_i$.

$Le_j = \sum_{i=1}^{m} A_j^i f_i$ and

$$\begin{aligned}
Le_j &= \sum_{i=1}^{m} A_j^i f_i = \sum_{i=1}^{m} A_j^i \left(\sum_{k=1}^{m} D_i^k \tilde{f}_k\right) \\
&= \sum_{k=1}^{m}\left(\sum_{i=1}^{m} D_i^k A_j^i\right) \tilde{f}_k
\end{aligned}$$

But also

$$\begin{aligned}
Le_j &= L\sum_{r=1}^{n} C_j^r \tilde{e}_r = \sum_{r=1}^{n} C_j^r L\tilde{e}_r \\
&= \sum_{r=1}^{n} C_j^r \left(\sum_{k=1}^{m} \tilde{A}_r^k \tilde{f}_k\right) = \sum_{k=1}^{m}\left(\sum_{r=1}^{n} \tilde{A}_r^k C_j^r\right)\tilde{f}_k
\end{aligned}$$

and so

$$\begin{aligned}
\sum_{i=1}^{m} D_i^k A_j^i &= \sum_{r=1}^{n} \tilde{A}_r^k C_j^r \\
\sum_{j=1}^{n}\sum_{i=1}^{m} D_i^k A_j^i \left(C^{-1}\right)_s^j &= \sum_{j=1}^{n}\sum_{r=1}^{n} \tilde{A}_r^k C_j^r \left(C^{-1}\right)_s^j \\
\sum_{j=1}^{n}\sum_{i=1}^{m} D_i^k A_j^i \left(C^{-1}\right)_s^j &= \sum_{r=1}^{n} \tilde{A}_r^k \delta_s^r = \tilde{A}_s^k \\
\tilde{A}_s^k &= \sum_{j=1}^{n}\sum_{i=1}^{m} D_i^k A_j^i \left(C^{-1}\right)_s^j
\end{aligned}$$

and so

$$\tilde{A}^k_s = \sum_{i=1}^{m} D^k_i A^i_j \left(C^{-1}\right)^j_s$$

Linear maps– kernel, image rank, rank-nullity theorem. Nonsingular, isomorphism, inverse

Vector spaces of linear maps

Group $GL(V)$

Examples

Matrix representatives, change of basis, similarity

$GL(n, \mathbb{F})$

Dual Spaces, dual basis

dual map

contragedient

multilinear maps

components, change of basis

tensor product

Inner product spaces, Indefinite Scalar Product spaces

$O(V)$, $O(n)$, $U(n)$

Normed vector spaces, continuity

Canonical Forms

0.3. General Topology (under construction)

metric

metric space

normed vector space

Euclidean space

topology (System of Open sets)

Neighborhood system

open set

neighborhood

closed set

closure

aherent point

interior point

discrete topology

metric topology
bases and subbases
continuous map (and compositions etc.)
open map
homeomorphism
weaker topology stronger topology
subspace
relatively open relatively closed
relative topology
product space
product topology
accumulation points
countability axioms
separation axioms
coverings
locally finite
point finite
compactness (finite intersection property)
image of a compact space
relatively compact
paracompactness

0.4. Shrinking Lemma

Theorem 0.15. *Let X be a normal topological space and suppose that $\{U_\alpha\}_{\alpha\in A}$ is a point finite open cover of X. Then there exists an open cover $\{V_\alpha\}_{\alpha\in A}$ of X such that $\overline{V}_\alpha \subset U_\alpha$ for all $\alpha \in A$.*

second proof. We use Zorn's lemma (see [**Dug**]). Let $\mathcal{T}$ denote the topology and let Φ be the family of all functions $\varphi : A \to \mathcal{T}$ such that
i) either $\varphi(\alpha) = U_\alpha$ or $\varphi(\alpha) = V_\alpha$ for some open V_α with $\overline{V}_\alpha \subset U_\alpha$
ii) $\{\varphi(\alpha)\}$ is an open cover of X.
We put a partial ordering on the family Φ by interpreting $\varphi_1 \prec \varphi_2$ to mean that $\varphi_1(\alpha) = \varphi_2(\alpha)$ whenever $\varphi_1(\alpha) = V_\alpha$ (with $\overline{V}_\alpha \subset U_\alpha$). Now let Ψ be a totally order subset of Φ and set

$$\psi^*(\alpha) = \bigcap_{\psi\in\Psi} \psi(\alpha)$$

Now since Ψ is totally ordered we have either $\psi_1 \prec \psi_2$ or $\psi_2 \prec \psi_1$. We now which to show that for a fixed α the set $\{\psi(\alpha) : \psi \in \Psi\}$ has no more than two members. In particular, $\psi^*(\alpha)$ is open for all α. Suppose that $\psi_1(\alpha), \psi_2(\alpha)$ and $\psi_3(\alpha)$ are distinct and suppose that $\psi_1 \prec \psi_2 \prec \psi_3$ without loss. If $\psi_1(\alpha) = V_\alpha$ with $\overline{V}_\alpha \subset U_\alpha$ then we must have $\psi_2(\alpha) = \psi_3(\alpha) = V_\alpha$ by definition of the ordering but this contradicts the assumption that $\psi_1(\alpha), \psi_2(\alpha)$ and $\psi_3(\alpha)$ are distinct. Thus $\psi_1(\alpha) = U_\alpha$. Now if $\psi_2(\alpha) = V_\alpha$ for with $\overline{V}_\alpha \subset U_\alpha$ then $\psi_3\ (\alpha) = V_\alpha$ also a contradiction. The only possibility left is that we must have both $\psi_1(\alpha) = U_\alpha$ and $\psi_2(\alpha) = U_\alpha$ again a contradiction. Now the fact that $\{\psi(\alpha) : \psi \in \Psi\}$ has no more than two members for every α means that ψ^* satisfies condition (i) above. We next show that ψ^* also satisfies condition (ii) so that $\psi^* \in \Phi$. Let $p \in X$ be arbitrary. We have $p \in \psi(\alpha_0)$ for some α_0.
By the locally finite hypothesis, there is a finite set $\{U_{\alpha_1}, ..., U_{\alpha_n}\} \subset \{U_\alpha\}_{\alpha \in A}$ which consists of exactly the members of $\{U_\alpha\}_{\alpha \in A}$ that contain p. Now $\psi^*(\alpha_i)$ must be U_{α_i} or V_{α_i}. Since p is certainly contained in U_{α_i} we may as well assume the worst case where $\psi^*(\alpha_i) = V_{\alpha_i}$ for all i. There must be $\psi_i \in \Psi$ such that $\psi^*(\alpha_i) = \psi_i(\alpha_i) = V_{\alpha_i}$. By renumbering if needed, we may assume that

$$\psi_1 \prec \cdots \prec \psi_n$$

Now since ψ_n is a cover we know that x is in the union of the set in the family

$$\{\psi_n(\alpha_1), ..., \psi_n(\alpha_n)\} \cup \{\psi_n(\alpha)\}_{\alpha \in A \setminus \{\alpha_1, ...\alpha_n\}}$$

but it must be that $x \in \psi_n(\alpha_i)$ for some i. Since $\psi_i \preceq \psi_n$ we have $x \in \psi_i(\alpha_i) = \psi^*(\alpha_i)$. Thus since x was arbitrary we see that $\{\psi^*(\alpha)\}_{\alpha \in A}$ is a cover and hence $\psi^* \in \Psi$. By construction, $\psi^* = \sup \Psi$ and so by Zorn's lemma Φ has a maximal element $\varphi_{\max}$. We now show that $\varphi_{\max}(\alpha) = \overline{V}_\alpha \subset U_\alpha$ for every α. Suppose by way of contradiction that $\varphi_{\max}(\beta) = U_\beta$ for some β. Let $X_\beta = X - \cup_{\alpha \neq \beta} \varphi_{\max}(\alpha)$. Since $\{\varphi_{\max}(\alpha)\}_{\alpha \in A}$ is a cover we see that X_β is a closed subset of U_β. Applying normality we obtain $X_\beta \subset V_\beta \subset \overline{V}_\beta \subset U_\beta$. Now if we define $\varphi^\beta : A \to \mathcal{T}$ by

$$\varphi^\beta(\alpha) := \begin{cases} \varphi_{\max}(\alpha) \text{ for } \alpha \neq \beta \\ V_\beta \text{ for } \alpha = \beta \end{cases}$$

then we obtain an element of Φ which contradicts the maximality of $\varphi_{\max}$. $\square$

0.5. Calculus on Euclidean coordinate spaces (underconstruction)

0.5.1. Review of Basic facts about Euclidean spaces. For each positive integer d, let $\mathbb{R}^d$ be the set of all d-tuples of real numbers. This means that $\mathbb{R}^d := \underbrace{\mathbb{R}\times\cdots\times\mathbb{R}}_{d\text{ times}}$. Elements of $\mathbb{R}^d$ are denoted by notation such as $x := (x^1, \ldots, x^d)$. Thus, in this context, x^i is the i-th component of x and not the i-th power of a number x. In the context of matrix algebra we interpret $x = (x^1, \ldots, x^d)$ as a column matrix

$$x = \begin{bmatrix} x^1 \\ \vdots \\ x^d \end{bmatrix}$$

Whenever possible we write d-tuples horizontally. We use both superscript and subscripts since this is so common in differential geometry. This takes some getting used to but it ultimately has a big payoff. The set $\mathbb{R}^d$ is endowed with the familiar real vector space structure given by

$$\begin{aligned} x + y &= (x^1, \ldots, x^d) + (y^1, \ldots, y^d) = (x^1 + y^1, \ldots, x^d + y^d) \\ ax &= (ax^1, \ldots, ax^d) \text{ for } a \in \mathbb{R}. \end{aligned}$$

We denote $(0, ..., 0)$ as simply 0. Recall that if $\{v_1, ..., v_k\} \subset \mathbb{R}^d$ then we say that a vector x is a linear combination of $v_1, ..., v_k$ if $x = c_1 v_1 + \cdots + c_k v_k$ for some real numbers $c_1, ..., c_k$. In this case we say that x is in the span of $\{v_1, ..., v_k\}$ or that it is a linear combination the elements $v_1, ..., v_k$. Also, a set of vectors, say $\{v_1, ..., v_k\} \subset \mathbb{R}^d$, is said to be **linearly dependent** if for some j, with $1 \le j \le d$ we have that v_j is a linear combination of the remaining elements of the set. If $\{v_1, ..., v_k\}$ is not linearly dependent then we say it is **linearly independent**. Informally, we can refer directly to the elements rather than the set; we say things like "$v_1, ..., v_k$ are linearly independent". An ordered set of elements of $\mathbb{R}^d$, say $(v_1, ..., v_d)$, is said to be a **basis** for $\mathbb{R}^d$ if it is a linearly independent set and every member of $\mathbb{R}^d$ is in the span of this set. In this case each element $x \in \mathbb{R}^d$ can be written as a linear combination $x = c_1 v_1 + \cdots + c_d v_d$ and the c_i are unique. A basis for $\mathbb{R}^d$ must have d members. The standard basis for $\mathbb{R}^d$ is $(\mathbf{e}_1, ..., \mathbf{e}_d)$ where all components of $\mathbf{e}_i$ are zero except i-th component which is 1. As column matrices we have

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \longleftarrow \quad i\text{-th position} \\ \\ \\ \\ \end{matrix}$$

Recall that a map $L : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if $L(ax + by) = aL(x) + bL(y)$. For any such linear map there is an $m \times n$ matrix such that $L(v) = Av$ where in the expression "Av" we must interpret v and Av as column matrices. The matrix A associated to L has ij-th entry a^i_j determined by the equations $L(\mathbf{e}_j) = \sum_{i=1}^m a^i_j \mathbf{e}_i$. Indeed, we have

$$\begin{aligned} L(v) &= L(\sum_{j=1}^n v^j \mathbf{e}_j) = \sum_{j=1}^n v^j L(\mathbf{e}_j) \\ &= \sum_{j=1}^n v^j \left(\sum_{i=1}^m a^i_j \mathbf{e}_i \right) = \sum_{i=1}^m \left(\sum_{j=1}^n a^i_j v^j \right) \mathbf{e}_i \end{aligned}$$

which shows that the i-th component of $L(v)$ is i-th entry of the column vector Av. Note that here and elsewhere we do not notationally distinguish the i-th standard basis element of $\mathbb{R}^n$ from the i-th standard basis element of $\mathbb{R}^m$. A special case arises when we consider the dual space to $\mathbb{R}^d$. The dual space is the set of all linear transformations from $\mathbb{R}^d$ to $\mathbb{R}$. We call elements of the dual space linear functionals or sometimes "covectors". The matrix of a linear functional is a row matrix. So we often use subscripts $(a_1, \ldots, a_d)$ and interpret such a d-tuple as a row matrix which gives the linear functional $x \mapsto \sum a_i x^i$. In matrix form

$$\begin{bmatrix} x^1 \\ \vdots \\ x^d \end{bmatrix} \mapsto [a_1, \ldots, a_d] \begin{bmatrix} x^1 \\ \vdots \\ x^d \end{bmatrix}.$$

The d-dimensional Euclidean space coordinate space[2] is $\mathbb{R}^d$ endowed with the usual inner product defined by $\langle x, y \rangle := x^1 y^1 + \cdots + x^d y^d$ and the corresponding norm $\|\cdot\|$ defined by $\|x\| := \sqrt{\langle x, x \rangle}$. Recall the basic properties: For all $x, y, z \in \mathbb{R}^d$ and $a, b \in \mathbb{R}$ we have

(1) $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ only if $x = 0$

(2) $\langle x, y \rangle = \langle y, x \rangle$

(3) $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$

We also have the Schwartz inequality

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$$

with equality if and only if x and y are linearly dependent. The defining properties of a norm are satisfied: If $x, y \in \mathbb{R}^d$ and $a \in \mathbb{R}$ then

(1) $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$.

(2) $\|x + y\| \leq \|x\| + \|y\|$ Triangle inequality (follows from the Schwartz inequality)

[2] A Euclidean space is more properly and affine space modeled on a an inner product space. We shall follow the tradition of using the standard model for such a space to cut down on notational complexity.

(3) $\|ax\| = |a| \, \|x\|$

We define the distance between two element $x, y \in \mathbb{R}^d$ to be $\operatorname{dist}(x,y) := \|x - y\|$. It follows that the map $\operatorname{dist} : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ is a distance function or **metric** on $\mathbb{R}^d$. More precisely, for any $x, y, z \in \mathbb{R}^d$ we have the following expected properties of distance:

(1) $\operatorname{dist}(x,y) \geq 0$

(2) $\operatorname{dist}(x,y) = 0$ if and only if $x = y$,

(3) $\operatorname{dist}(x,y) = \operatorname{dist}(y,x)$

(4) $\operatorname{dist}(x,z) \leq \operatorname{dist}(x,y) + \operatorname{dist}(y,z)$.

The open ball of radius r centered at $x_0 \in \mathbb{R}^d$ is denoted $B(x_0, r)$ and defined to be the set

$$B(x_0, r) := \{x \in \mathbb{R}^d : \operatorname{dist}(x, x_0) < r\}.$$

Since $\operatorname{dist}(.,.)$ defines a metric on $\mathbb{R}^d$ (making it a "metric space") we have the ensuing notions of open set, closed set and continuity. Suppose that S is a subset of $\mathbb{R}^d$. A point $x \in S$ is called an interior point of S if there exist an open ball centered at s that is contain in S That is, x is interior to S if there is an $r > 0$ such that $B(x,r) \subset S$. We recall that a subset O of $\mathbb{R}^d$ is called an open subset if each member of O is an interior point of O. A subset of $\mathbb{R}^d$ is called a closed subset if its compliment is an open set. If for every $r > 0$ the ball $B(x_0, r)$ contains both points of S and points of S^c then we call x_0 a boundary point of S. The set of all boundary points of S is the topological boundary of S.

Let $A \subset \mathbb{R}^n$. A map $f : A \to \mathbb{R}^m$ is **continuous** at $a \in A$ if for every open ball $B(f(a), \epsilon)$ centered at $f(a)$, there is an open ball $B(a, \delta) \subset \mathbb{R}^n$ such that $f(A \cap B(a, \delta)) \subset B(f(a), \epsilon)$. If $f : A \to \mathbb{R}^m$ is continuous at every point of its domain then we just say that f is continuous. This can be shown to be equivalent to an very simple condition: $f : A \to \mathbb{R}^m$ is continuous on A provided that for every open $U \subset \mathbb{R}^m$ we have $f^{-1}(U) = A \cap V$ for some open subset V of $\mathbb{R}^n$.

0.6. Topological Vector Spaces

We shall outline some of the basic definitions and theorems concerning topological vector spaces.

Definition 0.16. A **topological vector space** (**TVS**) is a vector space V with a Hausdorff topology such that the addition and scalar multiplication operations are (jointly) continuous. (Not all authors require the topology to be Hausdorff).

Exercise 0.17. Let $T : \mathsf{V} \to \mathsf{W}$ be a linear map between topological vector space. Show that T is continuous if and only if it is continuous at the zero element $0 \in \mathsf{V}$.

The set of all neighborhoods that contain a point x in a topological vector space V is denoted $\mathcal{N}(x)$. The families $\mathcal{N}(x)$ for various x satisfy the following **neighborhood axioms**:

(1) Every set that contains a set from $\mathcal{N}(x)$ is also a set from $\mathcal{N}(x)$
(2) If $\{N_i\}$is a finite family of sets from $\mathcal{N}(x)$ then $\bigcap_i N_i \in \mathcal{N}(x)$
(3) Every $N \in \mathcal{N}(x)$ contains x
(4) If $V \in \mathcal{N}(x)$ then there exists $W \in \mathcal{N}(x)$ such that for all $y \in W$, $V \in \mathcal{N}(y)$.

Conversely, let X be some set. If for each $x \in X$ there is a family $\mathcal{N}(x)$ of subsets of X that satisfy the above neighborhood axioms then there is a uniquely determined topology on X for which the families $\mathcal{N}(x)$ are the neighborhoods of the points x. For this a subset $U \subset X$ is open if and only if for each $x \in U$ we have $U \in \mathcal{N}(x)$.

It can be shown that a sequence $\{x_n\}$ in a TVS is a Cauchy sequence if and only if for every neighborhood U of 0 there is a number N_U such that $x_l - x_k \in U$ for all $k, l \geq N_U$.

A relatively nice situation is when V has a norm that induces the topology. Recall that a **norm** is a function $\|.\| : \mathsf{V} \to \mathbb{R}$ such that for all $v, w \in \mathsf{V}$ we have

(1) $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = 0$,
(2) $\|v + w\| \leq \|v\| + \|w\|$,
(3) $\|\alpha v\| = |\alpha| \, \|v\|$ for all $\alpha \in \mathbb{R}$.

In this case we have a metric on V given by $\mathrm{dist}(v, w) := \|v - w\|$. A **seminorm** is a function $\|.\| : \mathsf{V} \to \mathbb{R}$ such that 2) and 3) hold but instead of 1) we require only that $\|v\| \geq 0$.

Using a norm, we may define a metric $d(x, y) := \|x - y\|$. A **normed space** V is a vector space together with a norm. It becomes a TVS with the metric topology given by a norm.

Definition 0.18. A linear map $\ell : \mathsf{V} \to \mathsf{W}$ between normed spaces is called **bounded** if and only if there is a constant C such that for all $v \in \mathsf{V}$ we have $\|\ell v\|_{\mathsf{W}} \leq C \|v\|_{\mathsf{V}}$. If ℓ is bounded then the smallest such constant C is

$$\|\ell\| := \sup \frac{\|\ell v\|_{\mathsf{W}}}{\|v\|_{\mathsf{V}}} = \sup\{\|\ell v\|_{\mathsf{W}} : \|v\|_{\mathsf{V}} \leq 1\}.$$

The set of all bounded linear maps $\mathsf{V} \to \mathsf{W}$ is denoted $\mathcal{B}(V, W)$. The vector space $\mathcal{B}(V, W)$ is itself a normed space with the norm given as above.

Lemma 0.19. *A linear map is bounded if and only if it is continuous.*

0.6.1. Topology of Finite Dimensional TVS. The usual topology on $\mathbb{R}^n$ is given by the norm $\|x\| = \sqrt{x \cdot x}$. Many other norms give the same topology. In fact we have the following

Theorem 0.20. *Given $\mathbb{R}^n$ the usual topology. Then if V is any finite dimensional (Hausdorff) topological vector space of dimension n, there is a linear homeomorphism $\mathbb{R}^n \to \mathsf{V}$. Thus there is a unique topology on V that gives it the structure of a (Hausdorff) topological vector space.*

Proof. Pick a basis $v_1, ..., v_n$ for V. Let $T : \mathbb{R}^n \to \mathsf{V}$ be defined by

$$(a_1, ..., a_n) \mapsto a_1 v_1 + \cdots + a_n v_n$$

Since V is a TVS, this map is continuous and it is obviously a linear isomorphism. We must show that T^{-1} is continuous. By Exercise 0.17, it suffices to show that T is continuous at $0 \in \mathbb{R}^n$. Let $B := \overline{B}_1$ be the closed unit ball in $\mathbb{R}^n$. The set $T(B)$ is compact and hence closed (since V is Hausdorff). We have a continuous bijection $T|_B : B \to T(B)$ of compact Hausdorff spaces and so it must be a homeomorphism by Exercise **??**. The key point is that if we can show that $T(B)$ is a closed neighborhood of 0 (i.e. is has a nonempty interior containing 0) then the continuity of $T|_B$ at 0 imply the continuity of T at 0 and hence everywhere. We now show this to be the case. Let $S \subset B$ the unit sphere which is the boundary of B. Then $\mathsf{V} \backslash T(S)$ is open and contains 0. By continuity of scaling, there is an $\varepsilon > 0$ and an open set $U \subset \mathsf{V}$ containing $0 \in \mathsf{V}$ such that $(-\varepsilon, \varepsilon)U \subset \mathsf{V} \backslash T(S)$. By replacing U by $2^{-1}\varepsilon U$ if necessary, we may take $\varepsilon = 2$ so that $[-1, 1]U \subset (-2, 2)U \subset \mathsf{V} \backslash T(S)$.

We now claim that $U \subset T(B^o)$ where B^o is the interior of B. Suppose not. Then there must be a $v_0 \in U$ with $v_0 \notin T(B^o)$. In fact, since $v_0 \notin T(S)$ we must have $v_0 \notin T(B)$. Now $t \to T^{-1}(tv_0)$ gives a continuous map $[0, 1] \to \mathbb{R}^n$ connecting $0 \in B^o$ to $T^{-1}(v_0) \in \mathbb{R}^n \backslash B$. But since $[-1, 1]U \subset \mathsf{V} \backslash T(S)$ the curve never hits S! Since this is clearly impossible we conclude that $0 \in U \subset T(B^o)$ so that $T(B)$ is a (closed) neighborhood of 0 and the result follows. $\square$

Definition 0.21. A **locally convex topological vector space** V is a TVS such that its topology is generated by a family of seminorms $\{\|.\|_\alpha\}_\alpha$. This means that we give V the weakest topology such that all $\|.\|_\alpha$ are continuous. Since we have taken a TVS to be Hausdorff we require that the family of seminorms is **sufficient** in the sense that for each $x \in \mathsf{V}$ we have $\bigcap\{x : \|x\|_\alpha = 0\} = \emptyset$. A locally convex topological vector space is sometimes called a locally convex space and so we abbreviate the latter to **LCS**.

Example 0.22. Let Ω be an open set in $\mathbb{R}^n$ or any manifold. Let $C(\Omega)$ be the set of continuous real valued functions on Ω. For each $x \in \Omega$ define a seminorm ρ_x on $C(\Omega)$ by $\rho_x(f) = |f(x)|$. This family of seminorms makes $C(\Omega)$ a topological vector space. In this topology, convergence is pointwise convergence. Also, $C(\Omega)$ is not complete with this TVS structure.

Definition 0.23. An LCS that is complete (every Cauchy sequence converges) is called a **Frechet space**.

Definition 0.24. A complete normed space is called a **Banach space**.

Example 0.25. Suppose that (X, μ) is a σ-finite measure space and let $p \geq 1$. The set $L^p(X, \mu)$ of all measurable functions $f : X \to \mathbb{C}$ such that $\int |f|^p \, d\mu \leq \infty$ is a Banach space with the norm $\|f\| := \left(\int |f|^p \, d\mu\right)^{1/p}$. Technically, functions which are equal almost everywhere must be identified.

Example 0.26. The space $C_b(\Omega)$ of bounded continuous functions on Ω is a Banach space with norm given by $\|f\|_\infty := \sup_{x\in\Omega} |f(x)|$.

Example 0.27. Once again let Ω be an open subset of $\mathbb{R}^n$. For each compact $K \subset \Omega$ we have a seminorm on $C(\Omega)$ defined by $\|f\|_K := \sup_{x\in K} |f(x)|$. The corresponding convergence is the uniform convergence on compact subsets of Ω. It is often useful to notice that the same topology can be obtained by using $\|f\|_{K_i}$ obtained from a countable sequence of nested compact sets $K_1 \subset K_2 \subset ...$ such that

$$\bigcup K_n = \Omega.$$

Such a sequence is called an exhaustion of Ω.

If we have topological vector space V and a closed subspace S, then we can form the quotient V/S. The quotient can be turned in to a normed space by introducing as norm

$$\|[x]\|_{\mathrm{V/S}} := \inf_{v\in[x]} \|v\| .$$

If S is not closed then this only defines a seminorm.

Theorem 0.28. *If* V *is a Banach space and* S *a closed (linear) subspace then* V/S *is a Banach space with the norm* $\|.\|_{\mathrm{V/S}}$ *defined above.*

Proof. Let x_n be a sequence in V such that $\{[x_n]\}$ is a Cauchy sequence in V/S. Choose a subsequence such that $\|[x_n] - [x_{n+1}]\| \leq 1/2^n$ for $n = 1, 2, ...$ Setting s_1 equal to zero we find $s_2 \in$ S such that $\|x_1 - (x_2 + s_2)\| \leq 1/2^2$ and continuing inductively define a sequence s_i such that such that $\{x_n + s_n\}$ is a Cauchy sequence in V. Thus there is an element $y \in$ V with

$x_n + s_n \to y$. But since the quotient map is norm decreasing, the sequence $[x_n + s_n] = [x_n]$ must also converge;

$$[x_n] \to [y].$$

□

Remark 0.29. It is also true that if S is closed and V/S is a Banach space then so is V.

0.6.2. Hilbert Spaces.

Definition 0.30. A pre-Hilbert space $\mathcal{H}$ is a complex vector space with a Hermitian inner product $\langle .,. \rangle$. A Hermitian inner product is a bilinear form with the following properties:

1) $\langle v, w \rangle = \overline{\langle w, v \rangle}$
2) $\langle v, \alpha w_1 + \beta w_2 \rangle = \alpha \langle v, w_1 \rangle + \beta \langle v, w_2 \rangle$
3) $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ only if $v = 0$.

The inner product gives rise to an associate norm $\|v\| := \langle v, v \rangle^{1/2}$ and so every pre-Hilbert space is a normed space. If a pre-Hilbert space is complete with respect to this norm then we call it a **Hilbert space**.

One of the most fundamental properties of a Hilbert space is the projection property.

Theorem 0.31. *If K is a convex, closed subset of a Hilbert space $\mathcal{H}$, then for any given $x \in \mathcal{H}$ there is a unique element $p_K(x) \in \mathcal{H}$ which minimizes the distance* $\operatorname{dist}(x, y) = \|x - y\|$ *over $y \in K$. That is*

$$\|x - p_K(x)\| = \inf_{y \in K} \|x - y\|.$$

If K is a closed linear subspace then the map $x \mapsto p_K(x)$ is a bounded linear operator with the projection property $p_K^2 = p_K$.

Definition 0.32. For any subset $S \in \mathcal{H}$ we have the orthogonal complement $S^\perp$ defined by

$$S^\perp = \{x \in \mathcal{H} : \langle x, s \rangle = 0 \text{ for all } s \in S\}.$$

$S^\perp$ is easily seen to be a linear subspace of $\mathcal{H}$. Since $\ell_s : x \mapsto \langle x, s \rangle$ is continuous for all s and since

$$S^\perp = \cap_s \ell_s^{-1}(0)$$

we see that $S^\perp$ is closed. Now notice that since by definition

$$\|x - P_s x\|^2 \leq \|x - P_s x - \lambda s\|^2$$

for any $s \in S$ and any real λ we have $\|x - P_s x\|^2 \leq \|x - P_s x\|^2 - 2\lambda \langle x - P_s x, s \rangle + \lambda^2 \|s\|^2$. Thus we see that $p(\lambda) := \|x - P_s x\|^2 - 2\lambda \langle x - P_s x, s \rangle +$

$\lambda^2 \|s\|^2$ is a polynomial in λ with a minimum at $\lambda = 0$. This forces $\langle x - P_s x, s\rangle = 0$ and so we see that $x = P_s x$. From this we see that any $x \in \mathcal{H}$ can be written as $x = x - P_s x + P_s x = s + s^\perp$. On the other hand it is easy to show that $S^\perp \cap S = 0$. Thus we have $\mathcal{H} = S \oplus S^\perp$ for any closed linear subspace $S \subset \mathcal{H}$. In particular the decomposition of any x as $s + s^\perp \in S \oplus S^\perp$ is unique.

0.7. Differential Calculus on Banach Spaces

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[3] 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.

Definition 0.33. 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 0.34. 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 $v \in \mathsf{V}$.

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

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 0.35. 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 0.36. 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 0.37. *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 0.38 (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 0.39. 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, $1 \le i \le k$ 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

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

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

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

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 0.40. 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

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

and so if $\|(v_1,v_2)-(a_1,a_2)\| \to 0$ then $\|v_i-a_i\| \to 0$ and we see that

$$\|\mu(v_1,v_2)-\mu(a_1,a_2)\| \to 0.$$

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

(⇒) 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 $\|\mu(v_1,v_2)\|\le 1$ 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 $\|\mu(z_1,z_2)\|\le 1$. The case $(v_1,v_2)=(0,0)$ is trivial so assume $(v_1,v_2)\ne(0,0)$. Then we have

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

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

Notation 0.41. 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 0.42. 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 0.43. 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

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

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 0.44. 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 0.45. 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 0.44.

Notation 0.46. 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 0.47. 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 0.48. 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 0.49. 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[4] $c : I \subset \mathbb{R} \to \mathsf{E}$ at $t_0 \in I$, as a velocity vector and so we are identifying $Dc|_{t_0} \in L(\mathbb{R}, \mathsf{E})$ with

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

$Dc|_{t_0} \cdot 1 \in \mathsf{E}$. 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

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

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 $\|f\| = \sup_{a \le t \le b} \|f(t)\|$. 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

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

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 0.50. *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 0.51. *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 0.52. 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 0.53. 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 0.44. On the other hand, we have

Proposition 0.54. 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 $f^i : U \to \mathbb{R}$, $1 \le i \le m$ 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.

0.7.1. Chain Rule, Product rule and Taylor's Theorem.

Theorem 0.55 (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, $\mathrm{comp} : L(\mathsf{E}_1, \mathsf{E}_2) \times L(\mathsf{E}_2, \mathsf{E}_3) \to L(\mathsf{E}_1, \mathsf{E}_3)$, is bounded. Define a map on U_1 by

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

Consider the composition $\mathrm{comp} \circ m_{f,g}$. Since f and g are at least C^1 this composite map is clearly continuous. Now we may proceed inductively. Consider the r^{th} statement:

$$\text{compositions of } C^r \text{ maps are } C^r$$

Suppose f and g are C^{r+1} then Df is C^r and $Dg \circ f$ is C^r by the inductive hypothesis so that $m_{f,g}$ is C^r. A bounded bilinear functional is

C^∞. Thus comp is C^∞ and by examining comp $\circ m_{f,g}$ we see that the result follows. □

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

Lemma 0.56. *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 0.57. *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. $\square$

Theorem 0.58. *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 0.59 (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.

0.7.2. Local theory of differentiable maps.

0.7.2.1. *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 0.60. Let E and F be Banach spaces. A map will be called a C^r **diffeomorphism near** p if there is some open set $U \subset \mathrm{dom}(f)$ 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 0.61. 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 0.62 (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. $\square$

Remark 0.63. 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 0.64. 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 0.65. *Let* E *be a Banach space. If* $f : \mathsf{E} \to \mathsf{E}$ *is continuous map that is Lipschitz continuous with constant* k *where* $0 \le k < 1$, *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

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

Solving this inequality we get

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

$\square$

Lemma 0.66. *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 $\|A_1 \circ A_2\| \leq \|A_1\| \, \|A_2\|$ 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 0.67. *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

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

□

Theorem 0.68 (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* $Df^{-1}\big|_y = (Df|_{f^{-1}(y)})^{-1}$.

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 $\|g(x_2) - g(x_1)\| \le \frac{1}{2}\|x_2 - x_1\|$ for $x_2, x_1 \in \overline{B}(0, r)$ and that $g(\overline{B}(0, r)) \subset \overline{B}(0, r/2)$. Let $y_0 \in \overline{B}(0, r/2)$. It is not hard to show that the map $c : x \mapsto y_0 + x - f(x)$ is a contraction mapping $c : \overline{B}(0, r) \to \overline{B}(0, r)$ with constant $\frac{1}{2}$. The contraction mapping principle 0.62 says that c has a unique fixed point $x_0 \in \overline{B}(0, r)$. But $c(x_0) = x_0$ just translates to $y_0 + x_0 - f(x_0) = x_0$ and then $f(x_0) = y_0$. So x_0 is the unique element of $\overline{B}(0, r)$ satisfying this equation. But then since $y_0 \in \overline{B}(0, r/2)$ was an arbitrary element of $\overline{B}(0, r/2)$ it follows that the restriction $f : \overline{B}(0, r/2) \to f(\overline{B}(0, r/2))$ is invertible. But f^{-1} is also continuous since

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

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

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

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 0.67. This last step is left to the reader. □

Corollary 0.69. Let $U \subset \mathsf{E}$ be an open set. Suppose that $f : U \to \mathsf{F}$ is differentiable with $Df(p) : \mathsf{E} \to \mathsf{F}$ a (bounded) linear isomorphism for each $p \in U$. Then f is a local diffeomorphism.

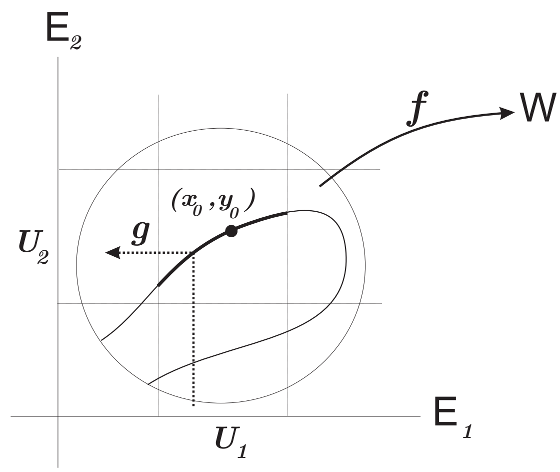
Example 0.70. Consider the map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ given by

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

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 $\phi|_{\mathbb{R}^2\setminus\{(0,0)\}}$ 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 0.71 (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

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

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 $\operatorname{graph}(g) \subset (U_1 \times U_2) \cap f^{-1}(0)$.

We now show that $(U_1 \times U_2) \cap f^{-1}(0) \subset \operatorname{graph}(g)$. 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 $(x,y) \in \operatorname{graph}(g)$. □

The simplest situation is that of a function $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(a,b) = 0$ and $D_2 f(a,b) \neq 0$. 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 0.72. Find a function $f : \mathbb{R}^2 \to \mathbb{R}$ with $D_2 f(0,0) = 0$ 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 $f(x, g(x)) = 0$.

0.7.3. Immersion.

Theorem 0.73. *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 0.74. 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 $\left(a^1, ..., a^k\right) \mapsto \left(a^1, ..., a^k, 0, ..., 0\right)$.

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 $A \circ f$. □

0.7.4. Submersion.

Theorem 0.75. *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 0.76. 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

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

0.7.5. 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 0.77 (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 $\operatorname{rank}(f) = \operatorname{rank}(f \circ h^{-1}) = r$ near 0. On the said (possibly smaller) neighborhood we now define the map $g_2 : (\mathbb{R}^m, q) \to (\mathbb{R}^m, 0)$ by

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

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

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

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

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

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

0.7.6. Existence and uniqueness for differential equations.

Theorem 0.78. *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 0.79. *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 $v \in \mathsf{E}$.

0.8. 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:

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

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 $\left.\delta F\right|_f h = \left(D_h F\right)[f]$. 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

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

It is well known that if $\delta S|_c(h) = 0$ for every h then c is a linear function; $c(t) = at + b$. The condition $\delta S|_c(h) = 0 = 0$ (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

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

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

$$\delta S = \int \frac{\delta S}{\delta q(t)} \delta q(t) dt \tag{0.1}$$

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

$$df|_p = \sum \frac{\partial f}{\partial x^i}(p)\, dx^i\big|_p \tag{0.2}$$

to some vector h we get

$$f'(p)h = \sum \frac{\partial f}{\partial x^i}(p)h^i. \tag{0.3}$$

In other words, $df|_p = D_h f(p) = \nabla f \cdot h = f'(p)$. Too many notations for the same concept. Equation 0.2 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:

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

Roughly, $\frac{\delta S}{\delta q(t)}$ is taken to be whatever function (or distribution) makes the equation 0.1 true. We often see the following type of calculation

$$\begin{aligned} \delta S &= \delta \int L dt \\ &= \int \left(\frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q}\right) dt \\ &= \int \left\{\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right\} \delta q dt \end{aligned} \tag{0.4}$$

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

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

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 $\delta S|_c$ instead of $\delta S|_{q(.)}$. The point is that the notation $\delta S|_q$ might leave one thinking that $q \in \mathbb{R}$ (which it is under one interpretation!) but then $\delta S|_q$ 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

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

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 $\delta S|_{q(.)} = 0$ (or even $\delta S|_{q(.)}\, h = 0$ 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 0.80. 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 $df = \frac{\partial f}{\partial xi} dx^i$ 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 0.81. 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 0.82. 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

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

0.8.1. 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 0.83. Let E and F and F_0 be as in the previous example. We define two functionals

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

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

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

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 0.84. *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 $q : \mathsf{E} \to L$

$$\begin{aligned} p^2 &= p \text{ and } q^2 = q \\ p + q &= id \end{aligned}$$

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

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

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.

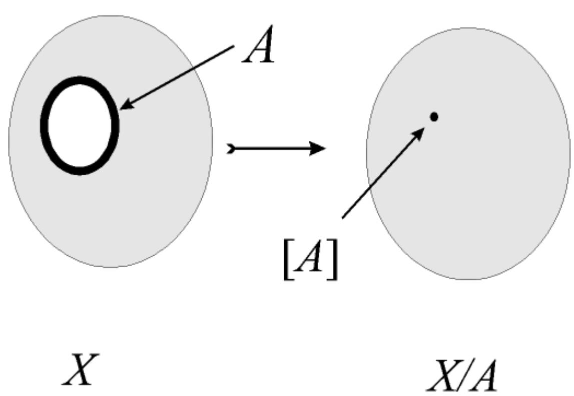
Now h and $\varepsilon(h)$ are in complementary subspaces and so ϕ must be injective. Thus its restriction to the set $V := \{x_0+h+\varepsilon(h) : h \in K \cap B_\delta(0)$ is invertible and in fact we have $\phi^{-1}(x_0 + h + \varepsilon(h)) = x_0 + h$. That V is open follows from the way we have 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. $\square$

0.9. Attaching Spaces and Quotient Topology

Suppose that we have a topological space X and a surjective set map $f : X \to S$ onto some set S. We may endow S with a natural topology according to the following recipe. A subset $U \subset S$ is defined to be open if and only if $f^{-1}(U)$ is an open subset of X. This is particularly useful when we have some equivalence relation on X which allows us to consider the set of equivalence classes $X/\sim$. In this case we have the canonical map $\varrho : X \to X/\sim$ that takes $x \in X$ to its equivalence class $[x]$. The quotient topology is then given as before by the requirement that $U \subset X/\sim$ is open if and only if and only if $\varrho^{-1}(U)$ is open in X. A common application of this idea is the identification of a subspace to a point. Here we have some subspace $A \subset X$ and the equivalence relation is given by the following two requirements:

$$\begin{array}{ll} \text{If } x \in X \backslash A & \text{then } x \sim y \text{ only if } x = y \\ \text{If } x \in A & \text{then } x \sim y \text{ for any } y \in A \end{array}$$

In other words, every element of A is identified with every other element of A. We often denote this space by X/A.



A hole is removed by identification

It is not difficult to verify that if X is Hausdorff (resp. normal) and A is closed then X/A is Hausdorff (resp. normal). The identification of a subset to a point need not simplify the topology but may also complicate the topology as shown in figure 0.1.

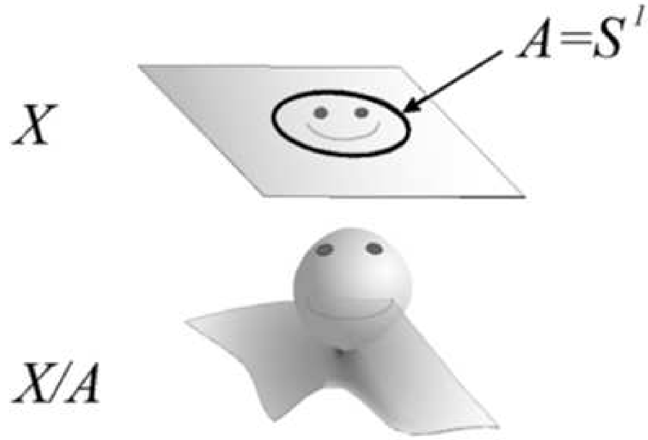
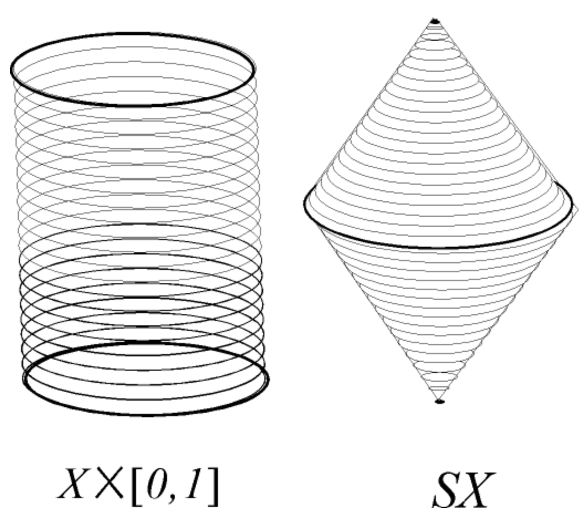
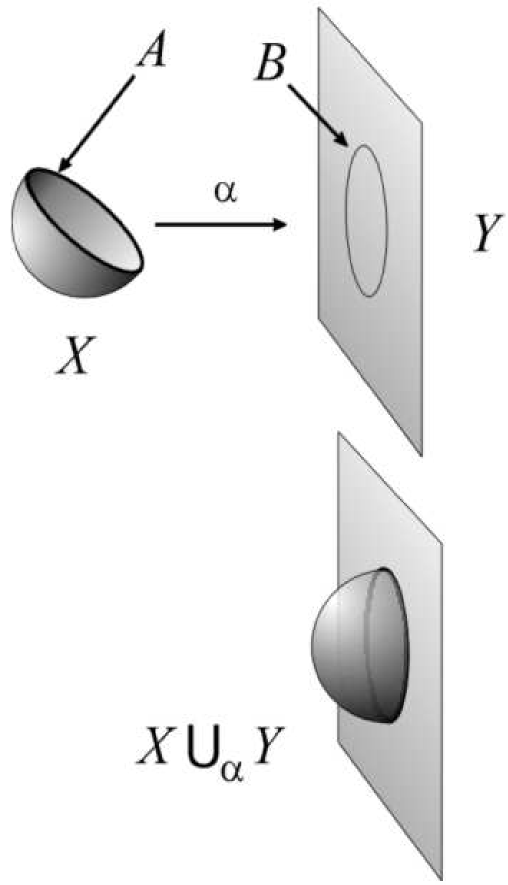


Figure 0.1. creation of a "hole"



An important example of this construction is the suspension. If X is a topological space then we define its suspension SX to be $(X \times [0,1])/A$ where $A := (X \times \{0\}) \cup (X \times \{1\})$. For example it is easy to see that $SS^1 \cong S^2$. More generally, $SS^{n-1} \cong S^n$.

Consider two topological spaces X and Y and a closed subset $A \subset X$. Suppose that we have a map $\alpha : A \to B \subset Y$. Using this map we may define an equivalence relation on the disjoint union $X \bigsqcup Y$ that is given by requiring that $x \sim \alpha(x)$ for $x \in A$. The resulting topological space is denoted $X \cup_\alpha Y$. A 2- cell is a topological space homeomorphic to a closed disk D in $\mathbb{R}^2$. Figure 0.9 shows the idea graphically



Attaching by a map

Another useful construction is that of the mapping cylinder of a map $f : X \to Y$. First we transfer the map f to a map on the base $X \times \{0\}$ of the cylinder $X \times I$ by

$$f(x, 0) := f(x)$$

and then we form the quotient $Y \cup_f (X \times I)$. We denote this quotient by M_f and call it the mapping cylinder of f.

0.10. 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.

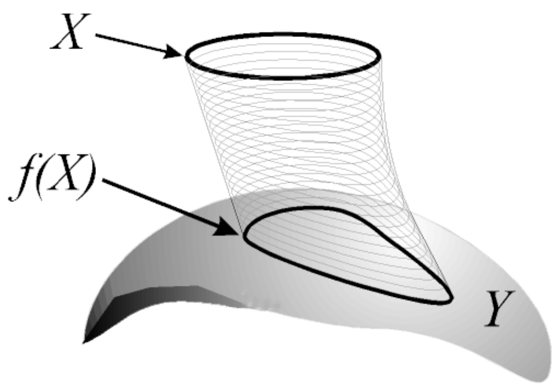


Figure 0.2. Mapping Cylinder

b) $f(x) = x^t A x$ 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 $D = \{|x| \leq 1\}$.

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 $\|f\|_k^S := \sum_{|\alpha| \le k} \sup_{x \in S} \left| \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x) \right|$. a) Show that (1) $\|rf\|_k^S = |r| \, \|f\|_k^S$ for any $r \in \mathbb{R}$, (2) $\|f_1 + f_2\|_k^S \le \|f_1\|_k^S + \|f_2\|_k^S$ for any $f_1, f_2 \in C^k(U)$, (3) $\|fg\|_k^S \le \|f\|_k^S \, \|g\|_k^S$ 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 1 Supplement

1.1. Comments and Errata

1.1.1. Comments. (Nothing yet)

1.1.2. Errata. (Nothing yet)

1.2. Rough Ideas I

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

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

After imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of n-tuples of numbers. In fact, since a Euclidean space in this sense is an object of intuition (at least in 2d and 3d) some may insist that to be sure such a space of points really exists that we should in fact *start* with $\mathbb{R}^n$ and "forget" the origin and all the vector space structure while retaining the notion of point and distance. The coordinatization of Euclidean space is then just a "remembering" of this forgotten structure. Thus our coordinates arise from a map $x : E^n \to \mathbb{R}^n$ which is just the identity map. This approach has much to recommend it and we shall more or less follow this canonical path. There is at least one regrettable aspect to this approach which is the psychological effect that results from always picturing Euclidean space as $\mathbb{R}^n$. The result is that when we introduce the idea of a manifold and describe what coordinates are in that setting it seems like a new idea. It might seem that this is a big abstraction and when the definition of charts and atlases and so on appear a certain notational fastidiousness sets in that somehow creates a psychological gap between open sets in $\mathbb{R}^n$ and the abstract space that we coordinatize. But what is now lost from sight is that we have already been dealing with an abstract manifolds E^n which we have identified with $\mathbb{R}^n$ via exactly such an idea of coordinates. Euclidean space was already always a manifold. But Euclidean space could just as easily support other coordinates such as spherical coordinates. In this case also we just get to work in the new coordinates without pondering what the exact definition of a coordinate system should be. We also picture the coordinates as being "on the space" and not as a map to some other Euclidean space. But how else should coordinates be defined? Consider how naturally a calculus student works in various coordinate systems. What competent calculus student would waste time thinking of polar coordinates as given by a map $E^2 \to \mathbb{R}^2$ (defined on a proper open subset of course) and then wonder whether something like $drd\theta$ lives on the original E^n or in the image of the coordinate map $E^n \to \mathbb{R}^n$? It has become an unfortunate consequence of the modern viewpoint that simple geometric ideas are lost from the notation. Ideas that allow one to think about "quantities and their variations" and then comfortably write things like

$$rdr \wedge d\theta = dx \wedge dy$$

without wondering if this shouldn't be a "pullback" $\psi_{12}^*(rdr \wedge d\theta) = dx \wedge dy$ where ψ_{12} is the change of coordinate map $\mathbb{R}^2 \to \mathbb{R}^2$ given by

$$x(r, \theta) = r \cos \theta$$
$$y(r, \theta) = r \cos \theta$$

Of course, the experienced differential geometer understands the various meanings and gleans from the context what the proper meaning of a notation should be. The new student of manifold theory on the other hand, is faced with a pedagogy that teaches notation, trains one to examine each equation for logical self consistency, but fails to teach geometric intuition. Having made this complaint the author must confess that he too will use the modern notation and will not stray far from standard practice. The student must learn how differential geometry is actually done. These remarks are meant to encourage the student to stop and seek the simplest most intuitive viewpoint whenever feeling overwhelmed by notation. The student is encouraged to experiment with abbreviated personal notation when checking calculations and to draw diagrams and schematics that encode the geometric ideas whenever possible. The maxim should be "Let the geometry write the equations".

Now this approach works fine as long as intuition is not allowed to overcome logic and thereby mislead us. On occasion intuition does mislead us and this is where the pedantic notation and the various abstractions can save us from error. The abstractions and pedantic notations are related to another issue: Why introduce abstract structures such as sheaves and modules and so forth. Couldn't differential geometry do without these notions being explicitly introduced? Perhaps, but there is a new force on the horizon the demands that we have a very firm grip on the logical and algebraic aspects of differential geometry. Namely, noncommutative geometry. Here the idea is to replace the algebra of continuous (and the subalgebra of smooth) functions by a noncommutative algebra. In this new setting one tries to study analogues of the objects of ordinary differential geometry. But this means having a very good grip on the exact role in ordinary geometry of function algebras and the various modules over these algebras. Things that seem so obvious like the naturality with respect to restrictions and so forth cannot be taken for granted, indeed, may not even have a meaning in the new world of noncommutative geometry. This is one reason that we shall not shy away from certain ideas such as the relation between finitely generated modules, locally free sheaves and vector bundles. We also intend to be a bit more pedantic than usual when we do tensor analysis[1]. What of the difference between $V \otimes V^*$ and $V^* \otimes V$; should they be identified?

[1] What about the choice of writing coefficients on the right or left in linear expressions. Should the tensor algebra distinguish $T_i{}^j\big|_{kl}$ from T^j_{ikl}? Is the covariant differential ∇T of a tensor $T \in \Gamma(TM \otimes TM^*)$ be thought of as an element of $\Gamma(TM \otimes TM^* \otimes TM^*)$ as the notation $T^i_{j;k}$ suggests or as an element of $\Gamma(TM^* \otimes TM \otimes TM^*)$ as the notation $\nabla_i T^i_j$ suggests? What effect does this have on the sign convention for the curvatuire tensor. If you think it doesn't matter ask yourself if you are still sure about that in the noncommutative setting.

So, as we said, after imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers. We will envision there to be a copy $\mathbb{R}^n_p$ of $\mathbb{R}^n$ at each of its points $p \in \mathbb{R}^n$. One way to handle this is to set $\mathbb{R}^n_p = \{p\} \times \mathbb{R}^n$ so that taking all points of a domain$U$ into account we should consider $U \times \mathbb{R}^n$. Thus if $(p, v) \in U \times \mathbb{R}^n$ we take this to represent the vector at p which is parallel to v. The elements of $\mathbb{R}^n_p$ are to be thought of as the vectors based at p, that is, the "tangent vectors". These tangent spaces are related to each other by the obvious notion of vectors being parallel (this is exactly what is not generally possible for tangents spaces of a manifold). For the standard basis vectors e_j (relative to the coordinates x_i) taken as being based at p we often write $\left.\frac{\partial}{\partial x_i}\right|_p$ instead of (p, e_j) and this has the convenient second interpretation as a differential operator acting on C^∞functions defined near $p \in \mathbb{R}^n$. Namely,

$$\left.\frac{\partial}{\partial x_i}\right|_p f = \frac{\partial f}{\partial x_i}(p).$$

An n-tuple of C^∞functions $X^1, ..., X^n$ defines a C^∞vector field $X = \sum X^i \frac{\partial}{\partial x_i}$ whose value at p is $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$. Thus a vector field assigns to each p in its domain, an open set U, a vector $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$ at p. We may also think of vector field as a differential operator via

$$f \mapsto Xf \in C^\infty(U)$$

$$(Xf)(p) := \sum X^i(p) \frac{\partial f}{\partial x_i}(p)$$

Example 1.1. $X = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}$ is a vector field defined on $U = \mathbb{R}^2 - \{0\}$ and $(Xf)(x, y) = y\frac{\partial f}{\partial x}(x, y) - x\frac{\partial f}{\partial y}(x, y)$.

Notice that we may certainly add vector fields defined over the same open set as well as multiply by functions defined there:

$$(fX + gY)(p) = f(p)X(p) + g(p)X(p)$$

The familiar expression $df = \frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_n}dx_n$ has the intuitive interpretation expressing how small changes in the variables of a function give rise to small changes in the value of the function. Two questions should come to mind. First, "what does 'small' mean and how small is small enough?" Second, "which direction are we moving in the coordinate" space? The answer to these questions lead to the more sophisticated interpretation of df as being a linear functional on each tangent space. Thus we must choose a direction v_p at $p \in \mathbb{R}^n$ and then $df(v_p)$ is a number depending linearly on our choice of vector v_p. The definition is determined by $dx_i(\frac{\partial}{\partial x_i}\big|_p) = \delta_{ij}$. In fact, this shall be the basis of our definition of df at p. We want

$$df|_p\,(\frac{\partial}{\partial x_i}\bigg|_p) := \frac{\partial f}{\partial x_i}(p).$$

Now any vector at p may be written $v_p = \sum_{i=1}^n v^i\ \frac{\partial}{\partial x_i}\big|_p$ which invites us to use v_p as a differential operator (at p):

$$v_p f := \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p) \in \mathbb{R}$$

This consistent with our previous statement about a vector field being a differential operator simply because $X(p) = X_p$ is a vector at p for every $p \in U$. This is just the directional derivative. In fact we also see that

$$\begin{aligned} df|_p\,(v_p) &= \sum_j \frac{\partial f}{\partial x_j}(p)dx_j\left(\sum_{i=1}^n v^i\ \frac{\partial}{\partial x_i}\bigg|_p\right) \\ &= \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p) = v_p f \end{aligned}$$

so that our choices lead to the following definition:

Definition 1.2. Let f be a C^∞function on an open subset U of $\mathbb{R}^n$. By the symbol df we mean a family of maps $df|_p$ with p varying over the domain U of f and where each such map is a linear functional of tangent vectors based at p given by $df|_p\,(v_p) = v_p f = \sum_{i=1}^n v^i \frac{\partial f}{\partial x_i}(p)$.

Definition 1.3. More generally, a smooth 1-form α on U is a family of linear functionals $\alpha_p : T_p\mathbb{R}^n \to \mathbb{R}$ with $p \in U$ that is smooth is the sense that $\alpha_p(\frac{\partial}{\partial x_i}\big|_p)$ is a smooth function of p for all i.

From this last definition it follows that if $X = X^i \frac{\partial}{\partial x_i}$ is a smooth vector field then $\alpha(X)(p) := \alpha_p(X_p)$ defines a smooth function of p. Thus an alternative way to view a $1-$form is as a map $\alpha : X \mapsto \alpha(X)$ that is defined on vector fields and linear over the algebra of smooth functions $C^\infty(U)$:

$$\alpha(fX + gY) = f\alpha(X) + g\alpha(Y).$$

Fixing a problem. It is at this point that we want to destroy the privilege of the rectangular coordinates and express our objects in an arbitrary coordinate system smoothly related to the existing coordinates. This means that for any two such coordinate systems, say $u^1, ..., u^n$ and $y^1, ..., y^n$ we want to have the ability to express fields and forms in either system and have for instance

$$X^i_{(y)}\frac{\partial}{\partial y_i} = X = X^i_{(u)}\frac{\partial}{\partial u_i}$$

for appropriate functions $X^i_{(y)}$, $X^i_{(u)}$. This equation only makes sense on the overlap of the domains of the coordinate systems. To be consistent with the chain rule we must have

$$\frac{\partial}{\partial y^i} = \frac{\partial u^j}{\partial y^i}\frac{\partial}{\partial u^j}$$

which then forces the familiar transformation law:

$$\sum \frac{\partial u^j}{\partial y^i} X^i_{(y)} = X^i_{(u)}$$

We think of $X^i_{(y)}$ and $X^i_{(u)}$ as referring to, or representing, the same geometric reality from the point of view of two different coordinate systems. No big deal right? Well, how about the fact, that there is this underlying abstract space that we are coordinatizing? That too is no big deal. We were always doing it in calculus anyway. What about the fact that the coordinate systems aren't defined as a 1-1 correspondence with the points of the space unless we leave out some points in the space? For example, polar coordinates must exclude the positive x-axis and the origin in order to avoid ambiguity in θ and have a nice open domain. Well if this is all fine then we may as well imagine other abstract spaces that support coordinates in this way. This is manifold theory. We don't have to look far for an example of a manifold other than Euclidean space. Any surface such as the sphere will do. We can talk about 1-forms like say $\alpha = \theta d\phi + \phi \sin(\theta) d\theta$, or a vector field tangent to the sphere $\theta \sin(\phi)\frac{\partial}{\partial\theta} + \theta^2\frac{\partial}{\partial\phi}$ and so on (just pulling things out of a hat). We just have to be clear about how these arise and most of all how to change to a new coordinate expression for the same object. This is the approach of tensor analysis. An object called a 2-tensor T is represented in two different coordinate systems as for instance

$$\sum T^{ij}_{(y)}\frac{\partial}{\partial y^i}\otimes\frac{\partial}{\partial y^j} = \sum T^{ij}_{(u)}\frac{\partial}{\partial u^i}\otimes\frac{\partial}{\partial u^j}$$

where all we really need to know for many purposes the transformation law

$$T^{ij}_{(y)} = \sum_{r,s} T^{rs}_{(u)}\frac{\partial y^i}{\partial u^r}\frac{\partial y^i}{\partial u^s}.$$

Then either expression is referring to the same abstract tensor T. This is just a preview but it highlight the approach wherein a transformation laws play a defining role.

In order to understand modern physics and some of the best mathematics it is necessary to introduce the notion of a space (or spacetime) that only locally has the (topological) features of a vector space like $\mathbb{R}^n$. Examples of two dimensional manifolds include the sphere or any of the other closed smooth surfaces in $\mathbb{R}^3$ such a torus. These are each locally like $\mathbb{R}^2$ and when sitting in space in a nice smooth way like we usually picture them, they support coordinates systems that allow us to do calculus on them. The reader will no doubt be comfortable with the idea that it makes sense to talk about directional rates of change in say a temperature distribution on a sphere representing the earth.

For a higher dimensional example we have the $3-$sphere S^3 which is the hypersurface in $\mathbb{R}^4$ given by the equation $x^2 + y^2 + z^2 + w^2 = 1$.

For various reasons, we would like coordinate functions to be defined on open sets. For closed surfaces like the sphere, it is not possible to define nice coordinates that are defined on the whole surface. By nice we mean that together the coordinate functions, say, θ, ϕ should combine to give a 1-1 correspondence with a subset of $\mathbb{R}^2$ that is continuous and has a continuous inverse. In general the best we can do is introduce several coordinate systems each defined on separate open subsets that together cover the surface. This will be the general idea for all manifolds.

Now suppose that we have some surface S and two coordinate systems

$$(\theta, \phi) : U_1 \to \mathbb{R}^2$$
$$(u, v) : U_2 \to \mathbb{R}^2$$

Imagine a real valued function f defined on S (think of f as temperature). Now if we write this function in coordinates (θ, ϕ) we have f represented by a function of two variables $f_1(\theta, \phi)$ and we may ask if this function is differentiable or not. On the other hand, f is given in (u, v) coordinates by a representative function $f_2(u, v)$. In order that our conclusions about differentiability at some point $p \in U_1 \cap U_2 \subset S$ should not depend on what coordinate system we use we had better have the coordinate systems themselves related differentiably. That is, we want the coordinate change functions in both directions to be differentiable. For example we may then relate the derivatives as they appear in different coordinates by chain rules expressions like

$$\frac{\partial f_1}{\partial \theta} = \frac{\partial f_2}{\partial u}\frac{\partial u}{\partial \theta} + \frac{\partial f_2}{\partial v}\frac{\partial v}{\partial \theta}$$

which have validity on coordinate overlaps. The simplest and most useful condition to require is that coordinates systems have C^∞ coordinate changes on the overlaps.

Definition 1.4. A set M is called a C^∞differentiable manifold of dimension n if M is covered by the domains of some family of coordinate mappings or charts $\{\mathtt{x}_\alpha : U_\alpha \to \mathbb{R}^n\}_{\alpha \in A}$ where $\mathtt{x}_\alpha = (x^1_\alpha, x^2_\alpha, ...x^n_\alpha)$. We require that the coordinate change maps $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ are continuously differentiable any number of times on their natural domains in $\mathbb{R}^n$. In other words, we require that the functions

$$\begin{aligned} x^1 &= x^1_\beta(x^1_\alpha, ..., x^n_\alpha) \\ x^2_\beta &= x^2_\beta(x^1_\alpha, ..., x^n_\alpha) \\ &\vdots \\ x^n_\beta &= x^n_\beta(x^1_\alpha, ..., x^n_\alpha) \end{aligned}$$

together give a C^∞ bijection where defined. The α and β are just indices from some set A and are just a notational convenience for naming the individual charts.

Note that we are employing the same type of abbreviations and abuse of notation as is common is every course on calculus where we often write things like $y = y(x)$. Namely, $(x^1_\alpha, ..., x^n_\alpha)$ denotes both an n-tuple of coordinate functions and an element of $\mathbb{R}^n$. Also, $x^1_\beta = x^1_\beta(x^1_\alpha, ..., x^n_\alpha)$ etc. could be thought of as an abbreviation for a functional relation that when evaluated at a point p on the manifold reads

$$(x^1_\beta(p), ...x^n_\beta(p)) = \mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}(x^1_\alpha(p), ..., x^n_\alpha(p)).$$

A function f on M will be deemed to be C^r if its representatives f_α are all C^r for every coordinate system $\mathtt{x}_\alpha = (x^1_\alpha, ..., x^n_\alpha)$ whose domain intersects the domain of f. Now recall our example of temperature on a surface. For an arbitrary pair of coordinate systems $\mathtt{x} = (x^1, ..., x^n)$ and $\mathtt{y} = (y^1, ..., y^n)$ the functions $f_1 := f \circ \mathtt{x}^{-1}$ and $f_2 := f \circ \mathtt{y}^{-1}$ represent the same function f with in the coordinate domains but the expressions $\frac{\partial f_1}{\partial x^i}$ and $\frac{\partial f_2}{\partial y^i}$ are not equal and do not refer to the same physical or geometric reality. The point is simply that because of our requirements on the smooth relatedness of our coordinate systems we know that on the overlap of the two coordinate systems if $f \circ \mathtt{x}^{-1}$ has continuous partial derivatives up to order k then the same will be true of $f \circ \mathtt{y}^{-1}$.

1.3. Pseudo-Groups and Models Spaces

Without much work we can generalize our definitions in such a way as to provide, as special cases, the definition of some common notions such as that *complex manifold* and *manifold with boundary*. If fact, we would also like 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 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 also study complex manifolds, foliated manifolds, manifolds with boundary, Hilbert manifolds and so on. Thus we need appropriate model spaces but also, significantly, we need a distinguished family on maps on the space. In this context the follow notion becomes useful:

Definition 1.5. 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 1.6. A **sub-pseudogroup** $\mathcal{S}$ of a pseudogroup is a subset of $\mathcal{G}$ and is also a pseudogroup (and so closed under composition and inverses).

We will be mainly interested in C^r-pseudogroups and the spaces that support them. Our main example will be the set $\mathcal{G}^r_{R^n}$ of all C^r maps between open subsets of $\mathbb{R}^n$. More generally, for a Banach space B we have the C^r-pseudo-group $\mathcal{G}^r_{\mathsf{B}}$ 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 1.7. A C^r- pseudogroup of transformations of a subset M of Banach space B is a pseudogroup arising as the restriction to M of a sub-pseudogroup of $\mathcal{G}^r_{\mathsf{M}}$. The set M with the relative topology and this C^r-pseudogroup is called a **model space**.

Example 1.8. Recall that a map $U \subset \mathbb{C}^n \to \mathbb{C}^n$ is holomorphic if the derivative (from the point of view of the underlying real space $\mathbb{R}^{2n}$) is in fact complex linear. A holomorphic map with holomorphic inverse is called biholomorphic. The set of all biholomorphic maps between open subsets of $\mathbb{C}^n$ is a pseudogroup. This is a C^r-pseudogroup for all r including $r = \omega$. In this case the subset M we restrict to is just $\mathbb{C}^n$ itself.

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 1.9. Let $\lambda \in \mathsf{M}^*$ be a continuous from 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 $\mathsf{M}_\lambda^+ = \{\mathsf{x} \in \mathsf{M}\colon\ \lambda(\mathsf{x}) \geq 0\}$ and $\mathsf{M}_\lambda^- = \{\mathsf{x} \in \mathsf{M}\colon\ \lambda(\mathsf{x}) \leq 0\}$ and $\partial\mathsf{M}_\lambda^+ = \partial\mathsf{M}_\lambda^- = \{\mathsf{x} \in \mathsf{M}\colon\ \lambda(\mathsf{x}) = 0\}$ is the kernel of λ. Clearly M_λ^+ and M_λ^- are homeomorphic and $\partial\mathsf{M}_\lambda^-$ is a closed subspace. [2]

Example 1.10. Let $\mathcal{G}^r_{\mathsf{M}_\lambda^-}$ be the restriction to M_λ^- of the set of C^r-diffeomorphisms ϕ from open subset of M to open subsets of M that have the following property

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

Notation 1.11. 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 study vector bundles.

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

Definition 1.12. 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 definition the model space M is a subset of a Banach space. If M 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 we just call the atlas a C^r atlas.

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

Exercise 1.13. Show that this definition of C^r atlas is the same as our original definition in the case where M is the finite dimensional Banach space $\mathbb{R}^n$.

In practice, a $\mathcal{G}-$manifold is just a space M (soon to be a topological manifold) together with an $\mathcal{G}-$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.

Now it is clear that every equivalence class of atlases contains a unique **maximal atlas** which is just the union of all the atlases in the equivalence class. Of course every member of the equivalence class determines the maximal atlas also–just toss in every possible compatible chart and we end up with the maximal atlas again.

Definition 1.14. A topological manifold M is called a C^r-**differentiable manifold** (or just C^r manifold) if it comes equipped with a differentiable structure. M is given the topology induced by the maximal atlas for the given C^r differentiable structure. Whenever we speak of a differentiable manifold we will have a fixed differentiable structure and therefore a maximal C^r-atlas $\mathcal{A}_M$ in mind. A chart from $\mathcal{A}_M$ will be called an **admissible chart**.

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

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

Definition 1.15. 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**.

Remark 1.16. It may be the case that there are two or more different differentiable structures on the same topological manifold. But see remark **??** below.

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

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

Remark 1.17. 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 verses y-coordinates" and so forth. Also, we use several rather self explanatory expressions such as "**coordinates", "coordinate charts**", "**coordinate systems"** and so onand 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 $\mathtt{x}(p) = 0 \in \mathsf{M}$.

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

Example 1.18. Each Banach space M is a differentiable manifold in a trivial way. Namely, there is a single chart that forms an atlas[3] which is just the identity map $\mathsf{M} \to \mathsf{M}$. In particular $\mathbb{R}^n$ with the usual coordinates

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

is a smooth manifold. Notice however that the map $\varepsilon : (x^1, x^2, ..., x^n) \mapsto ((x^1)^{1/3}, x^2, ..., x^n)$ is also a chart. It induces the usual topology again but the resulting maximal atlas is different! Thus we seem to have two manifolds $\mathbb{R}^n, \mathcal{A}_1$ and $\mathbb{R}^n, \mathcal{A}_2$. This is true but they are equivalent in another sense. Namely, they are diffeomorphic via the map ε. See definition **??** below. Actually, if V is any vector space with a basis $(f_1, ..., f_n)$ and dual basis $(f_1^*, ..., f_n^*)$ then one again, we have an atlas consisting of just one chart define on all of V defined by $\mathtt{x} : \mathrm{v} \mapsto (f_1^* \mathrm{v}, ..., f_n^* \mathrm{v}) \in \mathbb{R}^n$. On the other hand V may as well be 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 $M_1 \times M_2$. We may give $M_1 \times M_2$ a differentiable structure that induces this same product topology in the following way: Let $\mathcal{A}_{M_1}$ and $\mathcal{A}_{M_2}$ be atlases for M_1 and M_2. Take as charts on $M_1 \times M_2$ the maps of the form

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

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.

1.4. Sard's Theorem

Proof of Sard's theorem

In what follows we will need a technical result which we state without proof (See [**Stern**] or [**Bro-Jan**]). It is a version of the Fubini theorem.

Lemma 1.19 (Fubini). *Let $\mathbb{R}_t^{n-1}$ be defined to be the set $\{x \in \mathbb{R}^n : x^n = t\}$. Let C be a given set in $\mathbb{R}^n$. If $C_t := C \cap \mathbb{R}_t^{n-1}$ has measure zero in $\mathbb{R}_t^{n-1}$ for all $t \in \mathbb{R}$ then C has measure zero in $\mathbb{R}^n$.*

We now state and prove theorem of Sard.

Theorem 1.20 (Sard). *Let N be an n-manifold and M an m-manifold, both assumed second countable. For a smooth map $f : N \to M$, the set of critical values has measure zero.*

Proof. The proof is admittedly rather technical and perhaps not so enlightening. Some readers may reasonably decide to skip the proof on a first reading.

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

Let us use the following common notation: For any k-tuple of nonnegative integers $\alpha = (i_1, ..., i_k)$, we let

$$\frac{\partial^{|\alpha|} f}{\partial x^{\alpha}} := \frac{\partial^{i_1+\ldots+i_k} f}{\partial x^{i_1} \cdots \partial x^{i_k}}$$

where $|\alpha| := i_1 + ... + i_k$. Now let

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

Then

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

so we will be done if we can show that

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

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

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

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

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

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

$$f|_V : x \mapsto (x^1, g^2(x), ..., g^m(x))$$

for some perhaps smaller neighborhood V containing a. For each t, let f^t denote the restriction of f to the hyperplane $\left(\{t\} \times \mathbb{R}^{m-1}\right) \cap V$. Jacobian matrix for f at a point $(t, x^2, ..., x^n)$ in $\left(\{t\} \times \mathbb{R}^{n-1}\right) \cap V$ is of the form

$$Df = \begin{bmatrix} 1 & 0 \\ * & Df^t \end{bmatrix}$$

and so $(t, x^2, ..., x^n)$ is critical for f if and only if it is critical for f^t. Notice that

$$f(\left(\{t\} \times \mathbb{R}^{n-1}\right) \cap V) = f^t(\left(\{t\} \times \mathbb{R}^{n-1}\right) \cap V) \subset \{t\} \times \mathbb{R}^{m-1}.$$

But the domain of f^t is essentially an open set in a copy of $\mathbb{R}^{n-1}$ and so by the induction hypothesis $f(\left(\{t\} \times \mathbb{R}^{n-1}\right) \cap V)$ has measure zero in $\{t\} \times \mathbb{R}^{m-1}$. Thus the set of critical values of $f|_V$ has an intersection of measure zero with every set of the form $\{t\} \times \mathbb{R}^{m-1}$. By Fubini's theorem 1.19, $f((C \backslash C_1) \cap V)$ has measure zero. Since we may cover C by a countable number of sets of the form $(C \backslash C_1) \cap V$, we conclude that $f(C \backslash C_1)$ itself has measure zero.

Proof of (b): The proof of this part is quite similar to the proof of (a). Let $a \in C_{j-1} \backslash C_j$. It follows that some j-th partial derivative is not zero at a and after some permutation of the coordinate functions we may assume that

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

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

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

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

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

for some map $h : V \to \mathbb{R}^{m-1}$ where $V' = \phi(V)$. By the induction hypothesis, the restriction of g to

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

has a set of critical values of measure zero. But each point from $\phi(C_{j-1} \cap V) \subset 0 \times \mathbb{R}^{n-1}$ is critical for g_0 since diffeomorphisms preserve criticality and all partial derivatives of g and hence g_0, of order less than j vanish. Thus $g \circ \phi(C_{j-1} \cap V) = f(C_{j-1} \cap V)$ has measure zero. A countable number of set of the form $C_{j-1} \cap V$ covers $C_{j-1} \backslash C_j$ so $f(C_{j-1} \backslash C_j)$ has measure zero.

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

$$|f(a+h) - f(a)| \leq c\,|h|^{k+1} \tag{1.1}$$

for some constant c that depends only on f and $I^n(r)$. We now decompose the cube $I^n(r)$ into R^n cubes of side length r/R. Suppose that we label these cubes which contain critical points of f as $D_1, ...D_N$. Let D_i contain a critical point a of f. Now if $y \in D_i$, then $|y - a| \leq \sqrt{n}r/R$. Thus, using the Taylor's theorem remainder estimate above (1.1) with $y = a + h$, we see that $f(D_i)$ is contained in a cube $\widetilde{D}_i \subset \mathbb{R}^m$ of side

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

where the constant $b := 2c(\sqrt{n}r)^{k+1}$ is independent of the particular cube D from the decomposition. Furthermore depends only on f and $I^n(r)$. The sum of the volumes of all such cubes $\widetilde{D}_i$ is

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

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

Chapter 2 Supplement

2.1. Comments and Errata

2.1.1. Comments. (Nothing yet)

2.1.2. Errata. (Nothing yet)

2.2. Time Dependent Vector Fields

Definition 2.1. 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 2.2. 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 2.3. 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 2.4. *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 2.5. 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 2.6. 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, $X_t f$ must mean $X_t f_t \in C^\infty(M)$ 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 2.7. *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:*

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

and

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

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

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

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

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

Claim:

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

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 $\Phi_{s,t}^{X} = \left(\Phi_{t,s}^{X}\right)^{-1}$

$$
(***) \qquad \left(\Phi_{t,s}^{*}Y\right)f = \Phi_{t,s}^{*}\left(Y\left(\Phi_{s,t}^{*}f\right)\right)
$$

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

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

$\square$

Chapter 5 Supplement

5.1. Comments and Errata

5.1.1. Comments. (Nothing yet)

5.1.2. Errata. (Nothing yet)

5.2. Matrix Commutator Bracket

Recall that if V is a finite dimensional vector space, then each tangent space $T_x\mathrm{V}$ is naturally isomorphic to V. Now $\mathrm{GL}(n)$ is an open subset of the vector space of $n \times n$ real matrices $M_{n\times n}$ and so we obtain natural vector space isomorphisms $T_g\mathrm{GL}(n) \cong M_{n\times n}$ for all $g \in \mathrm{GL}(n)$. To move to the level of bundles, we reconstitute these isomorphisms to get maps $T_g\mathrm{GL}(n) \to \{g\} \times M_{n\times n}$ which we can then bundle together to get a trivialization $T\mathrm{GL}(n) \to \mathrm{GL}(n) \times M_{n\times n}$ (Recall Definition**??**). One could use this trivialization to identify $T\mathrm{GL}(n)$ with $\mathrm{GL}(n) \times M_{n\times n}$ and this trivialization is just a special case of the general situation: When U is an open subset of a vector space V, we have a trivialization $TU \cong U \times \mathrm{V}$. Further on, we introduce two more trivializations of $T\mathrm{GL}(n)$ defined using the (left or right) Maurer-Cartan form defined below. This will work for general Lie groups. One must be careful to keep track of which identification is in play.

Let us explicitly recall one way to describe the isomorphism $T_g\mathrm{GL}(n) \cong M_{n\times n}$. If $v_g \in T_gG$, then there is some curve (of matrices) $c^g : t \mapsto c(t)$ such that $c^g(0) = g$ and $\dot{c}^g(0) = v_g \in T_gG$. By definition $\dot{c}^g(0) := T_0c^g \cdot \frac{d}{dt}\big|_0$ which is based at g. If we just take the ordinary derivative we get the matrix that represents v_g: If $c(t)$ is given by

$$c(t) = \begin{bmatrix} g_1^1(t) & g_1^2(t) & \cdots \\ g_2^1(t) & \ddots & \\ \vdots & & g_n^n(t) \end{bmatrix}$$

then $\dot{c}(0) = v_g$ is identified with the matrix

$$a := \begin{bmatrix} \frac{d}{dt}\big|_{t=0} g_1^1 & \frac{d}{dt}\big|_{t=0} g_1^2 & \cdots \\ \frac{d}{dt}\big|_{t=0} g_2^1 & \ddots & \\ \vdots & & \frac{d}{dt}\big|_{t=0} g_n^n \end{bmatrix}.$$

As a particular case, we have a vector space isomorphism $\mathfrak{gl}(n) = T_I\mathrm{GL}(n) \cong M_{n\times n}$ where I is the identity matrix in $\mathrm{GL}(n)$. We want to use this to identify $\mathfrak{gl}(n)$ with $M_{n\times n}$. Now $\mathfrak{gl}(n) = T_I\mathrm{GL}(n)$ has a Lie algebra structure, and we would like to transfer the Lie bracket from $\mathfrak{gl}(n)$ to $M_{n\times n}$ in such a way that this isomorphism becomes a Lie algebra isomorphism. Below we discover that the Lie bracket for $M_{n\times n}$ is the commutator bracket defined by $[A, B] := AB - BA$. This is so natural that we can safely identify the Lie algebra $\mathfrak{gl}(n)$ with $M_{n\times n}$. Along these lines we will also be able to identify the Lie algebras of subgroups of $\mathrm{GL}(n)$ (or $\mathrm{GL}(n, \mathbb{C})$) with linear subspaces of $M_{n\times n}$ (or $M_{n\times n}(\mathbb{C})$).

Initially, the Lie algebra of $\mathrm{GL}(n)$ is given as the space of left invariant vectors fields on $GL(n)$. The bracket is the bracket of vector fields that we met earlier. This bracket is transferred to $T_I\mathrm{GL}(n)$ according to the isomorphism of $\mathfrak{X}(GL(n))$ with $T_IGL(n)$ given in this case by $X \mapsto X(I)$. The situation is that we have two Lie algebra isomorphisms

$$\mathfrak{X}(\mathrm{GL}(n)) \cong T_I\mathrm{GL}(n) \cong M_{n\times n},$$

and the origin of all of the Lie algebra structure is $\mathfrak{X}(\mathrm{GL}(n))$. The plan is then to find the left invariant vector field that corresponds to a given matrix from $M_{n\times n}$. This gives a direct route between $\mathfrak{X}(\mathrm{GL}(n))$ and $M_{n\times n}$ allowing us to see what the correct bracket on $M_{n\times n}$ should be. Note that a global coordinate system for $\mathrm{GL}(n)$ is given by the maps x_l^k, which are defined by the requirement that $x_l^k(A) = a_l^k$ whenever $A = (a_j^i)$. In what follows, it will be convenient to use the Einstein summation convention. Thus any vector fields $X, Y \in \mathfrak{X}(\mathrm{GL}(n))$ can be written

$$X = f_j^i \frac{\partial}{\partial x_j^i}$$

$$Y = g_j^i \frac{\partial}{\partial x_j^i}$$

for some functions f_j^i and g_j^i. Now let $(a_j^i) = A \in M_{n\times n}$. The corresponding element of $T_I\mathrm{GL}(n)$ can be written $A_I = a_j^i \frac{\partial}{\partial x_j^i}\Big|_I$. Corresponding to A_I is a left invariant vector field $X^A := L(A_I)$, which is given by $X^A(x) = T_I L_x \cdot A_I$ and which, in turn, corresponds to the matrix $\frac{d}{dt}\big|_0 xc(t) = xA$ where $c'(0) = A$. Thus X^A is given by $X^A = f_j^i \frac{\partial}{\partial x_j^i}$ where $f_j^i(x) = xA = \left(x_k^i a_j^k\right)$. Similarly, let B be another matrix with corresponding left invariant vector field X^B. The bracket $[X^A, X^B]$ can be computed as follows:

$$\begin{aligned}
[X^A, X^B] &= (f_j^i \frac{\partial g_l^k}{\partial x_j^i} - g_j^i \frac{\partial f_l^k}{\partial x_j^i})\frac{\partial}{\partial x_l^k} \\
&= \left(x_r^i a_j^r \frac{\partial(x_s^k b_l^s)}{\partial x_j^i} - x_r^i b_j^r \frac{\partial(x_s^k a_l^s)}{\partial x_j^i}\right)\frac{\partial}{\partial x_l^k} \\
&= \left(x_r^k a_s^r b_l^s - x_r^k b_s^r a_l^s\right)\frac{\partial}{\partial x_l^k}.
\end{aligned}$$

Evaluating at $I = (\delta_i^i)$ we have

$$\begin{aligned}
[X^A, X^B](I) &= \left(\delta_r^k a_s^r b_l^s - \delta_r^k b_s^r a_l^s\right)\frac{\partial}{\partial x_l^k}\Big|_I \\
&= \left(a_s^k b_l^s - b_s^k a_l^s\right)\frac{\partial}{\partial x_l^k}\Big|_I,
\end{aligned}$$

which corresponds to the matrix $AB - BA$. Thus the proper Lie algebra structure on $M_{n\times n}$ is given by the commutator $[A, B] = AB - BA$. In summary, we have

Proposition 5.1. Under the canonical of identification of $\mathfrak{gl}(n) = T_I GL(n)$ with $M_{n\times n}$ the Lie bracket is the commutator bracket

$$[A, B] = AB - BA.$$

Similarly, under the identification of $T_{\mathrm{id}}\mathrm{GL(V)}$ with $\mathrm{End(V)}$ the bracket is

$$[A, B] = A \circ B - B \circ A\,.$$

5.3. Spinors and rotation

The matrix Lie group SO(3) is the group of orientation preserving rotations of $\mathbb{R}^3$ acting by matrix multiplication on column vectors. The group SU(2) is the group of complex 2×2 unitary matrices of determinant 1. We shall now expose an interesting relation between these groups. First recall the Pauli matrices:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The real vector space spanned by $\sigma_1, \sigma_2, \sigma_3$ is isomorphic to $\mathbb{R}^3$ and is the space of traceless Hermitian matrices. Let us temporarily denote the latter by $\widehat{\mathbb{R}^3}$. Thus we have a linear isomorphism $\mathbb{R}^3 \to \widehat{\mathbb{R}^3}$ given by $(x^1, x^2, x^3) \mapsto x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3$ which we abbreviate to $\vec{x} \mapsto \widehat{x}$. Now it is easy to check that $\det(\widehat{x})$ is just $-|\vec{x}|^2$. In fact, we may introduce an inner product on $\widehat{\mathbb{R}^3}$ by the formula $\langle \widehat{x}, \widehat{y} \rangle := \frac{1}{2} tr(\widehat{x}\widehat{y})$ and then we have that $\vec{x} \mapsto \widehat{x}$ is an isometry. Next we notice that $SU(2)$ acts on $\widehat{\mathbb{R}^3}$ by $(g, \widehat{x}) \mapsto g\widehat{x}g^{-1} = g\widehat{x}g^*$ thus giving a representation ρ of $\mathrm{SU}(2)$ in $\widehat{\mathbb{R}^3}$. It is easy to see that $\langle \rho(g)\widehat{x}, \rho(g)\widehat{y} \rangle = \langle \widehat{x}, \widehat{y} \rangle$ and so under the identification $\mathbb{R}^3 \leftrightarrow \widehat{\mathbb{R}^3}$ we see that $\mathrm{SU}(2)$ act on $\mathbb{R}^3$ as an element of $\mathrm{O}(3)$.

Exercise 5.2. Show that in fact, the map $\mathrm{SU}(2) \to \mathrm{O}(3)$ is actually a group map onto $\mathrm{SO}(3)$ with kernel $\{\pm I\} \cong \mathbb{Z}_2$.

Exercise 5.3. Show that the algebra generated by the matrices $\sigma_0, -i\sigma_1, -i\sigma_2, -i\sigma_3$ is isomorphic to the quaternion algebra and that the set of matrices $-i\sigma_1, -i\sigma_2, -i\sigma_3$ span a real vector space which is equal as a set to the traceless skew Hermitian matrices $\mathfrak{su}(2)$.

Let $\mathbf{I} = -i\sigma_1$, $\mathbf{J} = -i\sigma_2$ and $\mathbf{K} = -i\sigma_3$. One can redo the above analysis using the isometry $\mathbb{R}^3 \to \mathfrak{su}(2)$ given by

$$\begin{aligned} (x^1, x^2, x^3) &\mapsto x^1\mathbf{I} + x^2\mathbf{J} + x^3\mathbf{K} \\ \vec{x} &\mapsto \widetilde{x} \end{aligned}$$

where this time $\langle \widetilde{x}, \widetilde{y} \rangle := \frac{1}{2}\mathrm{tr}(\widetilde{x}\widetilde{y}^*) = -\frac{1}{2}\mathrm{tr}(\widetilde{x}\widetilde{y})$. By the above exercise, that the real linear space $\{t\sigma_0 + x\mathbf{I} + y\mathbf{J} + z\mathbf{K} : t, x, y, z \in \mathbb{R}\}$ is a matrix realization of the quanternion algebra. Now let $A = t\sigma_0 + x\mathbf{I} + y\mathbf{J} + z\mathbf{K}$. It is easy to check that $\det(A) = t^2 + x^2 + y^2 + z^2$ and that

$$\begin{aligned} A^*A &= \begin{pmatrix} t + iz & ix + y \\ ix - y & t - iz \end{pmatrix} \begin{pmatrix} t - iz & -ix - y \\ y - ix & t + iz \end{pmatrix} \\ &= \begin{pmatrix} t^2 + x^2 + y^2 + z^2 & 0 \\ 0 & t^2 + x^2 + y^2 + z^2 \end{pmatrix}. \end{aligned}$$

It follows that $\left\{t\sigma_0 + x\mathbf{I} + y\mathbf{J} + z\mathbf{K} : t^2 + x^2 + y^2 + z^2 = 1\right\}$ is $\mathrm{SU}(2)$, exactly the set of unit quaternions, and a copy of S^3. Notice that $\mathfrak{su}(2) = \mathrm{span}_{\mathbb{R}}\{\mathbf{I}, \mathbf{J}, \mathbf{K}\}$ is the set of traceless skew-Hermitian matrices which we take as the Lie algebra of $\mathrm{SU}(2)$. The the action of $\mathrm{SU}(2)$ on $\mathfrak{su}(2)$ given by

$(g, \widehat{x}) \mapsto g\widehat{x}g^{-1} = g\widehat{x}g^*$ is just the adjoint action; $\mathrm{Ad}(g) : \widehat{x} \mapsto g\widehat{x}g^*$. But

$$\begin{aligned}\langle \mathrm{Ad}(g)\widehat{x}, \mathrm{Ad}(g)\widehat{y}\rangle &= -\frac{1}{2}\mathrm{tr}\,(\mathrm{Ad}(g)\widehat{x}\mathrm{Ad}(g)\widehat{y}) \\ &= -\frac{1}{2}\mathrm{tr}\left(g\widehat{x}g^{-1}g\widehat{y}g^{-1}\right) = -\frac{1}{2}\mathrm{tr}\,(\widehat{x}\widehat{y}) \\ &= \langle \widetilde{x}, \widetilde{y}\rangle\end{aligned}$$

and so $\mathrm{Ad}(g) \in \mathrm{SO}(\mathfrak{su}(2), \langle,\rangle)$ for all g. Thus, identifying $\mathrm{SO}(\mathfrak{su}(2), \langle,\rangle)$ with $\mathrm{SO}(3)$ we obtain the same map $\mathrm{SU}(2) \to \mathrm{SO}(3)$ which is a Lie group homomorphism and has kernel $\{\pm I\} \cong \mathbb{Z}_2$.

Exercise 5.4. Check the details here!

What is the differential of the map $\rho : \mathrm{SU}(2) \to \mathrm{SO}(3)$ at the identity? Let $g(t)$ be a curve in $\mathrm{SU}(2)$ with $\frac{d}{dt}\big|_{t=0} g = g'$ and $g(0) = \mathrm{id}$. We have $\frac{d}{dt}(g(t)Ag^*(t)) = (\frac{d}{dt}g(t))Ag^*(t) + g(t)A(\frac{d}{dt}g(t))^*$ and so

$$\mathrm{ad} : g' \mapsto g'A + Ag'^* = [g', A].$$

Then $\langle g(t)\widehat{x}, g(t)\widehat{y}\rangle = 1$ for all t and we have

$$\begin{aligned}0 = \left.\frac{d}{dt}\right|_{t=0} \langle g\widehat{x}, g\widehat{y}\rangle &= \left.\frac{d}{dt}\right|_{t=0} \frac{1}{2}\mathrm{tr}(g\widetilde{x}(g\widetilde{y})^*) \\ &= \frac{1}{2}\mathrm{tr}([g', \widetilde{x}]\widetilde{y}^*) + \frac{1}{2}\mathrm{tr}(\widetilde{x}([g', \widetilde{y}])^*) \\ &= \frac{1}{2}\mathrm{tr}([g', \widetilde{x}]\widetilde{y}^*) - \frac{1}{2}\mathrm{tr}(\widetilde{x}[g', \widetilde{y}]) \\ &= \langle [g', \widetilde{x}], \widetilde{y}\rangle - \langle \widetilde{x}, [g', \widetilde{y}]\rangle \\ &= \langle \mathrm{ad}(g')\widetilde{x}, \widetilde{y}\rangle - \langle \widetilde{x}, \mathrm{ad}(g')\widetilde{y}\rangle.\end{aligned}$$

From this it follows that the differential of the map $\mathrm{SU}(2) \to O(3)$ takes $\mathfrak{su}(2)$ isomorphically onto the space $\mathfrak{so}(3)$. We have

$$\begin{array}{ccc} \mathfrak{su}(2) & = & \mathfrak{su}(2) \\ d\rho \downarrow & & ad \downarrow \\ \mathfrak{so}(3) & \cong & \mathfrak{so}(\mathfrak{su}(2), \langle,\rangle) \end{array}$$

where $\mathfrak{so}(\mathfrak{su}(2), \langle,\rangle)$ denotes the linear maps $\mathfrak{su}(2) \to \mathfrak{su}(2)$ skew-symmetric with respect to the inner product $\langle \widetilde{x}, \widetilde{y}\rangle := \frac{1}{2}\mathrm{tr}(\widetilde{x}\widetilde{y}^*)$.

5.4. Lie Algebras

Let $\mathbb{F}$ denote one of the fields $\mathbb{R}$ or $\mathbb{C}$. In definition **??** we defined a real Lie algebra $\mathfrak{g}$ as a real algebra with a skew symmetric (bilinear) product called the Lie bracket, usually denoted $(v,w) \mapsto [v,w]$, such that the Jacobi identity holds

$$\text{(Jacobi Identity)} \qquad [x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0 \text{ for all } x,y,z \in \mathfrak{g}.$$

We also have the notion of a *complex* Lie algebra defined analogously.

Remark 5.5. We will assume that all the Lie algebras we study are finite dimensional unless otherwise indicated.

Let V be a finite dimensional vector space over $\mathbb{F}$ and recall that $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ is the set of all $\mathbb{F}-$linear maps $\mathrm{V} \to \mathrm{V}$. The space $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ is also denoted $\mathrm{End}_{\mathbb{F}}(\mathrm{V},\mathrm{V})$ or $L_{\mathbb{F}}(\mathrm{V},\mathrm{V})$ although in the context of Lie algebras we take $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ as the preferred notation. We give $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ its natural Lie algebra structure where the bracket is just the commutator bracket

$$[A,B] := A \circ B - B \circ A.$$

If the field involved is either irrelevant or known from context we will just write $\mathfrak{gl}(\mathrm{V})$. Also, we often identify $\mathfrak{gl}(\mathbb{F}^n)$ with the matrix Lie algebra $\mathbb{M}_{nxn}(\mathbb{F})$ with the bracket $AB - BA$.

For a Lie algebra $\mathfrak{g}$ we can associate to every basis $v_1, ..., v_n$ for $\mathfrak{g}$ the **structure constants** c_{ij}^k which are defined by

$$[v_i, v_j] = \sum_k c_{ij}^k v_k.$$

It then follows from the skew symmetry of the Lie bracket and the Jacobi identity that the structure constants satisfy

$$\text{(5.1)} \qquad \begin{array}{ll} \text{i)} & c_{ij}^k = -c_{ji}^k \\ \text{ii)} & \sum_k c_{rs}^k c_{kt}^i + c_{st}^k c_{kr}^i + c_{tr}^k c_{ks}^i = 0 \end{array}$$

Given a real Lie algebra $\mathfrak{g}$ we can extend it to a complex Lie algebra $\mathfrak{g}_{\mathbb{C}}$ by defining as $\mathfrak{g}_{\mathbb{C}}$ the complexification $\mathfrak{g}_{\mathbb{C}} := \mathbb{C} \otimes_{\mathbb{R}} \mathfrak{g}$ and then extending the bracket by requiring

$$[1 \otimes v, 1 \otimes w] = [v,w] \otimes 1.$$

Then $\mathfrak{g}$ can be identified with its image under the embedding map $v \mapsto 1 \otimes v$. In practice one often omits the symbol $\otimes$ and then the complexification just amounts to formally allowing complex coefficients.

Notation 5.6. Given two subsets S_1 and S_2 of a Lie algebra $\mathfrak{g}$ we let $[S_1, S_2]$ denote the linear span of the set defined by $\{[x,y] : x \in S_1$ and $y \in S_2\}$. Also, let $S_1 + S_2$ denote the vector space of all $x + y : x \in S_1$ and $y \in S_2$.

It is easy to verify that the following relations hold:

(1) $[S_1 + S_2, S] \subset [S_1, S] + [S_2, S]$
(2) $[S_1, S_2] = [S_2, S_1]$
(3) $[S, [S_1, S_2]] \subset [[S, S_1], S_2] + [S_1, [S, S_2]]$

where S_1, S_2 and S are subsets of a Lie algebra $\mathfrak{g}$.

Definition 5.7. A vector subspace $\mathfrak{a} \subset \mathfrak{g}$ is called a **subalgebra** if $[\mathfrak{a}, \mathfrak{a}] \subset \mathfrak{a}$ and an **ideal** if $[\mathfrak{g}, \mathfrak{a}] \subset \mathfrak{a}$.

If $\mathfrak{a}$ is a subalgebra of $\mathfrak{g}$ and $v_1, ..., v_k, v_{k+1}, ..., v_n$ is a basis for $\mathfrak{g}$ such that $v_1, ..., v_k$ is a basis for $\mathfrak{a}$ then with respect to this basis the structure constants are such that

$$c_{ij}^s = 0 \text{ if both } i, j \leq k \text{ and } s > k.$$

If $\mathfrak{a}$ is also an ideal then for any j we must have

$$c_{ij}^s = 0 \text{ when both } i \leq k \text{ and } s > k.$$

Remark 5.8. The numbers c_{ij}^s may be viewed as the components of an element of $T_{1,1}^1(\mathfrak{g})$ (i.e. as an algebraic tensor).

Example 5.9. Let $\mathfrak{su}(2)$ denote the set of all traceless and skew-Hermitian 2×2 complex matrices. This is a *real* Lie algebra under the commutator bracket $(AB - BA)$. A commonly used basis for $\mathfrak{su}(2)$ is e_1, e_2, e_3 where

$$e_1 = \frac{1}{2}\begin{bmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad e_2 = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad e_2 = \frac{1}{2}\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

The commutation relations satisfied by these matrices are

$$[e_i, e_j] = \epsilon_{ijk} e_k \text{ (no sum)}$$

where ϵ_{ijk} is the totally antisymmetric symbol given by

$$\epsilon_{ijk} := \begin{cases} 0 & \text{if } (i,j,k) \text{ is not a permutation of } (1,2,3) \\ 1 & \text{if } (i,j,k) \text{ is an even permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ is an odd permutation of } (1,2,3) \end{cases}.$$

Thus in this case the structure constants are $c_{ij}^k = \epsilon_{ijk}$. In physics it is common to use the Pauli matrices defined by $\sigma_i := 2\mathrm{i}e_i$ in terms of which the commutation relations become $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$ but now the Lie algebra is the isomorphic space of 2×2 traceless *Hermitian* matrices.

Example 5.10. The Weyl basis for $\mathfrak{gl}(n, \mathbb{R})$ is given by the n^2 matrices e_{sr} defined by

$$(e_{rs})_{ij} := \delta_{ri}\delta_{sj}.$$

Notice that we are now in a situation where "double indices" are convenient. For instance, the commutation relations read

$$[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta_{il} e_{kj}$$

while the structure constants are

$$c^{ij}_{sm,kr} = \delta^i_s \delta_{mk} \delta^j_r - \delta^i_k \delta_{rs} \delta^j_m.$$

5.4.0.1. *Classical complex Lie algebras.* If $\mathfrak{g}$ is a real Lie algebra we have seen that the complexification $\mathfrak{g}_{\mathbb{C}}$ is naturally a complex Lie algebra. As mentioned above, it is convenient to omit the tensor symbol and use the following convention: Every element of $\mathfrak{g}_{\mathbb{C}}$ may be written at $v + \mathrm{i}w$ for $v, w \in \mathfrak{g}$ and then

$$\begin{aligned}&[v_1 + \mathrm{i}w_1, v_2 + \mathrm{i}w_2]\\ &= [v_1, v_2] - [w_1, w_2] + \mathrm{i}([v_1, w_2] + [w_1, v_2]).\end{aligned}$$

We shall now define a series of complex Lie algebras sometimes denoted by A_n, B_n, C_n and D_n for every integer $n > 0$. First of all, notice that the complexification $\mathfrak{gl}(n, \mathbb{R})_{\mathbb{C}}$ of $\mathfrak{gl}(n, \mathbb{R})$ is really just $\mathfrak{gl}(n, \mathbb{C})$; the set of complex $n \times n$ matrices under the commutator bracket.

The algebra A_n: The set of all traceless $n \times n$ matrices is denoted A_{n-1} and also by $\mathfrak{sl}(n, \mathbb{C})$.

We call the readers attention to the following general fact: If $b(.,.)$ is a bilinear form on a complex vector space V then the set of all $A \in \mathfrak{gl}(n, \mathbb{C})$ such that $b(Az, w) + b(z, Aw) = 0$ for every $z, w \in V$ is a Lie subalgebra of $\mathfrak{gl}(n, \mathbb{C})$. This follows from the calculation

$$\begin{aligned} b([A,B]z, w) &= b(ABz, w) - b(BAz, w)\\ &= -b(Bz, Aw) + b(Az, Bw)\\ &= b(z, BAw) - b(z, ABw)\\ &= b(z, [B, A]w).\end{aligned}$$

The algebras B_n and D_n: Let $m = 2n + 1$ and let $b(.,.)$ be a non-degenerate symmetric bilinear form on an m dimensional complex vector space V. Without loss we may assume $V = \mathbb{C}^m$ and we may take $b(z, w) = \sum_{i=1}^m z_i w_i$. We define B_n to be $\mathfrak{o}(m, \mathbb{C})$ where

$$\mathfrak{o}(m, \mathbb{C}) := \{A \in \mathfrak{gl}(m, \mathbb{C}) : b(Az, w) + b(z, Aw) = 0\}.$$

Similarly, for $m = 2n$ we define D_n to be $\mathfrak{o}(m, \mathbb{C})$.

The algebra C_n: The algebra associated to a skew-symmetric non-degenerate bilinear form which we may take to be $b(z, w) = \sum_{i=1}^n z_i w_{n+i} - \sum_{i=1}^n z_{n+i} w_i$ on $\mathbb{C}^{2n}$. We obtain the complex symplectic algebra

$$C_n = \mathfrak{sp}(n, \mathbb{C}) := \{A \in \mathfrak{gl}(m, \mathbb{C}) : b(Az, w) + b(z, Aw) = 0\}.$$

5.4.0.2. *Basic Facts and Definitions.* The expected theorems hold for homomorphisms; the image $\mathrm{img}(\sigma) := \sigma(\mathfrak{a})$ of a homomorphism $\sigma : \mathfrak{a} \to \mathfrak{b}$ is a subalgebra of $\mathfrak{b}$ and the kernel $\ker(\sigma)$ is an ideal of $\mathfrak{a}$.

Definition 5.11. Let $\mathfrak{h}$ be an ideal in $\mathfrak{g}$. On the quotient vector space $\mathfrak{g}/\mathfrak{h}$ with quotient map π we can define a Lie bracket in the following way: For $\bar{v}, \bar{w} \in \mathfrak{g}/\mathfrak{h}$ choose $v, w \in \mathfrak{g}$ with $\pi(v) = \bar{v}$ and $\pi(w) = \bar{w}$ we define

$$[\bar{v}, \bar{w}] := \pi([v, w]).$$

We call $\mathfrak{g}/\mathfrak{h}$ with this bracket the **quotient Lie algebra**.

Exercise 5.12. Show that the bracket defined in the last definition is well defined.

Given two linear subspaces $\mathfrak{a}$ and $\mathfrak{b}$ of $\mathfrak{g}$ the (not necessarily direct) sum $\mathfrak{a} + \mathfrak{b}$ is just the space of all elements in $\mathfrak{g}$ of the form $a + b$ where $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$. It is not hard to see that if $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in $\mathfrak{g}$ then so is $\mathfrak{a} + \mathfrak{b}$.

Exercise 5.13. Show that for $\mathfrak{a}$ and $\mathfrak{b}$ ideals in $\mathfrak{g}$ we have a natural isomorphism $(\mathfrak{a} + \mathfrak{b})/\mathfrak{b} \cong \mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$.

If $\mathfrak{g}$ is a Lie algebra, $\mathfrak{s}$ a subset of $\mathfrak{g}$, then the **centralizer of $\mathfrak{s}$ in $\mathfrak{g}$** is $\mathfrak{z}(\mathfrak{s}) := \{x \in \mathfrak{g} : [x, y] = 0 \text{ for all } y \in \mathfrak{s}\}$. If $\mathfrak{a}$ is a (Lie) subalgebra of $\mathfrak{g}$ then the **normalizer** of $\mathfrak{a}$ in $\mathfrak{g}$ is $\mathfrak{n}(\mathfrak{a}) := \{v \in \mathfrak{g} : [v, \mathfrak{a}] \subset \mathfrak{a}\}$. One can check that $\mathfrak{n}(\mathfrak{a})$ is an ideal in $\mathfrak{g}$.

There is also a Lie algebra product. Namely, if $\mathfrak{a}$ and $\mathfrak{b}$ are Lie algebras, then we can define a Lie bracket on $\mathfrak{a} \times \mathfrak{b}$ by

$$[(a_1, a_2), (b_1, b_2)] := ([a_1, b_1], [a_2, b_2]).$$

With this bracket, $\mathfrak{a} \times \mathfrak{b}$ is a Lie algebra called the Lie algebra **product** of $\mathfrak{a}$ and $\mathfrak{b}$. The subspaces $\mathfrak{a} \times \{0\}$ and $\{0\} \times \mathfrak{b}$ are ideals in $\mathfrak{a} \times \mathfrak{b}$ that are clearly isomorphic to $\mathfrak{a}$ and $\mathfrak{b}$ respectively. Depending on the context this is also written as $\mathfrak{a} \oplus \mathfrak{b}$ and then referred to as the **direct sum** (external direct sum). If $\mathfrak{a}$ and $\mathfrak{b}$ are subalgebras of a Lie algebra $\mathfrak{g}$ such that $\mathfrak{a} + \mathfrak{b} = \mathfrak{g}$ and $\mathfrak{a} \cap \mathfrak{b} = 0$ then we have the vector space direct sum which, for reasons which will be apparent shortly, we denote by $\mathfrak{a} \dot{+} \mathfrak{b}$. If we have several subalgebras of $\mathfrak{g}$, say $\mathfrak{a}_1, ..., \mathfrak{a}_k$ such that $\mathfrak{a}_i \cap \mathfrak{a}_j = 0$ for $i \neq j$, and if $\mathfrak{g} = \mathfrak{a}_1 \dot{+} \cdots \dot{+} \mathfrak{a}_k$ which is the vector space direct sum. For the Lie algebra direct sum we need to require that $[\mathfrak{a}_i, \mathfrak{a}_j] = 0$ for $i \neq j$. In this case we write $\mathfrak{g} = \mathfrak{a}_1 \oplus \cdots \oplus \mathfrak{a}_k$ which is the Lie algebra (internal) direct sum. In this case, it is easy to verify that each $\mathfrak{a}_i$ is an ideal in $\mathfrak{g}$. With respect to such a decomposition the Lie product becomes $[\sum a_i, \sum a'_j] = \sum_i [a_i, a'_i]$. Clearly, the internal direct sum $\mathfrak{a}_1 \oplus \cdots \oplus \mathfrak{a}_k$ is isomorphic to $\mathfrak{a}_1 \times \cdots \times \mathfrak{a}_k$ which as we have seen is also denoted as $\mathfrak{a}_1 \oplus \cdots \oplus \mathfrak{a}_k$ (the external direct sum this time).

Definition 5.14. The **center** $\mathfrak{z}(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$ is the subspace $\mathfrak{z}(\mathfrak{g}) := \{v \in \mathfrak{g} : [v, y] = 0 \text{ for all } y \in \mathfrak{g}\}$.

5.4.0.3. *The Adjoint Representation.* In the context of abstract Lie algebras, the **adjoint map** $a \to \mathrm{ad}(a)$ is given by $\mathrm{ad}(a)(b) := [a, b]$. It is easy to see that $\mathfrak{z}(\mathfrak{g}) = \ker(\mathrm{ad})$.

We have $[a^i v_i, b^j v_j] = a^i b^j [v_i, v_j] = a^i b^j c_{ij}^k v_k$ and so the matrix of $\mathrm{ad}(a)$ with respect to the basis $(v_1, ..., v_n)$ is $(ad(a))_j^k = a^i c_{ij}^k$. In particular, $(ad(v_i))_j^k = c_{ij}^k$.

Definition 5.15. A derivation of a Lie algebra $\mathfrak{g}$ is a linear map $D : \mathfrak{g} \to \mathfrak{g}$ such that

$$D[v, w] = [Dv, w] + [v, Dw]$$

for all $v, w \in \mathfrak{g}$.

For each $v \in \mathfrak{g}$ the map $\mathrm{ad}(v) : \mathfrak{g} \to \mathfrak{g}$ is actually a derivation of the Lie algebra $\mathfrak{g}$. Indeed, this is exactly the content of the Jacobi identity. Furthermore, it is not hard to check that the space of all derivations of a Lie algebra $\mathfrak{g}$ is a subalgebra of $\mathfrak{gl}(\mathfrak{g})$. In fact, if D_1 and D_2 are derivations of $\mathfrak{g}$ then so is the commutator $D_1 \circ D_2 - D_2 \circ D_1$. We denote this subalgebra of derivations by $Der(\mathfrak{g})$.

Definition 5.16. A **Lie algebra representation** ρ of $\mathfrak{g}$ on a vector space V is a Lie algebra homomorphism $\rho : \mathfrak{g} \to \mathfrak{gl}(\mathrm{V})$.

One can construct Lie algebra representations in various ways from given representations. For example, if $\rho_i : \mathfrak{g} \to \mathfrak{gl}(\mathrm{V}_i)$ $(i = 1, .., k)$ are Lie algebra representations then $\oplus_i \rho_i : \mathfrak{g} \to \mathfrak{gl}(\oplus_i \mathrm{V}_i)$ defined by

$$(\oplus_i \rho_i)(x)(v_1 \oplus \cdots \oplus v_n) = \rho_1(x) v_1 \oplus \cdots \oplus \rho_1(x) v_n \tag{5.2}$$

for $x \in \mathfrak{g}$ is a Lie algebra representation called the **direct sum representation** of the ρ_i. Also, if one defines

$$\begin{aligned}(\otimes_i \rho_i)(x)(v_1 \otimes \cdots \otimes v_k) &:= \rho_1(x) v_1 \otimes v_2 \otimes \cdots \otimes v_k \\ &+ v_1 \otimes \rho_2(x) v_2 \otimes \cdots \otimes v_k + \cdots + v_1 \otimes v_2 \otimes \cdots \otimes \rho_k(x) v_k\end{aligned}$$

(and extend linearly) then $\otimes_i \rho_i$ is a representation $\otimes_i \rho_i : \mathfrak{g} \to \mathfrak{gl}(\otimes_i \mathrm{V}_i)$ is Lie algebra representation called a **tensor product representation**.

Lemma 5.17. *$ad : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ is a Lie algebra representation on $\mathfrak{g}$. The image of* ad *is contained in the Lie algebra* $\mathrm{Der}(\mathfrak{g})$ *of all derivation of the Lie algebra $\mathfrak{g}$.*

Proof. This follows from the Jacobi identity (as indicated above) and from the definition of ad. □

Corollary 5.18. $\mathfrak{r}(\mathfrak{g})$ is an ideal in $\mathfrak{g}$.

The image $\mathrm{ad}(\mathfrak{g})$ of ad in $\mathrm{Der}(\mathfrak{g})$ is called the adjoint algebra.

Definition 5.19. The Killing form for a Lie algebra $\mathfrak{g}$ is the bilinear form given by
$$K(X,Y) = Tr(\mathrm{ad}(X) \circ \mathrm{ad}(Y))$$

Lemma 5.20. *For any Lie algebra automorphism $\vartheta : \mathfrak{g} \to \mathfrak{g}$ and any $X \in \mathfrak{g}$ we have* $\mathrm{ad}(\vartheta X) = \vartheta \mathrm{ad} X \vartheta^{-1}$

Proof. $\mathrm{ad}(\vartheta X)(Y) = [\vartheta X, Y] = [\vartheta X, \vartheta\vartheta^{-1}Y] = \vartheta[X, \vartheta^{-1}Y] = \vartheta \circ \mathrm{ad}X \circ \vartheta^{-1}(Y)$. □

Clearly $K(X,Y)$ is symmetric in X and Y but we have more identities:

Proposition 5.21. The Killing forms satisfies identities:

1) $K([X,Y],Z) = K([Z,X],Y)$ for all $X,Y,Z \in \mathfrak{g}$

2) $K(\rho X, \rho Y) = K(X,Y)$ for any Lie algebra automorphism $\rho : \mathfrak{g} \to \mathfrak{g}$ and any $X, Y \in \mathfrak{g}$.

Proof. For (1) we calculate
$$\begin{aligned}
K([X,Y],Z) &= Tr(\mathrm{ad}([X,Y]) \circ \mathrm{ad}(Z)) \\
&= Tr([\mathrm{ad}X, \mathrm{ad}Y] \circ \mathrm{ad}(Z)) \\
&= Tr(\mathrm{ad}X \circ \mathrm{ad}Y \circ \mathrm{ad}Z - \mathrm{ad}Y \circ \mathrm{ad}X \circ \mathrm{ad}Z) \\
&= Tr(\mathrm{ad}Z \circ \mathrm{ad}X \circ \mathrm{ad}Y - \mathrm{ad}X \circ \mathrm{ad}Z \circ \mathrm{ad}Y) \\
&= Tr([\mathrm{ad}Z, \mathrm{ad}X] \circ \mathrm{ad}Y) \\
&= Tr(\mathrm{ad}[Z,X] \circ \mathrm{ad}Y) = K([Z,X],Y)
\end{aligned}$$
where we have used that $Tr(ABC)$ is invariant under cyclic permutations of A, B, C.

For (2) just observe that
$$\begin{aligned}
K(\rho X, \rho Y) &= Tr(\mathrm{ad}(\rho X) \circ \mathrm{ad}(\rho Y)) \\
&= Tr(\rho \mathrm{ad}(X)\rho^{-1}\rho \mathrm{ad}(Y)\rho^{-1}) \quad \text{(lemma 5.20)} \\
&= Tr(\rho \mathrm{ad}(X) \circ \mathrm{ad}(Y)\rho^{-1}) \\
&= Tr(\mathrm{ad}(X) \circ \mathrm{ad}(Y)) = K(X,Y).
\end{aligned}$$
□

Since $K(X,Y)$ is symmetric in X, Y and so there must be a basis $\{X_i\}_{1 \le i \le n}$ of $\mathfrak{g}$ for which the matrix (k_{ij}) given by
$$k_{ij} := K(X_i, X_j)$$

is diagonal.

Lemma 5.22. *If $\mathfrak{a}$ is an ideal in $\mathfrak{g}$ then the Killing form of $\mathfrak{a}$ is just the Killing form of $\mathfrak{g}$ restricted to $\mathfrak{a} \times \mathfrak{a}$.*

Proof. Let $\{X_i\}_{1 \leq i \leq n}$ be a basis of $\mathfrak{g}$ such that $\{X_i\}_{1 \leq i \leq r}$ is a basis for $\mathfrak{a}$. Now since $[\mathfrak{a}, \mathfrak{g}] \subset \mathfrak{a}$, the structure constants c^i_{jk} with respect to this basis must have the property that $c^k_{ij} = 0$ for $i \leq r < k$ and all j. Thus for $1 \leq i, j \leq r$ we have

$$\begin{aligned} K_{\mathfrak{a}}(X_i, X_j) &= Tr(\mathrm{ad}(X_i)\mathrm{ad}(X_j)) \\ &= \sum_{k=1}^{r} \sum_{s=1}^{r} c^k_{is} c^s_{jk} = \sum_{k=1}^{n} \sum_{s=1}^{n} c^i_{ik} c^k_{ji} \\ &= K_{\mathfrak{g}}(X_i, X_j). \end{aligned}$$

□

5.4.0.4. *The Universal Enveloping Algebra.* In a Lie algebra $\mathfrak{g}$ the product $[.,.]$ is not associative except in the trivial case that $[.,.] \equiv 0$. On the other hand, associative algebras play an important role in the study of Lie algebras. For one thing, if $\mathfrak{A}$ is an associative algebra then we can introduce the commutator bracket on $\mathfrak{A}$ by

$$[A, B] := AB - BA$$

which gives $\mathfrak{A}$ the structure of Lie algebra. From the other direction, if we start with a Lie algebra $\mathfrak{g}$ then we can construct an associative algebra called the universal enveloping algebra of $\mathfrak{g}$. This is done, for instance, by first forming the full tensor algebra on $\mathfrak{g}$;

$$T(\mathfrak{g}) = \mathbb{F} \oplus \mathfrak{g} \oplus (\mathfrak{g} \otimes \mathfrak{g}) \oplus \cdots \oplus \mathfrak{g}^{\otimes k} \oplus \cdots$$

and then dividing out by an appropriate ideal:

Definition 5.23. Associated to every Lie algebra $\mathfrak{g}$ there is an associative algebra $U(\mathfrak{g})$ called the **universal enveloping algebra** defined by

$$U(\mathfrak{g}) := T(\mathfrak{g})/J$$

where J is the ideal generated by elements in $T(\mathfrak{g})$ of the form $X \otimes Y - Y \otimes X - [X, Y]$.

There is the natural map of $\mathfrak{g}$ into $U(\mathfrak{g})$ given by the composition $\pi : \mathfrak{g} \hookrightarrow T(\mathfrak{g}) \to T(\mathfrak{g})/J = U(\mathfrak{g})$. For $v \in \mathfrak{g}$, let v^* denote the image of v under this canonical map.

Theorem 5.24. *Let* V *be a vector space over the field* $\mathbb{F}$. *For every* ρ *representation of* $\mathfrak{g}$ *on* V *there is a corresponding representation* ρ^* *of* $U(\mathfrak{g})$ *on* V *such that for all* $v \in \mathfrak{g}$ *we have*

$$\rho(v) = \rho^*(v^*).$$

This correspondence, $\rho \mapsto \rho^*$ *is a 1-1 correspondence.*

Proof. Given ρ, there is a natural representation $T(\rho)$ on $T(\mathfrak{g})$. The representation $T(\rho)$ vanishes on J since

$$T(\rho)(X \otimes Y - Y \otimes X - [X,Y]) = \rho(X)\rho(Y) - \rho(Y)\rho(X) - \rho([X,Y]) = 0$$

and so $T(\rho)$ descends to a representation ρ^* of $U(\mathfrak{g})$ on $\mathfrak{g}$ satisfying $\rho(v) = \rho^*(v^*)$. Conversely, if σ is a representation of $U(\mathfrak{g})$ on V then we put $\rho(X) = \sigma(X^*)$. The map $\rho(X)$ is linear and a representation since

$$\begin{aligned} \rho([X,Y]) &= \sigma([X,Y]^*) \\ &= \sigma(\pi(X \otimes Y - Y \otimes X)) \\ &= \sigma(X^*Y^* - Y^*X^*) \\ &= \rho(X)\rho(Y) - \rho(Y)\rho(X) \end{aligned}$$

for all $X, Y \in \mathfrak{g}$. $\square$

Now let $X_1, X_2, ..., X_n$ be a basis for $\mathfrak{g}$ and then form the monomials $X_{i_1}^* X_{i_2}^* \cdots X_{i_r}^*$ in $U(\mathfrak{g})$. The set of all such monomials for a fixed r span a subspace of $U(\mathfrak{g})$, which we denote by $U^r(\mathfrak{g})$. Let c_{ik}^j be the structure constants for the basis $X_1, X_2, ..., X_n$. Then under the map π the structure equations become

$$[X_i^*, X_j^*] = \sum_k c_{ij}^k X_k^*.$$

By using this relation we can replace the spanning set $\mathcal{M}_r = \{X_{i_1}^* X_{i_2}^* \cdots X_{i_r}^*\}$ for $U^r(\mathfrak{g})$ by spanning set $\mathcal{M}_{\leq r}$ for $U^r(\mathfrak{g})$ consisting of all monomials of the form $X_{i_1}^* X_{i_2}^* \cdots X_{i_m}^*$ where $i_1 \leq i_2 \leq \cdots \leq i_m$ and $m \leq r$. In fact one can then concatenate these spanning sets $\mathcal{M}_{\leq r}$ and it turns out that these combine to form a basis for $U(\mathfrak{g})$. We state the result without proof:

Theorem 5.25 (Birchoff-Poincarè-Witt)**.** *Let* $e_{i_1 \leq i_2 \leq \cdots \leq i_m} = X_{i_1}^* X_{i_2}^* \cdots X_{i_m}^*$ *where* $i_1 \leq i_2 \leq \cdots \leq i_m$. *The set of all such elements* $\{e_{i_1 \leq i_2 \leq \cdots \leq i_m}\}$ *for all* m *is a basis for* $U(\mathfrak{g})$ *and the set* $\{e_{i_1 \leq i_2 \leq \cdots \leq i_m}\}_{m \leq r}$ *is a basis for the subspace* $U^r(\mathfrak{g})$.

Lie algebras and Lie algebra representations play an important role in physics and mathematics, and as we shall see below, every Lie group has an associated Lie algebra that, to a surprisingly large extent, determines the structure of the Lie group itself. Let us first explore some of the important abstract properties of Lie algebras. A notion that is useful for constructing Lie algebras with desired properties is that of the **free Lie algebra** $\mathfrak{f}_n$ which is defined to be the quotient of the free algebra on n symbols by the smallest ideal such that we end up with a Lie algebra. Every Lie algebra can be realized as a quotient of one of these free Lie algebras.

Definition 5.26. The **descending central series** $\{\mathfrak{g}_{(k)}\}$ of a Lie algebra $\mathfrak{g}$ is defined inductively by letting $\mathfrak{g}_{(1)} = \mathfrak{g}$ and then $\mathfrak{g}_{(k+1)} = [\mathfrak{g}_{(k)}, \mathfrak{g}]$.

The reason for the term “descending” is the that we have the chain of inclusions

$$\mathfrak{g}_{(1)} \supset \cdots \supset \mathfrak{g}_{(k)} \supset \mathfrak{g}_{(k+1)} \supset \cdots$$

From the definition of Lie algebra homomorphism we see that if $\sigma : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism then $\sigma(\mathfrak{g}_{(k)}) \subset \mathfrak{h}_{(k)}$.

Exercise 5.27 (!)**.** Use the Jacobi identity to prove that for all positive integers i and j, we have $[\mathfrak{g}_{(i)}, \mathfrak{g}_{(i)}] \subset \mathfrak{g}_{(i+j)}$.

Definition 5.28. A Lie algebra $\mathfrak{g}$ is called **k-step nilpotent** if and only if $\mathfrak{g}_{(k+1)} = 0$ but $\mathfrak{g}_{(k)} \neq 0$.

The most studied nontrivial examples are the Heisenberg algebras which are 2-step nilpotent. These are defined as follows:

Example 5.29. The $2n+1$ dimensional **Heisenberg algebra** $\mathfrak{h}_n$ is the Lie algebra (defined up to isomorphism) with a basis $\{X_1, ..., X_n, Y_1, ..., Y_n, Z\}$ subject to the relations

$$[X_j, Y_j] = Z$$

and all other brackets of elements from this basis being zero. A concrete realization of $\mathfrak{h}_n$ is given as the set of all $(n+2) \times (n+2)$ matrices of the form

$$\begin{bmatrix} 0 & x_1 & ... & x_n & z \\ 0 & 0 & ... & 0 & y_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & y_n \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}$$

where x_i, y_i, z are all real numbers. The bracket is the commutator bracket as is usually the case for matrices. The basis is realized in the obvious way by putting a lone 1 in the various positions corresponding to the potentially nonzero entries. For example,

$$X_1 = \begin{bmatrix} 0 & 1 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}$$

and

$$Z = \begin{bmatrix} 0 & 0 & ... & 0 & 1 \\ 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}.$$

Example 5.30. The space of all upper triangular $n \times n$ matrices $\mathfrak{n}_n$ which turns out to be $n-1$ step nilpotent.

We also have the **free k-step nilpotent Lie algebra** given by the quotient $\mathfrak{f}_{n,k} := \mathfrak{f}_n/(\mathfrak{f}_n)_k$ where $\mathfrak{f}_n$ is the free Lie algebra mentioned above. (notice the difference between $\mathfrak{f}_{n,k}$ and $(\mathfrak{f}_n)_k$.

Lemma 5.31. *Every finitely generated k-step nilpotent Lie algebra is isomorphic to a quotient of the free k-step nilpotent Lie algebra.*

Proof. Suppose that $\mathfrak{g}$ is k-step nilpotent and generated by elements $X_1, ..., X_n$. Let $F_1, ..., F_n$ be the generators of $\mathfrak{f}_n$ and define a map $h : \mathfrak{f}_n \to \mathfrak{g}$ by sending F_i to X_i and extending linearly. This map clearly factors through $\mathfrak{f}_{n,k}$ since $h((\mathfrak{f}_n)_k) = 0$. Then we have a homomorphism $(\mathfrak{f}_n)_k \to \mathfrak{g}$ that is clearly onto and so the result follows. □

Definition 5.32. Let $\mathfrak{g}$ be a Lie algebra. We define the commutator series $\{\mathfrak{g}^{(k)}\}$ by letting $\mathfrak{g}^{(1)} = \mathfrak{g}$ and then inductively $\mathfrak{g}^{(k)} = [\mathfrak{g}^{(k-1)}, \mathfrak{g}^{(k-1)}]$. If $\mathfrak{g}^{(k)} = 0$ for some positive integer k, then we call $\mathfrak{g}$ a **solvable** Lie algebra.

Clearly, the statement $\mathfrak{g}^{(2)} = 0$ is equivalent to the statement that $\mathfrak{g}$ is abelian. Another simple observation is that $\mathfrak{g}^{(k)} \subset \mathfrak{g}_{(k)}$ so that nilpotency implies solvability.

Exercise 5.33 (!)**.** Every subalgebra and every quotient algebra of a solvable Lie algebra is solvable. In particular, the homomorphic image of a solvable Lie algebra is solvable. Conversely, if $\mathfrak{a}$ is a solvable ideal in $\mathfrak{g}$ and $\mathfrak{g}/\mathfrak{a}$ is solvable, then $\mathfrak{g}$ is solvable. Hint: Use that $(\mathfrak{g}/\mathfrak{a})^{(j)} = \mathfrak{g}^{(j)}/\mathfrak{a}$.

It follows from this exercise that we have

Corollary 5.34. Let $h : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra homomorphism. If $\mathrm{img}(h) := h(\mathfrak{g})$ and $\ker(h)$ are both solvable then $\mathfrak{g}$ is solvable. In particular, if $\mathrm{img}(\mathrm{ad}) := \mathrm{ad}(\mathfrak{g})$ is solvable then so is $\mathfrak{g}$.

Lemma 5.35. *If $\mathfrak{a}$ is a nilpotent ideal in $\mathfrak{g}$ contained in the center $\mathfrak{z}(\mathfrak{g})$ and if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then $\mathfrak{g}$ is nilpotent.*

Proof. First, the reader can verify that $(\mathfrak{g}/\mathfrak{a})_{(j)} = \mathfrak{g}_{(j)}/\mathfrak{a}$. Now if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then $\mathfrak{g}_{(j)}/\mathfrak{a} = 0$ for some j and so $\mathfrak{g}_{(j)} \subset \mathfrak{a}$ and if this is the case then we have $\mathfrak{g}_{(j+1)} = [\mathfrak{g}, \mathfrak{g}_{(j)}] \subset [\mathfrak{g}, \mathfrak{a}] = 0$. (Here we have $[\mathfrak{g}, \mathfrak{a}] = 0$ since $\mathfrak{a} \subset \mathfrak{z}(\mathfrak{g})$.) Thus $\mathfrak{g}$ is nilpotent. □

Trivially, the center $\mathfrak{z}(\mathfrak{g})$ of a Lie algebra a solvable ideal.

Corollary 5.36. Let $h : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra homomorphism. If $\mathrm{im}(\mathrm{ad}) := \mathrm{ad}(\mathfrak{g})$ is nilpotent then $\mathfrak{g}$ is nilpotent.

Proof. Just use the fact that $\ker(\mathrm{ad}) = \mathfrak{z}(\mathfrak{g})$. □

Theorem 5.37. *The sum of any family of solvable ideals in $\mathfrak{g}$ is a solvable ideal. Furthermore, there is a unique maximal solvable ideal that is the sum of all solvable ideals in $\mathfrak{g}$.*

Sketch of proof. The proof is a maximality argument based on the following idea : If $\mathfrak{a}$ and $\mathfrak{b}$ are solvable then $\mathfrak{a} \cap \mathfrak{b}$ is an ideal in the solvable $\mathfrak{a}$ and so is solvable. It is easy to see that $\mathfrak{a} + \mathfrak{b}$ is an ideal. We have by exercise 5.13 $(\mathfrak{a} + \mathfrak{b})/\mathfrak{b} \cong \mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$. Since $\mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$ is a homomorphic image of $\mathfrak{a}$ we see that $\mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b}) \cong (\mathfrak{a} + \mathfrak{b})/\mathfrak{b}$ is solvable. Thus by our previous result $\mathfrak{a} + \mathfrak{b}$ is solvable. □

Definition 5.38. The maximal solvable ideal in $\mathfrak{g}$ whose existence is guaranteed by the last theorem is called the **radical** of $\mathfrak{g}$ and is denoted $rad(\mathfrak{g})$

Definition 5.39. A Lie algebra $\mathfrak{g}$ is called **simple** if it contains no ideals other than $\{0\}$ and $\mathfrak{g}$. A Lie algebra $\mathfrak{g}$ is called **semisimple** if it contains no abelian ideals (other than $\{0\}$).

Theorem 5.40 (Levi decomposition)**.** *Every Lie algebra is the semi-direct sum of its radical and a semisimple Lie algebra.*

The map $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}) = \mathrm{End}(T_eG)$ is given as the tangent map at the identity of Ad which is a Lie algebra homomorphism. Thus by proposition **??** we have obtain

Proposition 5.41. $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ is a Lie algebra homomorphism.

Proof. This follows from our study of abstract Lie algebras and proposition **??**. □

Let's look at what this means. Recall that the Lie bracket for $\mathfrak{gl}(\mathfrak{g})$ is just $A \circ B - B \circ A$. Thus we have

$$\operatorname{ad}([v,w]) = [\operatorname{ad}(v), \operatorname{ad}(w)] = \operatorname{ad}(v) \circ \operatorname{ad}(w) - \operatorname{ad}(w) \circ \operatorname{ad}(v)$$

which when applied to a third vector z gives

$$[[v,w],z] = [v,[w,z]] - [w,[v,z]]$$

which is just a version of the Jacobi identity. Also notice that using the antisymmetry of the bracket we get $[z,[v,w]] = [w,[z,v]] + [v,[z,w]]$ which in turn is the same as

$$\operatorname{ad}(z)([v,w]) = [\operatorname{ad}(z)v, w] + [v, \operatorname{ad}(z)w]$$

so $\operatorname{ad}(z)$ is actually a derivation of the Lie algebra $\mathfrak{g}$ as explained before.

Proposition 5.42. The Lie algebra $Der(\mathfrak{g})$ of all derivation of $\mathfrak{g}$ is the Lie algebra of the group of automorphisms $\operatorname{Aut}(\mathfrak{g})$. The image $\operatorname{ad}(\mathfrak{g}) \subset Der(\mathfrak{g})$ is the Lie algebra of the set of all inner automorphisms $\operatorname{Int}(\mathfrak{g})$.

$$\begin{array}{ccc} \operatorname{ad}(\mathfrak{g}) & \subset & \operatorname{Der}(\mathfrak{g}) \\ \downarrow & & \downarrow \\ \operatorname{Int}(\mathfrak{g}) & \subset & \operatorname{Aut}(\mathfrak{g}) \end{array}$$

Let $\mu : G \times G \to G$ be the multiplication map. Recall that the tangent space $T_{(g,h)}(G \times G)$ is identified with $T_g G \times T_h G$. Under this identification we have

$$T_{(g,h)}\mu(v,w) = T_h L_g w + T_g R_h v$$

where $v \in T_g G$ and $w \in T_h G$. The following diagrams exhibit the relations:

$$\begin{array}{ccccc} & & G \times G, (g,h) & & \\ & \overset{pr_1}{\swarrow} & \| & \overset{pr_2}{\searrow} & \\ G, g & \to & G \times G, (g,h) & \leftarrow & G, h \\ & \underset{R_h}{\searrow} & \downarrow \mu & \underset{L_g}{\swarrow} & \\ & & G, gh & & \end{array}$$

The horizontal maps are the insertions $g \mapsto (g,h)$ and $h \mapsto (g,h)$. Applying the tangent functor to the last diagram gives.

$$\begin{array}{ccccc} Tpr_1 & & T_{(g,h)}(G \times G) & & Tpr_2 \\ & \swarrow & \updownarrow & \searrow & \\ T_g G & \to & T_g G \times T_h G & \leftarrow & T_h G \\ & \searrow & \downarrow T\mu & \swarrow & \\ T_g R_h & & T_{gh} G & & T_h L_g \end{array}$$

We will omit the proof but the reader should examine the diagrams and try to construct a proof on that basis.

We have another pair of diagrams to consider. Let $\nu : G \to G$ be the inversion map $\nu : g \mapsto g^{-1}$. We have the following commutative diagrams:

$$\begin{array}{ccccc} R_{g^{-1}} & & (G,g) & & L_{g^{-1}} \\ & \swarrow & & \searrow & \\ (G,e) & & \downarrow \nu & & (G,e) \\ & \searrow & & \swarrow & \\ L_{g^{-1}} & & \left(G,g^{-1}\right) & & R_{g^{-1}} \end{array}$$

Applying the tangent functor we get

$$\begin{array}{ccccc} TR_{g^{-1}} & & T_gG & & TL_{g^{-1}} \\ & \swarrow & & \searrow & \\ T_eG & & \downarrow T\nu & & T_eG \\ & \searrow & & \swarrow & \\ TL_{g^{-1}} & & T_{g^{-1}}G & & TR_{g^{-1}} \end{array} .$$

The result we wish to express here is that $T_g\nu = TL_{g^{-1}} \circ TR_{g^{-1}} = TR_{g^{-1}} \circ TL_{g^{-1}}$. Again the proof follows readily from the diagrams.

5.5. Geometry of figures in Euclidean space

We use differential forms in this section. We take as our first example the case of a Euclidean space. We intend to study figures in $\mathbf{E}^n$ but we first need to set up some machinery using differential forms and moving frames. Let e_1, e_2, e_3 be a moving frame on $\mathbf{E}^3$. Using the identification of $T_x\mathbf{E}^3$ with $\mathbb{R}^3$ we may think of each e_i as a function with values in $\mathbb{R}^3$. If x denotes the identity map then we interpret dx as the map $T\mathbf{E}^3 \to \mathbb{R}^3$ given by composing the identity map on the tangent bundle $T\mathbf{E}^3$ with the canonical projection map $T\mathbf{E}^3 = \mathbf{E}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$. If $\theta^1, \theta^2, \theta^3$ is the frame field dual to e_1, e_2, e_3 then we may write

$$dx = \sum e_i\theta^i. \tag{5.3}$$

Also, since we are interpreting each e_i as an $\mathbb{R}^3-$valued function we may take the componentwise exterior derivative to make sense of de_i. Then de_i is a vector valued differential form: If $e_i = f_1\mathbf{i} + f_2\mathbf{j} + f_3\mathbf{k}$ then $de_i = df_1\mathbf{i} + df_2\mathbf{j} + df_3\mathbf{k}$. We may write

$$de_j = \sum e_i\omega^i_j \tag{5.4}$$

for some set of 1−forms ω^i_j which we arrange in a matrix $\omega = (\omega^i_j)$. If we take exterior derivative of equations 5.3 and 5.4 For the first one we calculate

$$\begin{aligned}
0 = ddx &= \sum_{i=1}^n e_i \theta^i \\
&= \sum_{i=1}^n de_i \wedge \theta^i + \sum_{i=1}^n e_i \wedge d\theta^i \\
&= \sum_{i=1}^n \left(\sum_{j=1}^n e_j \omega^i_j \right) \wedge \theta^i + \sum_{i=1}^n e_i \wedge d\theta^i .
\end{aligned}$$

From this we get the first of the following two structure equations. The second one is obtained similarly from the result of differentiating 5.4.

$$\begin{aligned}
d\theta^i &= -\sum \omega^i_j \wedge \theta^j \\
d\omega^i_j &= -\sum \omega^i_k \wedge \omega^k_j
\end{aligned} \tag{5.5}$$

Furthermore, if we differentiate $e_i \cdot e_j = \delta_{ij}$ we find out that $\omega^i_j = -\omega^j_i$.

If we make certain innocent identifications and conventions we can relate the above structure equations to the group Euc(n) and its Lie algebra. We will identify $\mathbf{E}^n$ with the set of column vectors of the form

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

Then the group Euc(n) is presented as the set of all square matrices of the form

$$\begin{bmatrix} 1 & 0 \\ v & Q \end{bmatrix} \text{ where } Q \in \mathrm{O}(n) \text{ and } v \in \mathbb{R}^n.$$

The action $\mathrm{Euc}(n) \times \mathbf{E}^n \to \mathbf{E}^n$ is then simply given by matrix multiplication (see chapter **??**). One may easily check that the matrix Lie algebra that we identify as the Lie algebra $\mathfrak{euc}(n)$ of Euc(n) consists of all matrices of the form

$$\begin{bmatrix} 0 & 0 \\ v & A \end{bmatrix} \text{ where } v \in \mathbb{R}^n \text{ and } A \in \mathfrak{so}(n) \text{ (antisymmetric matrices)}$$

The isotropy of the point $o := \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is easily seen to be the subgroup G_o consisting of all elements of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \text{ where } Q \in \mathrm{O}(n)$$

which is clearly isomorphic to O(n). This isotropy group G_o is just the group of rotations about the origin. The origin is not supposed to special and so we should point out that any point would work fine for what we are about to do. In fact, for any other point $x \sim \begin{bmatrix} 1 \\ x \end{bmatrix}$ we have an isomorphism $G_o \cong G_x$ given by $h \mapsto t_x h t_x$ where

$$t_x = \begin{bmatrix} 1 & 0 \\ x & I \end{bmatrix}.$$

(see exercise **??**).

For each $x \in \mathbf{E}^n$ tangent space $T_x\mathbf{E}^n$ consists of pairs $x \times v$ where $v \in \mathbb{R}^n$ and so the dot product on $\mathbb{R}^n$ gives an obvious inner product on each $T_x\mathbf{E}^n$: For two tangent vectors in $T_x\mathbf{E}^n$, say $v_x = x \times v$ and $w_x = x \times w$ we have $\langle v_x, w_x \rangle = v \cdot w$.

Remark 5.43. The existence of an inner product in each tangent space makes $\mathbf{E}^n$ a Riemannian manifold (a smoothness condition is also needed). Riemannian geometry (studied in chapter **??**) is one possible generalization of Euclidean geometry (See figure 17 and the attendant discussion). Riemannian geometry represents an approach to geometry that is initially quite different in spirit from Klein's approach.

Now think of each element of the frame $e_1, ..., e_n$ as a column vector of functions and form a matrix of functions e. Let x denote the "identity map" given as a column vector of functions $x = (x^1, x^2, x^3) = (x, y, z)$. Next we form the matrix of functions

$$\begin{bmatrix} 1 & 0 \\ x & e \end{bmatrix}$$

This just an element of $\mathrm{Euc}(n)$! Now we can see that the elements of $\mathrm{Euc}(n)$ are in a natural $1-1$ correspondents with the set of all frames on E^n and the matrix we have just introduce corresponds exactly to the moving frame $x \mapsto (e_1(x), ..., e_n(x))$. The differential of this matrix gives a matrix of one forms

$$\varpi = \begin{bmatrix} 0 & 0 \\ \theta & \omega \end{bmatrix}$$

and it is not hard to see that θ is the column consisting of the same θ^i as before and also that $\omega = (\omega^i_j)$. Also, notice that $x \mapsto \varpi(x) = \begin{bmatrix} 0 & 0 \\ \theta & \omega \end{bmatrix}$ takes values in the Lie algebra $\mathfrak{euc}(n)$. This looking like a very natural state of affairs. In fact, the structure equations are encoded as a singe matrix equation

$$d\varpi = \varpi \wedge \varpi.$$

The next amazingly cool fact it that if we pull-back ϖ to $\mathrm{Euc}(n)$ via the projection $\pi : \begin{bmatrix} 0 & 0 \\ x & e \end{bmatrix} \mapsto x \in \mathbf{E}^n$ then we obtain the Maurer-Cartan form of the group $\mathrm{Euc}(n)$ and the equation $d\varpi = \varpi \wedge \varpi$ pulls back to the structure equations for $\mathrm{Euc}(n)$.

Chapter 6 Supplement

6.1. Comments and Errata

6.1.1. Comments. In the definition of structure groups for fibers bundles we cosidered only the case that the action is effective. Then since the corresponding homomorphism $G \to \mathrm{Diff}(F)$ given by $g \mapsto \lambda_g$ is injective, the reader may feel that the group G can be replaced by its image in $\mathrm{Diff}(F)$. Indeed, by transferring the Lie group structure by brute force onto this image, one may indeed replace G by its image. However, it is not always desirable nor convenient to do so.

6.1.2. Errata. (Nothing yet)

6.2. sections of the Mobius band

Consider the C^0-vector bundle obtained in the following way. Start with the first factor projection $[0,1] \times \mathbb{R} \to [0,1]$. On $[0,1] \times \mathbb{R}$ define the equivalence relation $(0,a) \sim (1,-a)$ and let $Mb = ([0,1] \times \mathbb{R})/\!\sim$ with the quotient topology. On $[0,1]$, define an equivalence relation by identifying 0 with 1. Then $[0,1]/\!\sim$ is identified with S^1. The first factor projection induces a continuous map

$$\pi : Mb \to S^1$$

which is easily seen to be a vector bundle of rank 1. In the category of general continuous bundles over S^1, this bundle is equivalent to the Mobius band bundle introduced in the main text. This bundle can be given a smooth vector bundle structure quite easily, but we only need the C^0 version to make our point about global nonzero sections. This is the vector bundle version of the Mobius band and is also referred to by the same name. By considering the quotient construction used to obtain $\pi : Mb \to S^1$ we see that a global continuous section σ of $\pi : Mb \to S^1$ must correspond, under the quotient, to a continuous map $\widetilde{\sigma} : [0,1] \to \mathbb{R}$ such that $\widetilde{\sigma}(0) = -\widetilde{\sigma}(1)$. By continuity, such a map must be zero at some point which forces σ to be zero also. Thus $\pi : Mb \to S^1$ is not trivial.

6.3. Etale space of a sheaf

Let $E \to M$ be a vector bundle and consider the sheal for local sections. Recall that $s_1 \in \Gamma_E(U)$ and $s_2 \in \Gamma_E(V)$ determine the same germ of sections at p if there is an open set $W \subset U \cap V$ such that $r^U_W s_1 = r^V_W s_2$. We impose an equivalence relation on the union

$$\bigcup_{p \in U} \Gamma_E(U)$$

by declaring that $s_1 \sim s_2$ if and only if s_1 and s_2 determine the same germ of sections at p. The set of equivalence classes (called germs of sections at p) is an abelian group in the obvious way and is denoted Γ^E_p. If we are dealing with a sheaf of rings, then Γ^E_p is a ring. The set $\Gamma_E((U)) := \bigcup_{p \in U} \Gamma^E_p$ is called the sheaf of germs and can be given a topology so that the projection

$$\mathrm{pr} : \Gamma_E((U)) \to M,$$

defined by the requirement that $\mathrm{pr}([s]) = p$ if $[s] \in \mathcal{S}^E_p$, is a local homeomorphism. The idea generalizes:

Definition 6.1. $\mathcal{M}_p$ is a set of equivalence classes called germs at p. Here $s_1 \in \mathcal{M}(U)$ and $s_2 \in \mathcal{M}(V)$ determine the same germ of sections at p if there is an open set $W \subset U \cap V$ containing p such that $r^U_W s_1 = r^V_W s_2$. The germ of $s \in \mathcal{M}(U)$ at p is denoted s_p.

Now we form the union $\widetilde{\mathcal{M}} = \bigcup_{p \in M} \mathcal{M}_p$ and define a surjection $\pi : \widetilde{\mathcal{M}} \to M$ by the requirement that $\pi(s_p) = p$ for $s_p \in \mathcal{M}_p$. The space $\widetilde{\mathcal{M}}$ is called the **sheaf of germs** generated by $\mathcal{M}$. We want to topologize $\widetilde{\mathcal{M}}$ so that π is continuous and a local homeomorphism, but first we need a definition:

Definition 6.2 (étalé space)**.** A topological space Y together with a continuous surjection $\pi : Y \to M$ that is a local homeomorphism is called an **étalé space**. A local section of an étalé space over an open subset $U \subset M$ is a map $s_U : U \to Y$ such that $\pi \circ s_U = \mathrm{id}_U$. The set of all such sections over U is denoted $\Gamma(U, Y)$.

Definition 6.3. For each $s \in \mathcal{M}(U)$, we can define a map (of sets) $\widetilde{s} : U \to \widetilde{\mathcal{M}}$ by

$$\widetilde{s}(x) = s_x$$

and we give $\widetilde{\mathcal{M}}$ the smallest topology such that the images $\widetilde{s}(U)$ for all possible U and s are open subsets of $\widetilde{\mathcal{M}}$.

With the above topology $\widetilde{\mathcal{M}}$ becomes an étalé space and all the sections $\widetilde{s}$ are continuous open maps. If we let $\widetilde{\mathcal{M}}(U)$ denote the sections over U for this étalé space, then the assignment $U \to \widetilde{\mathcal{M}}(U)$ is a presheaf that is always a sheaf. We give the following without proof:

Proposition 6.4. If $\mathcal{M}$ is a sheaf, then $\widetilde{\mathcal{M}}$ is isomorphic as a sheaf to $\mathcal{M}$.

6.4. Discussion on G bundle structure

It is well understood that one starting point for defining a G-bundle structure on a locally trivial fibration is the idea that a G-bundle structure is determined by a bundle atlas that in some sense takes values in a Lie group or topological group G. This is similar to the idea that a C^{∞}-atlas determines a smooth structure on a locally Euclidean paracompact topological space. But when do two G-bundle atlases determine the very same G-bundle structure? There are several ways one may describe the situation but we will show that one way which occurs in the literature is actually wrong. This wrong way actually describes a special case of a different equivalence relation that says when two *different* G-bundles are equivalent but not necessarily the same. It does *not* reduce to the correct notion in this special case. This is analogous to the notion of two *different* smooth structures on a fixed topological manifold being nevertheless diffeomorphic. One does not want to confuse the notion of diffeomorphism class with the notion of a smooth structure. One would also like to avoid a similar confusion in the definition and classification of G-bundles. That is we do not want to confuse G-bundle structures with the notion of equivalence classes of G-bundle structures (even on a single fibration).

The bottom line will be that Steenrod got it right in his famous book "The topology of fiber bundles". The key overlooked notion is that of **strict equivalence** defined on page 8 of the 1999 Princeton edition. Strict equivalence is to be contrasted with the second important notion of equivalence mentioned above that is cohomological in nature. There are several places in the literature, including textbooks, where the notion of strict equivalence is conflated with the weaker cohomological notion. We shall convince the reader that these are importantly different and have different roles and do not coincide even if one is dealing with a single fibration (fixed total space and base space).

6.4.1. Two notions of equivalence. For concreteness let us work in the smooth category. Let $\pi : E \to M$ be a locally trivial fibration with typical fiber F. Suppose that $\rho : G \times F \to F$ is an appropriately smooth action of a Lie group G on F so that G acts by diffeomorphisms of F. (We will think in terms of actions rather than representations).

Recall that each bundle chart $(\varphi_\alpha, U_\alpha)$ gives a diffeomorphism $\varphi_{\alpha,x} : E_x \to F$ for each $x \in U_\alpha$. Notice that each φ_α must have the form $\varphi_\alpha = (\pi, \Phi_\alpha)$ for some smooth map $\Phi_\alpha : \pi^{-1}(U_\alpha) \to F$. Then $\varphi_{\alpha,x} := \Phi_\alpha|_{E_x}$. If $(\varphi_\alpha, U_\alpha)$ and (φ_β, U_β) are bundle charts for $\pi : E \to M$, then the "raw" transition functions (between the charts) are defined by

$$\varphi_{\alpha\beta}(x) := \varphi_{\alpha,x} \circ \varphi_{\beta,x}^{-1}$$

and for each $x \in U_\alpha \cap U_\beta$, the map $\varphi_{\alpha\beta}(x) : F \to F$ is a diffeomorphism of F.

Definition 6.5. A (G, ρ)-atlas consists of a cover of the bundle by local trivializations or bundle charts $(\varphi_\alpha, U_\alpha)$ such that there exist smooth maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ defined whenever $U_\alpha \cap U_\beta \neq \emptyset$, such that

$$\varphi_{\alpha\beta}(x)(y) = \rho(g_{\alpha\beta}(x), y) \text{ for } y \in F.$$

We further require that the family $\{g_{\alpha\beta}\}$ satisfy the cocycle conditions

$$\begin{aligned} g_{\alpha\alpha}(p) &= e \quad \text{for } p \in U_\alpha \\ g_{\alpha\beta}(p) &= (g_{\beta\alpha}(p))^{-1} \quad \text{for } p \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(p) g_{\beta\gamma}(p) g_{\gamma\alpha}(p) &= e \quad \text{for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned}$$

Notice that the cocycle condition would be automatic if the action $\rho : G \times F \to F$ were assumed to be an effective action. Otherwise the cocycle condition is necessary to keep contact with the important notions of principal bundle and associated bundle which are the basic notions in an alternative development of G-bundles (which seems to be due to Ehresmann). Also notice that for an ineffective action, the functions $\varphi_{\alpha\beta}(x)$ do not uniquely determine the G-valued functions $g_{\alpha\beta}(x)$! Thus if we do not restrict ourselves to effective actions (as does Steenrod) we must actually include the the cocycle as part of the data. Thus we must specify

$$\text{Data} := (E, \pi, M, F, G, \rho, \{(\varphi_\alpha, U_\alpha)\}_{\alpha \in A}, \{g_{\alpha\beta}\}).$$

Indeed, it may be the case that there is another quite different cocycle $\{g'_{\alpha\beta}\}$ such that $\varphi_{\alpha\beta}(x)(y) = \rho(g'_{\alpha\beta}(x), y)$ for all $x \in U_\alpha \cap U_\beta$ and $y \in F$. One only needs to have $g_{\alpha\beta}(x) \left(g'_{\alpha\beta}(x) \right)^{-1}$ in the kernel of the action for each x. This causes all the definitions to be far more complicated in many places than those found in Steenrod since he restricts himself to effective actions. This is true even for the notion of strict equivalence that we wish to explore next.

So as to not obscure the main issue let us restrict ourselves effective actions.

We now come to the main question:

When do two (G,ρ)-atlases define the same (G,ρ)-bundle structure?

This is a question of how to make an appropriate definition. We can phrase the question by asking when the data

$$\mathrm{Data}_1 := (E, \pi, M, F, G, \rho, \{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A})$$

is appropriately equivalent to similar data

$$\mathrm{Data}_2 := (E, \pi, M, , F, G, \rho, \{(\psi_j, V_j)\}_{j\in J}).$$

The most straightforward notion of equivalence for this purpose is the following

Definition 6.6 (strict equivalence)**.** Let ρ be effective. If $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ and $\{(\psi_j, V_j)\}_{j\in J}$ are both (G,ρ)-atlases on a bundle (E,π,M) then $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ is **strictly equivalent** to $\{(\psi_j, V_j)\}_{j\in J}$ if $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A} \bigcup \{(\psi_j, V_j)\}_{j\in J}$ is a (G,ρ)-atlas (with new larger and uniquely determined cocycle since ρ is effective).

This is equivalent to Steenrod's definition:

Definition 6.7 (strict equivalence 2)**.** Let ρ be effective. If $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ and $\{(\psi_j, V_j)\}_{j\in J}$ are both (G,ρ)-atlases on a bundle (E,π,M) then $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ is **strictly equivalent** to $\{(\psi_j, V_j)\}_{j\in J}$ if for each $\alpha \in A$ and $j \in J$ such that $U_\alpha \cap U_\beta \neq \emptyset$ we have that there are smooth maps $g_{\alpha j} : U_\alpha \cap V_j \to G$ and $g_{j\alpha} : U_\alpha \cap V_j \to G$ such that

$$\varphi_{\alpha,x} \circ \psi_{j,x}^{-1}(y) = \rho(g_{\alpha j}(x), y)$$
$$\psi_{j,x} \circ \varphi_{\alpha,x}^{-1}(y) = \rho(g_{j\alpha}(x), y)$$

for all $x \in U_\alpha \cap V_j$ and $y \in F$.

It is not hard to show that the above two definitions are equivalent and we can define a *(G,ρ)-bundle structure* to be a **strict** equivalence class of *(G,ρ)-atlases.*

Definition 6.8. A *(G,ρ)-bundle structure* is a **strict** equivalence class of *(G,ρ)-atlases.*

Now consider the following notion of equivalence, which we claim is weaker:

Definition 6.9. Let ρ be effective. If $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ and $\{(\psi_j, V_j)\}_{j\in J}$ are both (G,ρ)-atlases on a bundle (E,π,M) then $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ is **equivalent$_2$** to $\{(\psi_j, V_j)\}_{j\in J}$ if there are smooth maps $\widetilde{g}_{j\alpha} : U_\alpha \cap V_j \to G$ and $\widetilde{g}_{\alpha j} : U_\alpha \cap V_j \to G$ for each $\alpha \in A$ and $j \in J$ such that

$$\widetilde{g}_{j\alpha}(x) = \widetilde{g}_{j\beta}(x) g_{\beta\alpha}(x) \text{ for all } x \in U_\alpha \cap U_\beta \cap V_j$$
$$\widetilde{g}_{l\alpha}(x) = g'_{lk}(x) \widetilde{g}_{k\alpha}(x) \text{ for all } x \in U_\alpha \cap V_k \cap V_l$$

and similarly for $\widetilde{g}_{\alpha j}$. Here $g_{\beta\alpha}(x)$ and $g'_{lk}(x)$ are the cocycles uniquely determined by $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ and $\{(\psi_j, V_j)\}_{j\in J}$ respectively (**notice that we need effectiveness of the action here!**)

This definition is a bit cumbersome and tough to think about, but notice that the sets of the form $U_\alpha \cap V_j$ cover M and form a refinement of both $\{U_\alpha\}$and $\{V_j\}$ so that $\{(\varphi_\alpha, U_\alpha)\}_{\alpha\in A}$ and $\{(\psi_j, V_j)\}_{j\in J}$ are each strictly equivalent to atlases with the same cover (**but not necessarily strictly equivalent to each other as we shall see**). Call these atlases $\{(\varphi_i, U_i)\}_{i\in I}$ and $\{(\psi_i, U_i)\}_{i\in I}$ and notice that we have a new indexing set I and we have renamed the cover for simplicity.

It is not hard to show that the following is an alternative characterization of **equivalence$_2$** in this case.

Definition 6.10. Let ρ be effective. If $\{(\varphi_i, U_i)\}_{i\in I}$ and $\{(\psi_i, U_i)\}_{i\in I}$ are both (G,ρ)-atlases on a bundle (E,π,M) then $\{(\varphi_i, U_i)\}_{i\in I}$ is **equivalent**$_2$ to $\{(\psi_j, U_j)\}_{j\in J}$ if and only if there exist smooth functions $\lambda_i : U_i \to G$ such that

$$g'_{ji}(x) = \lambda_j(x)^{-1} g_{ji}(x) \lambda_i(x) \text{ for all } x \in U_i \cap U_j$$

We say that $\{(\varphi_i, U_i)\}_{i\in I}$ and $\{(\psi_i, U_i)\}_{i\in I}$ give rise to **cohomologous cocycles.**

Now we come to our first main assertion:

If $\{(\varphi_i, U_i)\}_{i\in I}$ and $\{(\psi_i, U_i)\}_{i\in I}$ are strictly equivalent then they are equivalent$_2$ **but the converse is not necessarily true**.

The fact that strict equivalence implies the weaker equivalence is an easy exercise. The fact the converse is not true becomes apparent already for trivial bundles and even bundles over a single point.

Example 6.11. Let us consider the case of a $Gl(n)$ structure on a bundle over a zero dimensional manifold. Thus we are essentially working with a single fiber X. The $Gl(n)$ structure on X is certainly supposed to be a vector space structure. Indeed, if we have a family of diffeomorphisms $\{\varphi_i : X \to \mathbb{R}^n\}_{i\in I}$ such that $\varphi_{\iota j} := \varphi_i \circ \varphi_j^{-1}$ is in $Gl(n)$ for every i, j then we have a $Gl(n)$-atlas and we get a well defined vector space structure on X.

(In fact all we need is one such chart.) Clearly two families of maps that are strictly equivalent will give the same vector space structure. Indeed, strict equivalence of $\{\psi_i\}$ and $\{\varphi_i\}$ would just mean that $\varphi_i \circ \psi_j^{-1}$ is always in $Gl(n)$ for all $i, j \in I$.
But now let $h : X \to X$ be any randomly chosen diffeomorphism[1] and let $\{\psi_i : X \to \mathbb{R}^n\}_{i \in I}$ be defined by $\psi_i := \varphi_i \circ h$. Then

$$\begin{aligned} g'_{ij} &:= \psi_i \circ \psi_j^{-1} = (\varphi_i \circ h) \circ (\varphi_j \circ h)^{-1} \\ &= \varphi_i \circ h \circ h^{-1} \circ \varphi_j^{-1} = \varphi_i \circ \varphi_j^{-1} = g_{ij} \end{aligned}$$

so not only does $\{\psi_i\}$ give rise to a cocycle cohomologous to that of $\{\varphi_i\}$, it gives rise to the very same cocycle! But we certainly don't expect $\{\psi_i\}$ and $\{\varphi_i\}$ to give the same vector space structure on X since h was chosen arbitrarily.

Notice that due to the arbitrary nature of h, the cohomologous equivalence is essentially vacuous in this case. This is as it should be since we are working over a point and the same would be true over a contractible base space. On the other hand, the strict equivalence defines the "geometrically" significant notion of vector space structure on X. If one tried to use the second weaker notion of equivalence then we wouldn't end up with a vector space structure–we wouldn't even have a consistent way of adding elements of X.

Remark 6.12. If we were to compare atlases $\{\varphi_i\}$ and $\{\psi_i = f \circ \varphi_i\}$ where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a linear isomorphism then we would have cohomologous, but in this case, also strictly equivalence atlases. If f is only assumed to be a diffemorphism then $g'_{ij} = f \circ \varphi_i \circ \varphi_j^{-1} \circ f^{-1} = f \circ g_{ij} \circ f^{-1}$ is not necessarily linear so the atlas $\{\psi_i = f \circ \varphi_i\}$ is not even a $Gl(n)$-atlas at all.

The example above, can be generalized to bundles over disks and then we see that using the wrong notion of equivalence would result in a false notion of $Gl(n)$-bundle structure that would leave us without a well defined module structure on the space of sections of the bundle–the fibers don't even become vector spaces.

It is often said that an $O(n)$ structure on vector bundle is just a choice of metric tensor. If the bundle has only one fiber again then this is just a question of when one can use these maps to define an inner product structure on a set X.

[1] We are in the smooth category but it is the only bijective nature of the map that will be used.

Example 6.13. Let us consider the case of a $O(n)$ structure on a bundle over zero dimensional manifold. Thus we are essentially working with a single fiber X. The $O(n)$-bundle structure on X is certainly supposed to be a an inner product space structure. Indeed, if we have a family of diffeomorphisms $\{\varphi_i : X \to \mathbb{R}^n\}_{i \in I}$ such that $\varphi_{ij} := \varphi_i \circ \varphi_j^{-1}$ is in $O(n)$ for every i, j then we have well defined vector space structure on X and an inner product can be defined by

$$\langle x, y \rangle := \varphi_i^{-1}(x) \cdot \varphi_i^{-1}(y) \text{ (dot product)}$$

and this is independent of i. Clearly, two such family of maps that are strictly equivalent, will give the same inner product space structure. Indeed, strict equivalence of $\{\psi_i\}$ and $\{\varphi_i\}$ would just mean that $\varphi_i \circ \psi_j^{-1}$ is always in $O(n)$.
But now let $h : X \to X$ be any randomly chosen diffeomorphism and let $\{\psi_i : X \to \mathbb{R}^n\}_{i \in I}$ be defined by $\psi_i := \varphi_i \circ h$. Then

$$\begin{aligned} g'_{ij} &:= \psi_i \circ \psi_j^{-1} = (\varphi_i \circ h) \circ (\varphi_j \circ h)^{-1} \\ &= \varphi_i \circ h \circ h^{-1} \circ \varphi_j^{-1} = \varphi_i \circ \varphi_j^{-1} = g_{ij}. \end{aligned}$$

So, not only does $\{\psi_i\}$ give rise to a cocycle cohomologous to that of $\{\varphi_i\}$, it gives rise to the very same cocycle! But we certainly don't expect $\{\psi_i\}$ and $\{\varphi_i\}$ to give the same inner product structure on X since h was chosen arbitrarily.

Example 6.14. Let D be the 2-disk $\{(x, y, 0) \in \mathbb{R}^3 : x^2 + y^2 < 1\}$. Give this manifold the flat metric it inherits from $\mathbb{R}^3$. This gives the tangent bundle TD an $O(2)$-bundle structure which is the same one that we get if we use the single vector bundle chart (φ, D) coming from the standard o.n. frame field on D. But now consider the projection of the upper hemisphere S^2_+ onto D. Use this to transfer an orthonormal frame from the curved space S^2_+ to D. This puts a different metric on D which now has curvature w.r.t the associated Levi-Civita connection. This is the same $O(2)$-bundle structure that we get if we use the single vector bundle chart (ψ, D) coming from this new frame field on D. But while the $O(2)$-atlas $\{(\varphi, D)\}$ is not strictly equivalent to $\{(\psi, D)\}$, it is equivalent in the cohomologous sense by an argument formally the same as the one in the last example. We certainly do not want to say that they define the very same $O(n)$-bundle structure. Indeed, they are two different metrics. Here we see that the $O(2)$-bundle structure itself can have a rigid geometric significance while the weaker cohomological notion retains a mainly topological significance.

The point is that one needs to first define a G-bundle structure using the notion of strict equivalence and then talk about when two such structures (possibly on the very same fibation) are equivalent.

6.4.2. Explanation. The notion of strict equivalence for G-atlases (with effective action) is used to define a G-bundle structure on a bundle. It is analogous to the equivalence of atlases for smooth manifolds where two such are equivalent if they are both subsets of the maximal atlas. The cohomological equivalence which we have called *equivalence*$_2$ is a special case of equivalence of two different G-bundles, where in this case, we just happen to have the same underlying fibration. It is similar to having two atlases which induce two *different* smooth structures on the same topological manifold which may nevertheless be diffeomorphic structures.

Let us put the weaker notion of equivalence in its proper context. In the general case, one considers two *different total spaces* and then the cohomologous cycle equivalence is the same as the following notion of G-bundle equivalence:

Definition 6.15. Let $\xi_1 = (E_1, \pi_1, M_1, F)$ be a (G, ρ)-bundle with its (G, ρ)-bundle structure determined by the strict equivalence class of the (G, ρ)-atlas $\{(\varphi_\alpha, U_\alpha)\}_{\alpha \in A}$. Let $\xi_2 = (E_2, \pi_2, M_2, F)$ be a (G, ρ)-bundle with its (G, ρ)-bundle structure determined by the strict equivalence class of the (G, ρ)-atlas $\{(\psi_\beta, V_\beta)\}_{\beta \in B}$. Then a bundle morphism $(\widehat{h}, h) : \xi_1 \to \xi_2$ is called a **(G, ρ)-bundle morphism** along $h : M_1 \to M_2$ if
i) $\widehat{h}$ carries each fiber of E_1 diffeomorphically onto the corresponding fiber of E_2;
ii) whenever $U_\alpha \cap h^{-1}(V_\beta)$ is not empty, there is a smooth map $h_{\alpha\beta} : U_\alpha \cap h^{-1}(V_\beta) \to G$ such that for each $p \in U_\alpha \cap h^{-1}(V_\beta)$ we have

$$\left(\Psi_\beta \circ \widehat{h} \circ \left(\Phi_\alpha|_{\pi_1^{-1}(p)} \right)^{-1} \right)(y) = \rho(h_{\alpha\beta}(p), y) \text{ for all } y \in F,$$

where as usual $\varphi_\alpha = (\pi_1, \Phi_\alpha)$ and $\psi_\beta = (\pi_2, \Psi_\beta)$. If $M_1 = M_2$ and $h = \mathrm{id}_M$ then we call $\widehat{h}$ a **(G, ρ)-bundle equivalence** over M. (In this case, $\widehat{h}$ is a diffeomorphism).

For this definition to be good it must be shown to be well defined. That is, one must show that condition (iii) is independent of the choice of representatives $\{(\varphi_\alpha, U_\alpha)\}_{\alpha \in A}$ and $\{(\varphi_i, V_i)_{i \in J}\}$ of the *strict* equivalence classes of atlases on the repsective bundles.

Now we see that the notion we have called "equivalence$_2$" is just the assertion that two different (G, ρ)-bundle structures on the very same fibration are (G, ρ)-bundle equivalent. We have also seen that *equivalence*$_2$ is not what should be used to define (G, ρ)-bundle structure in the first place.

Finally, we comment that one can approach (G,ρ)-bundle structures via principal bundles which can be defined without recourse to equivalence classes of bundle atlases. One must specify the right action since a single space can have more than one right action. In this case, we may have two different principal bundles (P,π,M,G,r_1) and (P,π,M,G,r_2) which differ only in the right action while the spaces are the same. Then (P,π,M,G,r_1) is equivalent to (P,π,M,G,r_2) if there is a fiber preserving map $P \to P$ that is also a G-map. This corresponds to *equivalence*$_2$ while *equivalence*$_1$ is now unnecessary since the principal G-bundle is now our starting point. In some sense the principal bundle just is the G-bundle structure for itself and all its associated vector bundles. (More on this to come). In fact, a G-cocyle is just what one needs to construct a principal G-bundle in the case where all one knows is information about overlaps for some atlas of fiber bundle local trivializations. Alternatively, in the case of vector bundles where ρ can be taken as a representation of G on the typical fiber V, then we can think of a vector bundle $E \to M$ as having a (G,ρ)-bundle structure if there is a principal G-bundle P and a bundle map $\widehat{f} : P \to F(E)$ such that $\widehat{f}(ug) = \widehat{f}(u) \circ \rho(g)$.

Remark 6.16. (G-bundle structures vs. "G-bundle structures") We have deliberately opted to use the term "G-bundle structure" rather that simply "G-structure" which could reasonably be taken to mean the same thing. Perhaps the reader is aware that there is a theory of G-structures *on* a smooth manifold (see). One may rightly ask whether a G-structure *on* M is nothing more than a G-bundle structure on the tangent bundle TM where G is a Lie subgroup of $GL(n)$ acting in the standard way. The answer is both yes and no. First, one could indeed say that a G-structure on M is a kind of G-bundle structure on TM even though the theory is usually fleshed out in terms of the frame bundle of M (defined below). However, the notion of equivalence of two G-structure on M is different than what we have given above. Roughly, G-structures on M are equivalent in this sense if there is a diffeomorphism ϕ such that $(T\phi,\phi)$ is a type II bundle isomorphism that is also a **G-bundle morphism** along ϕ.

6.4.3. Effectiveness of the Action. We wish to make some comments on the fact that we have chosen to assume the action ρ is effective. Let us return to the "coordinate bundle" viewpoint again and consider the task of developing good definitions without the effectivelness assumption. This is the assumption made in Steenrod's book on page 7. He keeps this assumption throughout and so for Steenrod, the structure group of the bundle $\bigwedge^k TM$ cannot be taken to be $Gl(n)$ since the latter acts ineffectively on $\bigwedge^k \mathbb{R}^n$ Without the assumption of an effective action, we must include the cocyle as part of the data. Indeed, the equations

$$\varphi_{\alpha\beta}(x)(y) = \rho(g_{\alpha\beta}(x), y) \text{ for } y \in F$$

used in the definition do not determine the $g_{\alpha\beta}$ and they may only satisfy the cocyle conditions modulo the kernel of the action! This is actually a serious problem.

Remark 6.17. On page 63 of the 3rd edition Husemoller's book we find the following sentence:
"Properties (T1) to (T2) [cocycle conditions] follow from the fact that $g_{i,j}$ is the only map satisfying the relation $h_j(b, y) = h_i(b, g_{i,j}(b)y)$."
The statement only makes sense if G acts effectively. But Husemoller has not made such an assumption prior to this nor after this statement. Much of what follows needs the effectiveness and yet many of the actions he uses are not effective (consider how the bundle $\bigwedge^k TM$ is an associated bundle for the frame bundle via an ineffective action $\bigwedge^k \rho_0$ where ρ_0 is the standard action of $Gl(n)$ on $\mathbb{R}^n$. Not that the way Steenrod would contruct a tensor bundle is by starting with a homomorphism $h : Gl(n) \to G$ where G acts effectively on $\bigwedge^k \mathbb{R}^n$ and G is the structure group (See section 6.3 on page 21 of Steenrod). Steenrod takes a $Gl(n)-$valued cocyle and composed with h to obtain a G-valued cocyle and only then does he use the bundle construction theorem to obtain a G-bundle. In the associated bundle construction as presented in Husemoller's book, one would use the ineffective representation $\rho : Gl(n) \to Gl(\bigwedge^k \mathbb{R}^n)$ and form the associated bundle

$$F(M) \times_\rho \bigwedge^k \mathbb{R}^n.$$

One is then tempted to say that $Gl(n)$ is the structure group or that ρ is the ineffective structure representation. To get more in line with Steenrod one should quotient both ρ and the frame bundle by $K = \ker \rho$ and only then produce an associated bundle. The quotiented frame bundle would by a $Gl(n)/K$ bundle where $Gl(n)/K$ isomorphic to G.

So, if we wish to proceed using atlases a la Steenrod, how shall we define (G, ρ)-structure in case ρ is ineffective? At the very least we must assume that each atlas is associated to a true G-cocyle and not just a cocyle modulo the kernel of the action (A G-principal bundle must be lurking) . Lessons from the theory of spin structures show that this is a nontrivial assumption. Second; how shall we formulate a notion of strict equivalence? If we simply take the union of two (G, ρ)-atlases then we may not end up with an atlas that is a associated to a larger inclusive G-cocyle (one again we only get the cocycle conditions up to the kernal of ρ). So we must just assume that the combined (G, ρ)-atlas comes with a cocycle. Can we obtain a maximal such atlas with a single maximal canonical cocyle?

The authors hope that the final version of this note will be able to sort this out by taking clues from the theory of principal and associated bundles. But, so far, it has proved rather slippery; there always seems to be a defeating loop hole that traces back to the ineffectiveness issue.

Chapter 8 Supplement

8.1. Comments and Errata

8.1.1. Comments. The exterior derivative is indeed special. It was proved by R. Palais that if $D : \Omega(M) \to \Omega(M)$ is a map defined for all smooth manifolds such that $D(\Omega^k(M)) \subset \Omega^\ell(M)$ and such that for every smooth map $f : M \to N$ the diagram below commutes,

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{f^*} & \Omega^k(M) \\ D\downarrow & & \downarrow D \\ \Omega^\ell(N) & \xrightarrow{f^*} & \Omega^\ell(M) \end{array}$$

then one of the following holds:

(1) $D = 0$,

(2) D is a multiple of the identity map (so $k = \ell$),

(3) D is a multiple of exterior differentiation or $k = \dim M$, $\ell = 0$ and D is a multiple of integration as defined and studied in the next chapter.

8.1.2. Errata. (Nothing yet)

8.2. 2-Forms over a surface in Euclidean Space Examples.

We just parametrize the surface as $\mathbf{r}(u, v) = (x(u, v), y(u, v), z(u, v))$ for some region of u, v. Let

$$\omega = F_1(x, y, z)dy \wedge dz + F_2(x, y, z)dz \wedge dx + F_3(x, y, z)dx \wedge dy$$

be our 2-form. Then $dy = \frac{\partial y}{\partial u}du + \frac{\partial y}{\partial v}dv$ and $dz = \frac{\partial z}{\partial u}du + \frac{\partial z}{\partial v}dv$ so

$$\begin{aligned}
dy \wedge dz &= \left(\frac{\partial y}{\partial u}du + \frac{\partial y}{\partial v}dv\right) \wedge \left(\frac{\partial z}{\partial u}du + \frac{\partial z}{\partial v}dv\right)\\
&= \frac{\partial y}{\partial u}\frac{\partial z}{\partial v}du \wedge dv + \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}dv \wedge du\\
&= \frac{\partial y}{\partial u}\frac{\partial z}{\partial v}du \wedge dv - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}dv \wedge du\\
&= \left(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u}\right) du \wedge dv\\
&= \begin{vmatrix} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{vmatrix} du \wedge dv = \frac{\partial (y,z)}{\partial (u,v)} du \wedge dv\\
&\text{(only involves } u \text{ and } v\text{)}
\end{aligned}$$

We have seen that before! Likewise we figure out $dz \wedge dx$ and $dx \wedge dy$ in terms of u and v. So we get

$$\begin{aligned}
\int \omega &= \int F_1(x,y,z)dy \wedge dz + F_2(x,y,z)dz \wedge dx + F_3(x,y,z)dx \wedge dy\\
&= \int\int \left(F_1 \frac{\partial (y,z)}{\partial (u,v)} + F_2 \frac{\partial (z,x)}{\partial (u,v)} + F_3 \frac{\partial (x,y)}{\partial (u,v)}\right) dudv\\
&= \int\int \langle F_1, F_2, F_3 \rangle \cdot (\mathbf{r}_u \times \mathbf{r}_v)\, du\, dv = \iint\limits_S \mathbf{F} \cdot \mathbf{n}\, dS
\end{aligned}$$

So $\int \omega = \iint\limits_S \mathbf{F} \cdot \mathbf{n}\, dS$ as long as we use the correspondence $\mathbf{F} \leftrightarrows \omega$ or in detail

$$F_1(x,y,z)\mathbf{i} + F_2(x,y,z)\mathbf{j} + F_3(x,y,z)\mathbf{k} \quad \leftrightarrows \quad F_1(x,y,z)\mathbf{dy} \wedge \mathbf{dz} + F_2(x,y,z)\mathbf{dz} \wedge \mathbf{dx} + F_3(x,y,z)\mathbf{dx} \wedge$$

Example 8.1. Let $\omega = xdy \wedge dz$ and consider the surface which is the graph of $z = 1 - x - y$ where x and y only vary over the region $0 \le x \le 1$ and $0 \le y \le 1 - x$. Then we can use x and y as our parameters. So

$$\begin{aligned}
\frac{\partial (y,z)}{\partial (x,y)} &= \left(\frac{\partial y}{\partial x}\frac{\partial z}{\partial y} - \frac{\partial y}{\partial y}\frac{\partial z}{\partial x}\right)\\
&= 0 - \frac{\partial z}{\partial x} = 1
\end{aligned}$$

$$\begin{aligned}
\int \omega &= \int xdy \wedge dz\\
&= \int_0^1 \int_0^1 \left(x \frac{\partial (y,z)}{\partial (x,y)}\right) dxdy\\
&= \int_0^1 \left(\int_0^{1-x} xdy\right) dx = \frac{1}{6}
\end{aligned}$$

Example 8.2. Let $\omega = xdy \wedge dz$ and consider the hemisphere S^+ of radius 1 centered ot the origin. This is parameterized by

$$\begin{aligned}\mathbf{r}(\phi,\theta) &= (x(\phi,\theta), y(\phi,\theta), z(\phi,\theta)) \\ &= (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)\end{aligned}$$

with $0 \leq \theta \leq 2\pi$ and $0 \leq \phi \leq \pi/2$. Then $dy = \frac{\partial y}{\partial \phi}d\phi + \frac{\partial y}{\partial \theta}d\theta$ and $dz = \frac{\partial z}{\partial \phi}d\phi + \frac{\partial z}{\partial \theta}d\theta$ or

$$\begin{aligned}dy &= \cos\phi\sin\theta\, d\phi + \sin\phi\cos\theta\, d\theta \\ dz &= -\sin\phi\, d\phi + 0\, d\theta\end{aligned}$$

so

$$\begin{aligned}ydy \wedge dz &= (\sin\phi\cos\theta)\left(\cos\phi\sin\theta\, d\phi + \sin\phi\cos\theta\, d\theta\right) \wedge (-\sin\phi\, d\phi) \\ &= -\sin^2\phi\sin\theta\cos^2\theta\, d\theta \wedge d\phi = \sin^2\phi\sin\theta\cos^2\theta\ d\phi \wedge d\theta\end{aligned}$$

Thus

$$\int_{S^+}\omega = \int_0^{\pi}\int_0^{\pi/2}\sin^2\phi\sin\theta\cos^2\theta\, d\phi d\theta = \frac{1}{6}\pi$$

8.3. Pseudo-Forms

Let M be a smooth n-manifold. Recall that the set of frames $\{e_1, ..., e_n\}$ in T_pM are divided into two equivalence classes called orientations. Equivalently, an orientation at p is an equivalence class of elements of $\wedge^n T_p^*M$. For each p there are exactly two such orientations and the set of all orientations $\mathrm{or}(M)$ at all points $p \in M$ form a set with a differentiable structure and the map π_{or} that takes both orientations at p to the point p is a two fold covering map. Rather than a covering space what we would like at this point is a rank one vector bundle. We now construct the needed bundle with is called the **flat orientation line bundle** and denoted $O_{\mathrm{flat}}(TM)$. Sections of $O_{\mathrm{flat}}(TM)$ are sometimes called pseudoscalar fields. For each $p \in M$ we construct a one dimensional vector space O_p. Let

$$S_p = \{(p, \mu_p, a) : a \in \mathbb{R}, \quad [\mu_p] \text{ an orientation on } T_pM\}$$

Thus μ_p is one of the two orientations of T_pM. The define an equivalence relation on S_p by declaring that $(p, [\mu_p], a)$ is equivalent to $(p, [-\mu_p], -a)$. Denote the equivalence class of $(p, [\mu_p], a)$ by $[p, [\mu_p], a]$. Then we let $O_p := S_p/\sim$. The set O_p has a one dimensional vector space structure such that

$$\begin{aligned}[p, [\mu_p], a] + [p, [\mu_p], b] &= [p, [\mu_p], a + b] \\ r[p, [\mu_p], a] &= [p, [\mu_p], ra]\end{aligned}$$

for $a, b, r \in \mathbb{R}$. Now let

$$O_{\mathrm{flat}}(TM) := \bigcup_{p \in M} O_p\,.$$

We have the obvious projection $\pi : O_{\text{flat}}(TM) \to M$ given by $[p, [\mu_p], a] \mapsto p$. An atlas $\{(U_\alpha, \mathtt{x}_\alpha)\}$ for M induces a vector bundle atlas for $O_{\text{flat}}(TM)$ as follows: For a chart $(U_\alpha, \mathtt{x}_\alpha)$ define

$$\psi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}$$

by $\psi_\alpha([p, [\mu_p], a]) = (p, \varepsilon a)$ where $\varepsilon = 1$ or -1 according to whether $T_p\psi_\alpha : (T_pM, [\mu_p]) \to \mathbb{R}^n$ is orientation preserving or not. Associated with such a chart we have the local section

$$o_{\mathtt{x}_\alpha} : p \mapsto [p, [\mu_p(\mathtt{x}_\alpha)], a]$$

where $\mu_p(\mathtt{x}_\alpha) = dx^1 \wedge \cdots \wedge dx^n$. In fact, if U is any orientable open set in M with orientation given by $[\mu]$ where $\mu \in \Omega^n(U)$ then we can regard $[\mu]$ as a local section of $O_{\text{flat}}(TM)$ as follows:

$$[\mu] : p \mapsto [p, [\mu(p)], 1]$$

With this this VB-atals the transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \to Gl(1, \mathbb{R})$ defined by

$$g_{\alpha\beta}(p) := \frac{\det\left(T\mathtt{x}_\alpha \circ T\mathtt{x}_\beta^{-1}\right)}{\left|\det\left(T\mathtt{x}_\alpha \circ T\mathtt{x}_\beta^{-1}\right)\right|} = \pm 1$$

and so we see that we have a structure group $O(1)$. Thus the transition functions are locally constant (which is why we call this bundle flat).

A pseudo-k-form (or twisted k-form) is a cross section of $O_{\text{flat}}(TM) \otimes \wedge^k T^*M$. The set of all twisted k-forms will be denoted by $\Omega_o^k(M)$. Sections of $\Omega_o^k(M)$ can be written locally as sums of elements of the form $o \otimes \theta$ where o is an orientation of U regarded as a local section of $O_{\text{flat}}(TM)$ as described above. Now we can extend the exterior product to maps

$$\wedge : \Omega_o^k(M) \times \Omega^l(M) \to \Omega_o^{k+l}(M)$$

by the rule $(o_1 \otimes \theta_1) \wedge \theta_2 = o_1 \otimes \theta_1 \wedge \theta_2$ with a similar and obvious map $\wedge : \Omega^k(M) \times \Omega_o^l(M) \to \Omega_o^{k+l}(M)$. Similarly, we have a map $\wedge : \Omega_o^k(M) \times \Omega_o^l(M) \to \Omega^{k+l}(M)$ given by

$$(o_1 \otimes \theta_1) \wedge (o_2 \otimes \theta_2) = (o_1 o_2)\, \theta_1 \wedge \theta_2$$

and where $(o_1 o_2)$ is equal to $+1$ wherever $o_1 = o_2$ and -1 otherwise. Now we can extend the exterior algebra to $\sum_{k,l=0}^n \left(\Omega^k(M) \oplus \Omega_o^l(M)\right)$. If $\omega \in \Omega_o^k(M)$ then with respect to the chart $(U_\alpha, \mathtt{x}_\alpha)$, ω has the local expression

$$\omega = o_{\mathtt{x}_\alpha} \otimes a_{\vec{I}}^\alpha dx_\alpha^{\vec{I}}$$

and if $\omega = o_{\mathtt{x}_\beta} \otimes a^\beta_{\vec{J}} dx^{\vec{J}}_\beta$ for some other chart $(U_\beta, \mathtt{x}_\beta)$ then

$$a^\alpha_{\vec{I}} = \frac{\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)}{\left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right|} \frac{\partial x^{\vec{J}}_\alpha}{\partial x^{\vec{I}}_\beta} a^\beta_{\vec{J}}$$

In particular if ω is a pseudo n-form (a volume pseudo-form) where $n = \dim M$, then $\vec{I} = (1, 2, ..., n) = \vec{J}$ and $\frac{\partial x^{\vec{J}}_\alpha}{\partial x^{\vec{I}}_\beta} = \det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)$ and so the rule becomes

$$a^\alpha_{12..n} = \frac{\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)}{\left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right|} \det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right) a^\beta_{12..n}$$

or

$$a^\alpha_{12..n} = \left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right| a^\beta_{12..n}.$$

There is another way to think about pseudo forms that has the advantage of having a clearer global description. The group $\mathbb{Z}/2\mathbb{Z} = \{1, -1\}$ acts as deck transformations on $\mathrm{or}(M)$ so that -1 sends each orientation to its opposite. Denote the action of $g \in \mathbb{Z}/2\mathbb{Z}$ by $l_g : \mathrm{or}(M) \to \mathrm{or}(M)$. Now we think of a pseudo k-form as being nothing more than a k-form η on the manifold $\mathrm{or}(M)$ with the property that $l^*_{-1}\eta = -\eta$. Now we would like to be able to integrate a k-form over a map $h : N \to M$ where N is a k-manifold. By definition h is orientable if there is a lift $\widetilde{h} : \mathrm{or}(N) \to \mathrm{or}(M)$

$$\begin{array}{ccc} \mathrm{or}(N) & \overset{\widetilde{h}}{\to} & \mathrm{or}(M) \\ \downarrow \pi_{Or_N} & & \downarrow \pi_{Or_M} \\ N & \overset{h}{\to} & M \end{array}$$

We will say that $\widetilde{h}$ is said to orient the map. In this case we define the integral of a pseudo k-form η over h to be

$$\int_h \eta := \frac{1}{2} \int_{\mathrm{Or}(N)} \widetilde{h}^* \eta$$

Now there is clearly another lift $\widetilde{h_-}$ which sends each $\widetilde{n} \in \mathrm{or}(N)$ to the opposite orientation of $\widetilde{h}(\widetilde{n})$. This is nothing more that saying $\widetilde{h_-} = l_{-1} \circ \widetilde{h} = \widetilde{h} \circ l_{-1}$.

Exercise 8.3. Assume M is connected. Show that there are at most two such lifts $\widetilde{h}$.

Now

$$\int_{\mathrm{or}(N)} \widetilde{h}^*_- \eta = \int_{\mathrm{or}(N)} \widetilde{h}^* l^*_g \eta = \int_{\mathrm{or}(N)} \widetilde{h}^* \eta$$

and so the definition of $\int_h \eta$ is independent of the lift $\widetilde{h}$.

If $S \subset M$ is a regular k-submanifold and if the inclusion map $\iota_S : S \hookrightarrow M$ map is orientable than we say that S has a transverse orientation in M. In this case we define the integral of a pseudo k-form η over S to be

$$\int_S \eta := \int_{\iota_S} \eta = \frac{1}{2} \int_{\mathrm{or}(N)} \widetilde{\iota_S}^* \eta$$

Exercise 8.4. Show that the identity map $\mathrm{id}_M : M \to M$ is orientable in two ways:

$$\begin{array}{ccc} \mathrm{or}(M) & \overset{\widetilde{\mathrm{id}}}{\to} & \mathrm{or}(M) \\ \downarrow \wp_{Or_M} & & \downarrow \wp_{Or_M} \\ M & \overset{\mathrm{id}}{\to} & M \end{array}$$

where $\widetilde{\mathrm{id}} = \pm \,\mathrm{id}_{Or(M)}$.

Now finally, if ω is a pseudo-n-form on M then by definition

$$\int_M \omega := \frac{1}{2} \int_{\mathrm{or}(M)} \widetilde{\mathrm{id}}^* \omega = \frac{1}{2} \int_{\mathrm{or}(M)} \omega$$

If $(U, \mathtt{x})$ is a chart then the map $\sigma_{\mathtt{x}} : p \mapsto [dx^1 \wedge \cdots \wedge dx^n]$ is a local cross section of the covering $\wp_{Or} : \mathrm{or}(M) \to M$ meaning that the following diagram commutes

$$\begin{array}{ccc} & & \mathrm{or}(M) \\ & \overset{\sigma_{\mathtt{x}}}{\nearrow} & \downarrow \wp_{Or} \\ U & \hookrightarrow & M \end{array}$$

and we can define the integral of ω locally using a partition of unity $\{\rho_\alpha\}$ subordinate to a cover $\{U_\alpha, \rho_\alpha\}$:

$$\int_M \omega := \sum_\alpha \int_{U_\alpha} \rho_\alpha \sigma_{\mathtt{x}_a}^* \omega.$$

Now suppose we have a vector field $X \in \mathfrak{X}(M)$. Since $\wp_{Or} : \mathrm{or}(M) \to M$ is a surjective local diffeomorphism there is a vector field $\widetilde{X} \in \mathfrak{X}(\widetilde{M})$ such that $T\wp \cdot \widetilde{X}_{\widetilde{p}} = X_p$ (where $\wp_{\mathrm{Or}}(\widetilde{p}) = p$) for all p. Similarly, if μ is a volume form on M then there is a volume pseudo-form $\widetilde{\mu}$ on M, i.e. a $\mathbb{Z}/2\mathbb{Z}$ anti-invariant $n-$form $\widetilde{\mu}$ on $\mathrm{or}(M)$ such $\widetilde{\mu} = \wp_{\mathrm{or}}^* \mu$. In this case, it is easy to show that the divergence $\operatorname{div} \widetilde{X}$ of $\widetilde{X}$ with respect to $\widetilde{\mu}$ is the lift of $\operatorname{div} X$ (with respect to μ). Thus if M is not orientable and so has no volume form we may still define $\operatorname{div} X$ (with respect to the pseudo-volume form $\widetilde{\mu}$) to be the unique vector field on M which is $\wp_{\mathrm{or}}$-related to $\operatorname{div} \widetilde{X}$ (with respect to volume form $\widetilde{\mu}$ on $\mathrm{or}(M)$).

Chapter 11
Supplement

11.1. Comments and Errata

11.1.1. Comments. (Nothing yet)

11.1.2. Errata. (Nothing yet)

11.2. Singular Distributions

Lemma 11.1. *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 11.2. Let $\mathfrak{X}_{loc}(M)$ denote the set of all sections of the presheaf $\mathfrak{X}_M$. That is

$$\mathfrak{X}_{loc}(M) := \bigcup_{\text{open } U \subset M} \mathfrak{X}_M(U).$$

Also, for a distribution Δ let $\mathfrak{X}_\Delta(M)$ denote the subset of $\mathfrak{X}_{loc}(M)$ consisting of local fields X with the property that $X(x) \in \Delta_x$ for every x in the domain of X.

Definition 11.3. 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 11.4. 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 11.5. 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 11.6. 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 11.7. Show that $\mathcal{S}(\mathcal{X})$ is the smallest stable family of local vector fields containing $\mathcal{X}$.

Definition 11.8. 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 11.9. 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 11.10. Convince yourself that $\mathrm{aut}_{loc}(\Delta)$ is stable.

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

Theorem 11.11. *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

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

Now if Y is in $\mathfrak{X}_\Delta$ then at an arbitrary x we have $Y(x) \in \Delta_x$ and so the above shows that $((\mathrm{Fl}_t^X)^*Y)(x) \in \Delta$ so $(\mathrm{Fl}_t^X)^*Y)$ is in $\mathfrak{X}_\Delta$. We conclude that $\mathfrak{X}_\Delta$ is stable and have shown that $(1) \Rightarrow (2)$.

Next, if (2) hold then $\mathfrak{X}_\Delta \subset \mathrm{aut}_{loc}(\Delta)$ and so we have (3).

If (3) holds then we let $\mathcal{X} := \mathrm{aut}_{loc}(\Delta) \cap \mathfrak{X}_\Delta$. Then for $Y, Y \in \mathcal{X}$ we have $(\mathrm{Fl}_t^X)^* Y \in \mathfrak{X}_\Delta$ and so $\mathcal{X} \subset \mathcal{S}(\mathcal{X}) \subset \mathfrak{X}_\Delta$. from this we see that since $\mathcal{X}$ and $\mathfrak{X}_\Delta$ both span Δ so does $\mathcal{S}(\mathcal{X})$.

Finally, we show that (4) implies (1). Let $x \in M$. Since $\mathcal{S}(\mathcal{X})$ spans the distribution and is also stable by construction we have

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

for all fields X from $\mathcal{S}(\mathcal{X})$. Let the dimension Δ_x be k and choose fields $X_1, ..., X_k \in \mathcal{S}(\mathcal{X})$ such that $X_1(x), ..., X_k(x)$ is a basis for Δ_x. Define a map $f :: \mathbb{R}^k \to M$ by

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

which is defined (and smooth) near $0 \in \mathbb{R}^k$. As in lemma 11.1 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 $T_x S$ is spanned by $\frac{\partial f}{\partial t^j}(0)$ for $j = 1, 2, ..., k$ and

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

But $\mathcal{S}(\mathcal{X})$ is stable so each $\frac{\partial f}{\partial t^j}(0)$ lies in $\Delta_{f(t)}$. From the construction of f and remembering **??** we see that $\mathrm{span}\{\frac{\partial f}{\partial t^j}(0)\} = T_{f(t)}S = \Delta_{f(t)}$ and we are done. □

Chapter 12
Supplement

12.1. Comments and Errata

12.1.1. Comments. (Nothing yet)

12.1.2. Errata. (Nothing yet)

12.2. More on Actions

Let $\alpha : P \times G \to P$ be a right Lie group action of the Lie group G on a smooth manifold P. We will write $x \cdot g$ or xg in place of $\alpha(x, g)$ whenever convenient. For each $A \in \mathfrak{g} = T_eG$, let $\lambda(A)$ denote the left invariant vector field such that $\lambda(A)(e) = A$. Recall that $\lambda(A)(g) = T_eL_g(A)$ for each $g \in G$ and where $L_g : G \to G$ is left translation by g. We have

$$\lambda(A)(g) = \left.\frac{d}{dt}\right|_{t=0} g \exp(tA)$$

Let $\ell_x : G \to P$ be defined by $\ell_x(g) := xg$. Given $A \in \mathfrak{g}$ and $x \in P$, a vector $\upsilon_x(A) \in T_xP$ is defined by

$$\upsilon_x(A) = T_e\ell_x(A).$$

Then we have

$$\upsilon_x(A) := \left.\frac{d}{dt}\right|_{t=0} x \cdot \exp(tA)$$

for $x \in P$. This shows that the flow given by $\varphi(t, x) := x \cdot \exp(tA)$ generates a smooth vector field $\upsilon(A) : x \mapsto \upsilon(A)(x) := \upsilon_x(A)$. The map $\upsilon : \mathfrak{g} \to \mathfrak{X}(P)$

given by $A \mapsto \upsilon(A)$ is linear since

$$\begin{aligned}\upsilon(aA+bB)(x) &= T_e\ell_x(aA+bB) = aT_e\ell_x(A) + bT_e\ell_x(B)\\ &= a\upsilon(A)(x) + b\upsilon(B)(x)\end{aligned}$$

for $a,b \in \mathbb{R}$ and $A,B \in \mathfrak{g}$ and all x. In the special case that $P = G$ and where the action is given by group multiplication, this construction gives $\lambda(A)$ defined above for each $A \in \mathfrak{g}$.

Let us identify $T(P \times G)$ with $TP \times TG$ as usual. Then $(0_x, \lambda(A)_g) \in T_xP \times T_gG$ is represented by the curve

$$t \mapsto (x, g\exp(tA))$$

and so

$$\alpha(x, g\exp(tA)) = x \cdot g\exp(tA)$$

and

$$T\alpha(0_x, \lambda(A)_g) = \left.\frac{d}{dt}\right|_{t=0} x \cdot g\exp(tA) = \upsilon_{x\cdot g}(A).$$

From this we see that the vector field $(0, \lambda(A)) \in \mathfrak{X}(P \times G)$ is α-related to $\upsilon(A)$ for any A. Thus

$$[(0,\lambda(A)),(0,\lambda(B))] \text{ is } \alpha\text{-related to } [\upsilon(A),\upsilon(B)].$$

But $[(0,\lambda(A)),(0,\lambda(B))] = (0,[\lambda(A),\lambda(B)]) = (0,\lambda([A,B]))$ and so $(0,\lambda([A,B]))$ is α-related to $[\upsilon(A),\upsilon(B)]$. In other words,

$$[\upsilon(A),\upsilon(B)]_{x\cdot g} = T\alpha(0_x, \lambda([A,B])_g) = \upsilon([A,B])_{x\cdot g}$$

or

$$\upsilon([A,B]) = [\upsilon(A),\upsilon(B)].$$

Thus υ gives a Lie algebra homomorphism of $\mathfrak{g}$ onto its image $\upsilon(\mathfrak{g}) \subset \mathfrak{X}(P)$.

Lemma 12.1. *If the action $\alpha : P \times G \to P$ is effective then the map $\mathfrak{g} \to T_xP$ given by $\upsilon_x : A \longmapsto \upsilon_x(A)$ is an injective linear map.*

Proof. That υ_x is linear is clear since it is the composition $A \mapsto \upsilon(A) \mapsto \upsilon_x.(A)$. Now suppose that $\upsilon_x(A) = 0$. Then the flow of $\upsilon(A)$ is stationary at x and we must have $x \cdot \exp(tA) = x$ for all t. But then $\exp(tA) = e$ for all t since the action is assumed effective. It follows $A = 0$ since exp is injective on a sufficiently small neighborhood of 0. □

Lemma 12.2. *If the action $\alpha : P \times G \to P$ is effective,the map $\upsilon : \mathfrak{g} \to \upsilon(\mathfrak{g}) \subset \mathfrak{X}(P)$ is a Lie algebra isomorphism onto its image $\upsilon(\mathfrak{g})$.*

Proof. We already know that the map is a Lie algebra homomorphism and it is onto. Now suppose that $\upsilon(A)$ is identically zero. Then since $A \longmapsto \upsilon_x(A)$ is injective for any $x \in P$, we see that $A = 0$. The result follows.

Lemma 12.3. *If the action is free and $A \neq 0$ then $\upsilon(A)$ is nowhere vanishing.*

□

Proof. We leave this as an exercise. □

Now let μ be the left Maurer-Cartan form[1] which is the $\mathfrak{g}$-valued 1-form on G defined by

$$\mu(g)(X_g) = T_g\left(L_{g^{-1}}\right)(X_g),$$

Theorem 12.4. *Let $\alpha : P \times G \to P$ be a Lie groups action as above. Then $T\alpha : TP \times TG \to TP$ is given by*

$$T\alpha \cdot (X_p, Y_g) = Tr_g \cdot X_p + \upsilon_{pg}(\mu(Y_g))$$

for $X_p \in T_pP$, $Y_g \in T_gG$ and where μ is the left Maurer-Cartan form of G.

12.3. Connections on a Principal bundle.

We recall our definition of a smooth principal G-bundle:

Definition 12.5. Let $\wp : P \to M$ be a smooth fiber bundle with typical fiber a Lie group G. The bundle $(P, \wp, M, G)$ is called a **principal G-bundle** if there is a smooth free right action of G on P such that

(i) The action preserves fibers; $\wp(ug) = \wp(u)$ for all $u \in P$ and $g \in G$;

(ii) For each $p \in M$, there exists a bundle chart (U, ϕ) with $p \in U$ and such that if $\phi = (\wp, \Phi)$, then

$$\Phi(ug) = \Phi(u)g$$

for all $u \in \wp^{-1}(U)$ and $g \in G$. If the group G is understood, then we may refer to $(P, \wp, M, G)$ simply as a **principal bundle**. We call these charts **principal G-bundle charts.**

Let $\wp : P \to M$ be a smooth principal G-bundle. Thus we a have a free and effective right action of G on P whose orbits are exactly the fibers. The vertical subspace $\mathcal{V}_u$ at $u \in P$ is is the subspace of T_uP consisting of vectors tangent to the fiber that contains u. Thus $\mathcal{V}_u = \ker(T_u\wp)$ and $\mathcal{V} = \bigcup_{u \in P} \mathcal{V}_u$ is a subbundle of TP which is an integrable distribution on P called the vertical distribution or vertical bundle. Its leaves are the fibers of the principal bundle. Let $\mathfrak{X}(\mathcal{V})$ denote the subspace of $\mathfrak{X}(P)$ consisting of vector fields taking valued in $\mathcal{V}$, that is, vector fields tangent to the fibers of P. From the last section we know that the right action of G on P gives a map $\upsilon : \mathfrak{g} \to \mathfrak{X}(P)$ that is a Lie algebra homomorphsim. Its image $\upsilon(\mathfrak{g})$ is Lie algebra and is clearly contained in $\mathfrak{X}(\mathcal{V})$.

[1] Elsewhere we denoted Maurer-Cartan form by ω_G.

From the lemmas of the previous section we know that $\upsilon(A)$ is nowhere vanishing whenever $A \neq 0$ and the map from $\mathfrak{g}$ to $\mathcal{V}_u$ given by $A \longmapsto \upsilon(A)_u$ is one to one. But since $\dim(\mathfrak{g}) = \dim \mathcal{V}_u$ this map is actually a linear isomorphism of $\mathfrak{g}$ onto $\mathcal{V}_u$. The vertical $\upsilon(A)$ vector field is called the **fundamental vertical vector field** corresponding to A. From the previous section we know that the map $\upsilon : \mathfrak{g} \to \upsilon(\mathfrak{g})$ is a Lie algebra isomorphism.

Definition 12.6. For each $g \in G$, let the maps $r_g : P \to P$, $R_g : G \to G$, be defined by

$$\begin{aligned} r_g(u) &:= u \cdot g \\ R_g(a) &= ag \end{aligned}$$

We have already met the left translation L_g. Each of the maps r_g, R_g, and L_g are diffemorphisms. Recall that for each $g \in G$ we have the maps $C_g : G \to G$ and $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ given by

$$\begin{aligned} C_g(x) &= gxg^{-1} \\ \mathrm{Ad}_g &:= T_e C_g \end{aligned}$$

Note that $C_g = R_{g^{-1}} \circ L_g = L_g \circ R_{g^{-1}}$ and since $\frac{d}{dt}\big|_{t=0} \exp tA = A$ we have

$$\mathrm{Ad}_g A = \left.\frac{d}{dt}\right|_{t=0} g(\exp tA)\, g^{-1}.$$

Proposition 12.7. Let $\wp : P \to M$ be a smooth principal G-bundle and let $\upsilon : \mathfrak{g} \to \upsilon(\mathfrak{g})$ be as above. For each $A \in \mathfrak{g}$ we have

$$T_u r_g(\upsilon(A)_u) = \upsilon(\mathrm{Ad}_{g^{-1}}(A))_{u \cdot g}$$

Proof.

$$\begin{aligned} T_u r_g(\upsilon(A)_u) = T_e C_g(\upsilon(A)_u) &= \left.\frac{d}{dt}\right|_{t=0} r_g(u(\exp tA)) \\ &= \left.\frac{d}{dt}\right|_{t=0} u(\exp tA)\, g = \left.\frac{d}{dt}\right|_{t=0} ugg^{-1}(\exp tA)\, g \\ &= \left.\frac{d}{dt}\right|_{t=0} ug \exp\left(t\, \mathrm{Ad}_{g^{-1}}(A)\right) =: \upsilon(\mathrm{Ad}_{g^{-1}}(A))_{u \cdot g} \end{aligned}$$

□

The result of the previous proposition can also be written

$$Tr_g \circ \upsilon(A) = \upsilon(\mathrm{Ad}_{g^{-1}}(A)) \circ r_g^{-1}$$

or

$$(r_g)_* \upsilon(A) = \upsilon(\mathrm{Ad}_{g^{-1}}(A))$$

where $(r_g)_* : \mathfrak{X}(P) \to \mathfrak{X}(P)$ is the push-forward map.

Basically, a connection is a choice of subbundle of TP, that is complementary to the vertical bundle. In other words, a distribution complementary to $\mathcal{V}$. However, we would also like this distribution to respect the action of the Lie group.

Definition 12.8. A connection on a principal G-bundle P is a distribution $\mathcal{H} \subset TP$ such that
(i) $T_uP = \mathcal{H}_u \oplus \mathcal{V}_u$ for all $u \in P$;
(ii) $Tr_g(\mathcal{H}_u) = \mathcal{H}_{ug}$ for all $u \in P$ and $g \in G$.

The distribution $\mathcal{H}$ is also refered to as the horizontal distribution.

Definition 12.9. A tangent vector $Y_u \in T_uP$ called a **horizontal vector** at u if $Y_u \in \mathcal{H}_u$ and is called a **vertical vector** at u if $Y_u \in \mathcal{V}_u$.

For each u, we have two projections $\mathrm{p_h} : T_uP \to \mathcal{H}_u$ and $\mathrm{p_v} : T_uP \to \mathcal{V}_u$. These combine to give vectro bundle morphisms $\mathrm{p_h} : TP \to \mathcal{H}$ and $\mathrm{p_v} : TP \to \mathcal{V}$. For for any tangent vector $Y_u \in T_uP$ we have a unique decomposition into horizontal and vertical vectors

$$Y_u = \mathrm{p_h}(Y_u) + \mathrm{p_v}(Y_u).$$

If, for each $u \in O$, we apply the projection $T_uP \to \mathcal{V}_u$ and then the inverse of isomorphsim $\mathfrak{g} \to \mathcal{V}_u$ given by $A \longmapsto \upsilon(A)_u$, then we obtain linear map $T_uP \to \mathfrak{g}$. In this way we obtain a $\mathfrak{g}$-vlaued 1-form on P that we denote by ω. It is characterized by the following:

$$\omega(\upsilon(A)_u) = A \text{ for all } u.$$

This form is called the connection form. It is clear that $\omega \circ \mathrm{p_h} = 0$

Lemma 12.10. *The connection form satisfies*

$$r_g^*\omega = \mathrm{Ad}_{g^{-1}}\,\omega.$$

In other words, $\omega(Tr_g(Y_u)) = \mathrm{Ad}_{g^{-1}}\,\omega(Y_u)$ *for all* $Y_u \in TP$.

Proof. Notice that since $\mathrm{p_h}(Y_u) \in \mathcal{H}_u$ we have $Tr_g(\mathrm{p_h}(Y_u)) \in \mathcal{H}_{ug}$. Now $\mathrm{p_v}(Y_u) = \upsilon(A)_u$ for a unique $A \in \mathfrak{g}$ and so

$$\omega(Y_u) = \omega(\mathrm{p_h}(Y_u) + \mathrm{p_v}(Y_u)) = \omega(\mathrm{p_v}(Y_u)) = \omega(\upsilon(A)_u) = A.$$

Thus,

$$\begin{aligned}
\omega(Tr_g(Y_u)) &= \omega(Tr_g(\mathrm{p_h}(Y_u) + \mathrm{p_v}(Y_u))) = \omega(Tr_g(\mathrm{p_h}(Y_u) + \upsilon(A)_u)) \\
&= \omega(Tr_g(\mathrm{p_h}(Y_u)) + Tr_g(\upsilon(A)_u)) = \omega(Tr_g(\upsilon(A)_u)) \\
&= \omega\left(\upsilon(\mathrm{Ad}_{g^{-1}}(A))\right) = \mathrm{Ad}_{g^{-1}}(A) = \mathrm{Ad}_{g^{-1}}(\omega(Y_u)).
\end{aligned}$$

$\square$

Conversly we can start with an appropriate $\mathfrak{g}$-valued 1-form on P on and recover a connection $\mathcal{H}$.

Theorem 12.11. *Let $\wp : P \to M$ be a principal G-bundle as above. Let ω be a $\mathfrak{g}$-valued 1-form on P such that*
(i) $\omega(\upsilon(A)_u) = A$ for all u;
(ii) $r_g^*\omega = \mathrm{Ad}_{g^{-1}}\,\omega$.
If $\mathcal{H}$ defined by $\mathcal{H}_u := \ker(\omega_u)$ for all $u \in P$, then $\mathcal{H}$ is a connection on $\wp : P \to M$.

Proof. If $Y_u \in T_uP$ then $Y_u - \upsilon(\omega_u(Y_u))$ is in $\mathcal{H}_u = \ker(\omega_u)$. Indeed, $\omega\left(Y_u - \upsilon(\omega_u(Y_u))\right) = \omega(Y_u) - \omega(Y_u) = 0$. Thus we may write $Y_u = Y_u - \upsilon(\omega_u(Y_u)) + \upsilon(\omega_u(Y_u))$. We conclude that $T_uP = \mathcal{H}_u + \mathcal{V}_u$. But if $Y_u \in \mathcal{H}_u \cap \mathcal{V}_uM$, then $Y_u = \upsilon(A)$ for a unique A and also $0 = \omega(Y_u) = \omega(\upsilon(A)) = A$ so that $Y_u = \upsilon(0) = 0$. Thus we have $T_uP = \mathcal{H}_u \oplus \mathcal{V}_u$.
Now suppose that $Y_u \in \mathcal{H}_u$. Then $\omega(Tr_g(Y_u)) = r_g^*\omega(Y_u) = \mathrm{Ad}_{g^{-1}}\,\omega(Y_u) = \mathrm{Ad}_{g^{-1}}(0) = 0$. So $Tr_g\left(\mathcal{H}_u\right) \subset \mathcal{H}_{ug}$. But since $\dim \mathcal{H}_u = \dim \mathcal{H}_{ug}$ we conclude that $Tr_g\left(\mathcal{H}_u\right) = \mathcal{H}_{ug}$. □

Definition 12.12. We shall call a $\mathfrak{g}$-valued 1-form that satisifies the hypotheses Theorem 12.11 a **connection form**.

We recall that if (ϕ, U) be a principal bundle chart with $\phi = (\wp, \Phi)$ then $s_\phi : U \to \wp^{-1}(U)$ is the associated local section defined by $s_\phi : p \mapsto \phi^{-1}(p, e)$. Conversely, if $s : U \to \wp^{-1}(U)$ is a local section then $\phi^{-1}(p, g) = ps(g)$ defines a principal bundle chart with (ϕ, U). Let ω be the connection form of a connection P. A local section is sometimes called a **local gauge**, especially in the physics literature. Given a local section s defined on U, we obtain the $\mathfrak{g}$-valued 1-form $\mathcal{A} = s^*\omega$ defined on U. The form $\mathcal{A}$ is called a **local connection form** or a **local gauge potential**.

Proposition 12.13. Suppose that s_1 and s_2 are local sections of P defined on U_1 and U_2 respectively. Let $\mathcal{A}_1 = s_1^*\omega$ and $\mathcal{A}_2 = s_2^*\omega$ be the corresponding local gauge potentials. Then on $U_1 \cap U_2$ we have $s_2 = s_1 g$ for some smooth function $g : U_1 \cap U_2 \to G$. We have

$$\mathcal{A}_2 = \mathrm{Ad}_{g^{-1}}\,\mathcal{A}_1 + g^*\mu \text{ on } U_1 \cap U_2.$$

Proof. $\mathcal{A}_2(X_p) = s_2^*\omega(X_p) = \omega(Ts_2 \cdot X_p)$. Now let $X_p = \dot\gamma(0)$ for a curve γ. Then if $\alpha : P \times G \to P$ denotes the action, then using Theorem 12.4, we have

$$\begin{aligned}
Ts_2 \cdot X_p &= \left.\frac{d}{dt}\right|_{t=0} s_2(\gamma(t)) = \left.\frac{d}{dt}\right|_{t=0} s_1(\gamma(t))g(\gamma(t)) \\
&= \left.\frac{d}{dt}\right|_{t=0} \alpha\left(s_1(\gamma(t))\,,\, g(\gamma(t))\right) = T\alpha \cdot (Ts_1 \cdot X_p, Tg \cdot X_p) \\
&= Tr_g \cdot (Ts_1 \cdot X_p) + \upsilon_{s_1(p)g(p)}(\mu(Tg \cdot X_p)) \\
&= Tr_g \cdot (Ts_1 \cdot X_p) + \upsilon_{s_1(p)g(p)}\left(g^*\mu(X_p)\right).
\end{aligned}$$

Now apply ω to both sides to get

$$\begin{aligned}
\mathcal{A}_2(X_p) &= \omega(Tr_g \cdot (Ts_1 \cdot X_p) + g^*\mu(X_p)) \\
&= r_g^*\omega(Ts_1 \cdot X_p) + g^*\mu(X_p) \\
&= \mathrm{Ad}_{g^{-1}}\,\omega(Ts_1 \cdot X_p) + g^*\mu(X_p) \\
&= \mathrm{Ad}_{g^{-1}}\,\mathcal{A}_1(X_p) + g^*\mu(X_p).
\end{aligned}$$

□

Theorem 12.14. *Let ω be a connection form on a principal G-bundle $\wp : P \to M$. Let $\{(U_\alpha, \phi_\alpha)\}_{\alpha\in A}$ be a principal G-bundle atlas with associated local sections $s_\alpha : U_\alpha \to P$. Let $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ the assocated maps defined by $s_\beta = s_\alpha g_{\alpha\beta}$. Then for $\mathcal{A}_\alpha := s_\alpha^*\omega$, $\alpha \in A$ we have*

$$\mathcal{A}_\beta = \mathrm{Ad}_{g_{\alpha\beta}^{-1}}\,\mathcal{A}_\alpha + g_{\alpha\beta}^*\mu$$

on $U_\alpha \cap U_\beta$. Conversely, if to every (U_α, ϕ_α) in the principal bundle atlas with associated sections s_α, there is assigned a $\mathfrak{g}$-valued 1-form $\mathcal{A}_\alpha$ such that $\mathcal{A}_\beta = \mathrm{Ad}_{g_{\alpha\beta}^{-1}}\,\mathcal{A}_\alpha + g_{\alpha\beta}^*\mu$ holds for whenever $U_\alpha \cap U_\beta$ is nonempty, then there exits a connection form ω such that $\mathcal{A}_\alpha := s_\alpha^*\omega$ for all α.

Proof. The first part of the theorem follows directly from Proposition 12.13. For the second part we define a $\mathfrak{g}$-valued 1-form on $\wp^{-1}(U_\alpha)$ for each α as follows: For $p \in U_\alpha$, $u = s_\alpha(p)$, $X_p \in T_pM$, and $A \in \mathfrak{g}$, define $\omega_\alpha(T_ps_\alpha \cdot X_p + \upsilon_u(A)) := \mathcal{A}_\alpha(X_p) + A$. This defines ω_α on the set $s_\alpha(U_\alpha) \subset \wp^{-1}(U_\alpha)$. Now any other $u_1 \in \wp^{-1}(U_\alpha)$ is of the form $u_1 = s_\alpha(p)g$ for a unique g that depends smoothly on p. We then let

$$\omega_\alpha(X_{u_1}) := \mathrm{Ad}_{g^{-1}}\,\omega_\alpha(Tr_{g^{-1}} \cdot X_{u_1}).$$

We leave it to the reader to check that ω_α is well defined on $\wp^{-1}(U_\alpha)$ and satisfies $r_g^*\omega_\alpha = \mathrm{Ad}_{g^{-1}}\,\omega_\alpha$ and $\omega_\alpha(\upsilon(A)_u) = A$ for $u \in \wp^{-1}(U_\alpha)$. It then remains to show that if $U_\alpha \cap U_\beta$ is nonempty then $\omega_\alpha = \omega_\beta$ for then we can piece together to define a global connection form. If ω_α and ω_β agree on $s_\beta(U_\alpha \cap U_\beta)$ then they must agree on $\wp^{-1}(U_\alpha \cap U_\beta)$ because we have both $r_g^*\omega_\alpha = \mathrm{Ad}_{g^{-1}}\,\omega_\alpha$ and $r_g^*\omega_\beta = \mathrm{Ad}_{g^{-1}}\,\omega_\beta$ for all g. Also, since

$$\omega_\alpha(\upsilon(A)_u) = A = \omega_\beta(\upsilon(A)_u)$$

for any $u \in \wp^{-1}(U_\alpha \cap U_\beta)$ we need only check that $\omega_\beta(T_p s_\beta \cdot X_p) = \omega_\alpha(T_p s_\alpha \cdot X_p)$ for any arbitrary $X_p \in T_p M$. Note first that $\omega_\beta(T_p s_\beta \cdot X_p) = \mathcal{A}_\beta(X_p)$. But using the calculation of Proposition 12.13 have $Ts_\beta \cdot X_p = Tr_{g_{\alpha\beta}} \cdot (Ts_\alpha \cdot X_p) + \upsilon_{s_\alpha(p) g_{\alpha\beta}(p)} \left(g_{\alpha\beta}^* \mu(X_p) \right)$ and so

$$\begin{aligned}\omega_\alpha(T_p s_\beta \cdot X_p) &= \omega_\beta \left(Tr_{g_{\alpha\beta}} \cdot (Ts_\alpha \cdot X_p) + \upsilon_{s_\alpha(p) g_{\alpha\beta}(p)} \left(g_{\alpha\beta}^* \mu(X_p) \right) \right) \\ &= r_{g_{\alpha\beta}}^* \omega_\alpha(Ts_\alpha \cdot X_p) + g_{\alpha\beta}^* \mu(X_p) \\ &= \mathrm{Ad}_{g_{\alpha\beta}^{-1}} \mathcal{A}_\alpha(X_p) + g_{\alpha\beta}^* \mu(X_p) = \mathcal{A}_\beta(X_p) = \omega_\beta(T_p s_\beta \cdot X_p).\end{aligned}$$

We leave it to the reader to check that $\mathcal{A}_\alpha := s_\alpha^* \omega$ for all α. □

In summary, we may define a connection on a principal G-bundle by an appropriate distribution on the total space P, by a connection form on or by a system of local gauge potentials.

Remark 12.15. In practice G is almost always a matrix group and $\mathfrak{g}$ a Lie algebra of matrices with commutator of the bracket. In this case the transformation law for local change of gauge $\mathcal{A}_\beta = \mathrm{Ad}_{g_{\alpha\beta}^{-1}} \mathcal{A}_\alpha + g_{\alpha\beta}^* \mu$ becomes

$$\mathcal{A}_\beta = g_{\alpha\beta}^{-1} \mathcal{A}_\alpha \, g_{\alpha\beta} + g_{\alpha\beta}^{-1} \, g_{\alpha\beta}.$$

For example, If γ is a curve with $\dot\gamma(0) = X_p$ then

$$\begin{aligned}g_{\alpha\beta}^* \mu(X_p) &= \mu(T_p g_{\alpha\beta} \cdot X_p) = TL_{g_{\alpha\beta}(p)}^{-1} \cdot (T_p g_{\alpha\beta} \cdot X_p) = TL_{g_{\alpha\beta}(p)}^{-1} \left.\frac{d}{dt}\right|_0 g_{\alpha\beta}(\gamma(t)) \\ &= \left.\frac{d}{dt}\right|_0 L_{g_{\alpha\beta}(p)}^{-1} g_{\alpha\beta}(\gamma(t)) = L_{g_{\alpha\beta}(p)}^{-1} \left.\frac{d}{dt}\right|_0 g_{\alpha\beta}(\gamma(t)) = g_{\alpha\beta}^{-1}(p) \, dg_{\alpha\beta}(X_p)\end{aligned}$$

so that, with a proper interpretation, we may write $g_{\alpha\beta}^* \mu = g_{\alpha\beta}^{-1} \, dg_{\alpha\beta}$

12.4. Horizontal Lifting

Let $I \subset \mathbb{R}$ be some compact interval containing 0, and let $\wp : P \to M$ be a principal G-bundle with connection with connection for ω as before. A curve $c : I \to P$ is called horizontal if $\dot c(t) \in \mathcal{H}_{c(t)}$ for all $t \in I$. A lift of a curve $\gamma : I \to M$ is a curve $\widetilde\gamma : I \to P$ such that $\wp \circ \widetilde\gamma = \gamma$. If a lift is a horizontal we call it a horizontal lift. We would like to show that given a curve $\gamma : I \to M$ and a point $u_0 \in \wp^{-1}(\gamma(0))$, there is a unique horizontal lift $\widetilde\gamma : I \to P$ such that $\widetilde\gamma(0) = u_0$.

Exercise 12.16. Show that given $\gamma : I \to M$ and $u_0 \in \wp^{-1}(\gamma(0))$ as above, there is a not necessarily horizontal lift (smooth) $\overline\gamma : I \to P$ such that $\overline\gamma(0) = u_0$. Hint: Use local trivializations (bundle charts) and the Lebesgue number lemma.

Let $\overline{\gamma} : I \to P$ be any lift of γ. Then since the action is free we see that, for any other lift $\widetilde{\gamma} : I \to P$, there must be a smooth function $g : I \to G$ such that $\widetilde{\gamma}(t) = \overline{\gamma}(t)g(t)$ for $t \in I$. Our goal is to show that g can be chosen so that $\widetilde{\gamma}$ is horizontal.

Using Theorem 12.4 we have

$$\frac{d}{dt}\widetilde{\gamma}(t) = \frac{d}{dt}\overline{\gamma}(t)g(t) = T\alpha\cdot(\frac{d}{dt}\overline{\gamma}(t), \frac{d}{dt}g(t)) = Tr_g\cdot\frac{d}{dt}\overline{\gamma}(t)+\upsilon_{\overline{\gamma}(t)g(t)}\left(\mu(\frac{d}{dt}g(t))\right).$$

Now we want $\frac{d}{dt}\widetilde{\gamma}(t)$ to be horizontal so we want $\omega\left(\frac{d}{dt}\widetilde{\gamma}(t)\right) = 0$ for all t. So apply ω to the right hand side of the displayed equation above and set it to zero to get

$$\begin{aligned}
\omega\left(T\alpha\cdot(\frac{d}{dt}\overline{\gamma}(t), \frac{d}{dt}g(t)) = Tr_g\cdot\frac{d}{dt}\overline{\gamma}(t)+\upsilon_{\overline{\gamma}(t)g(t)}\left(\mu(\frac{d}{dt}g(t))\right)\right) &= 0\\
\omega\left(Tr_{g(t)}\cdot\frac{d}{dt}\overline{\gamma}(t)+\upsilon_{\overline{\gamma}(t)g(t)}\left(\mu(\frac{d}{dt}g(t))\right)\right) &= 0\\
\omega\left(Tr_{g(t)}\cdot\frac{d}{dt}\overline{\gamma}(t)\right)+\omega\left(\upsilon_{\overline{\gamma}(t)g(t)}\left(\mu(\frac{d}{dt}g(t))\right)\right) &= 0\\
\mathrm{Ad}_{g(t)^{-1}}\,\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right)+\mu(\frac{d}{dt}g(t)) &= 0\\
TL_{g(t)^{-1}}Tr_{g(t)}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right)+L_{g(t)^{-1}}\frac{d}{dt}g(t) &= 0
\end{aligned}$$

So the equation we want to hold is

$$\begin{aligned}
Tr_{g(t)}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right)+\frac{d}{dt}g(t) &= 0\\
\frac{d}{dt}g(t) &= -Tr_{g(t)}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right).
\end{aligned}$$

Now for any $A \in \mathfrak{g}$, we have a right invariant vector field $g \mapsto \rho_g(A)$ defined by $\rho_g(A) = r_gA$. Thus if we define a time dependent vector field Y on G by $(t, g) \mapsto Y(t, g) := -Tr_g\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right)$ then we are trying to solve $\frac{d}{dt}g(t) = Y(t, g(t))$. In other words, we seek integral curves of the time dependent vector field Y. We know by general existence and uniqueness theory that such curves exist with $g(0) = e$ for t in some interval $(-\epsilon, \epsilon) \subset I$. But if we choose $g_0 \in G$ then $h(t) = g(t)g_0$ is also an integral curve defined on the same interval but with $g(0) = g_0$. Indeed, $h(t) = r_{g_0} \circ g(t)$ and

$$\begin{aligned}\frac{d}{dt}h(t) &= Tr_{g_0} \cdot \frac{d}{dt}g(t) \\ &= -Tr_{g_0} \cdot Tr_{g(t)}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right) = -Tr_{g(t)g_0}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right) \\ &= -Tr_{h(t)}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right)\end{aligned}$$

and so $\frac{d}{dt}h(t) = -Tr_{h(t)}\omega\left(\frac{d}{dt}\overline{\gamma}(t)\right) = Y(t, h(t))$. We conclude that given any u_0 in P there is a horizontal lift through u_0.

Chapter 13
Supplement

13.1. Alternate proof of test case

Theorem 13.1. *Let (M, g) be a semi-Riemannian manifold of dimension n and index ν. If (M, g) is flat, that is, if the curvature tensor is identically zero, then (M, g) is locally isometric to the semi-Euclidean space $\mathbb{R}^n_\nu$.*

Proof. Since this is clearly an entirely local question, we may as well assume that M is some open neighborhood of 0 in $\mathbb{R}^n$ with some metric g of index ν. Let $y_1, ..., y_n$ be standard coordinates on $\mathbb{R}^n$. First, if X_0 is a given nonzero vector in $T_0\mathbb{R}^n$ then we may find a vector field X with $X(0) = X_0$ and $\nabla_Z X \equiv 0$ for all Z. It is enough to show that we can choose such an X with $\nabla_{\partial/\partial y^i} X \equiv 0$ for all i. We first obtain X along the y^1-axis by parallel transport. This defines $X(y^1, 0, ..., 0)$ for all y^1. Then we parallel translate $X(y^1, 0, ..., 0)$ along $y \mapsto (y^1, y, 0, ..., 0)$ for each y^1. This gives us a vector field X along the surface $(y^1, y^2) \mapsto (y^1, y^2, 0, ..., 0)$. By construction, $\nabla_{\partial/\partial y^2} X$ is zero on this whole surface while $\nabla_{\partial/\partial y^1} X$ is zero at least along $y \mapsto (y, 0, ..., 0)$. But Exercise **??** we have

$$\nabla_{\partial/\partial y^1} \nabla_{\partial/\partial y^2} X - \nabla_{\partial/\partial y^2} \nabla_{\partial/\partial y^1} X = R\left(\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}\right) X = 0,$$

so that $\nabla_{\partial/\partial y^2} \nabla_{\partial/\partial y^1} X = 0$ and $\nabla_{\partial/\partial y^1} X$ is parallel along each curve $y \mapsto (y^1, y, 0, ..., 0)$. Since $\nabla_{\partial/\partial y^1} X$ is zero at $(y^1, 0, 0, ..., 0)$ we see that $\nabla_{\partial/\partial y^1} X$ is actually zero on the surface $(y^1, y^1) \mapsto (y^1, y^2, 0, ..., 0)$. An analogous argument allows us to extend X to the 3-dimensional submanifold $\{(y^1, y^2, y^3, ..., 0)\}$ and eventually to the whole neighborhood of 0.

Given that we can extend any vector in $T_0\mathbb{R}^n$ to a parallel field, we can extend a whole orthonormal basis of $T_0\mathbb{R}^n$ to parallel frame fields $X_1, ..., X_n$ in some neighborhood of 0.

Next we use the fact that the Levi-Civita connection is symmetric (torsion zero). We have $\nabla_{X_i} X_j - \nabla_{X_j} X_i - [X_i, X_j] = 0$ for all i, j. But since the X_i are parallel this means that $[X_i, X_j] \equiv 0$. Of course, this means that there exists coordinates $x^1, ..., x^n$ such that

$$\frac{\partial}{\partial x^i} = X_i \text{ for all } i.$$

On the other hand we know that these fields are orthonormal since parallel translation preserves scalar products. The result is that these coordinates give a chart which is an isometry with a neighborhood of 0 in the semi-Riemannian space $\mathbb{R}^n_\nu$. $\square$

Chapter 14

Complex Manifolds

14.1. Some complex linear algebra

The set of all n-tuples of complex $\mathbb{C}^n$ numbers is a complex vector space and by choice of a basis, every complex vector space of finite dimension (over $\mathbb{C}$) is linearly isomorphic to $\mathbb{C}^n$ for some n. Now multiplication by $i := \sqrt{-1}$ is a complex linear map $\mathbb{C}^n \to \mathbb{C}^n$ and since $\mathbb{C}^n$ is also a real vector space $\mathbb{R}^{2n}$ under the identification

$$(x^1 + iy^1, ..., x^n + iy^n) \rightleftharpoons (x^1, y^1, ..., x^n, y^n)$$

we obtain multiplication by i as a real linear map $J_0 : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by the matrix

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

Conversely, if V is a real vector space of dimension $2n$ and there is a map $J : \mathrm{V} \to \mathrm{V}$ with $J^2 = -1$ then we can define the structure of a complex vector space on V by defining the scalar multiplication by complex numbers via the formula

$$(x + \mathrm{i}y)v := xv + yJv \text{ for } v \in \mathrm{V}.$$

Denote this complex vector space by V_J. Now if $e_1, ...e_n$ is a basis for V_J (over $\mathbb{C}$) then we claim that $e_1, ..., e_n, Je_1, ..., Je_n$ is a basis for V over $\mathbb{R}$. We only need to show that $e_1, ..., e_n, Je_1, ..., Je_n$ span. For this let

$v \in \mathrm{V}$ and then for some complex numbers $c^i = a^i + \mathrm{i}b^j$ we have $\sum c^i e_i = \sum (a^j + \mathrm{i}b^j)e_j = \sum a^j e_j + \sum b^j J e_j$.

Next we consider the complexification of V which is $\mathrm{V}_{\mathbb{C}} := \mathbb{C} \otimes \mathrm{V}$. Now any real basis $\{f_j\}$ of V is also a basis for $\mathrm{V}_{\mathbb{C}}$ iff we identify f_j with $1 \otimes f_j$. Furthermore, the linear map $J : \mathrm{V} \to \mathrm{V}$ extends to a complex linear map $J : \mathrm{V}_{\mathbb{C}} \to \mathrm{V}_{\mathbb{C}}$ and still satisfies $J^2 = -1$. Thus this extension has eigenvalues i and $-$i. Let $\mathrm{V}^{1,0}$ be the i eigenspace and $\mathrm{V}^{0,1}$ be the $-$i eigenspace. Of course we must have $\mathrm{V}_{\mathbb{C}} = \mathrm{V}^{1,0} \oplus \mathrm{V}^{0,1}$. The reader may check that the set of vectors $\{e_1 - \mathrm{i}Je_1, ..., e_n - \mathrm{i}Je_n\}$ span $\mathrm{V}^{1,0}$ while $\{e_1 + \mathrm{i}Je_1, ..., e_n + \mathrm{i}Je_n\}$ span $\mathrm{V}^{0,1}$. Thus we have a convenient basis for $\mathrm{V}_{\mathbb{C}} = \mathrm{V}^{1,0} \oplus \mathrm{V}^{0,1}$.

Lemma 14.1. *There is a natural complex linear isomorphism* $\mathrm{V}_J \cong \mathrm{V}^{1,0}$ *given by* $e_i \mapsto e_i - \mathrm{i}Je_i$. *Furthermore, the conjugation map on* $\mathrm{V}_{\mathbb{C}}$ *interchanges the spaces* $\mathrm{V}^{1,0}$ *and* $\mathrm{V}^{0,1}$.

Let us apply these considerations to the simple case of the complex plane $\mathbb{C}$. The realification is $\mathbb{R}^2$ and the map J is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

If we identify the tangent space of $\mathbb{R}^{2n}$ at 0 with $\mathbb{R}^{2n}$ itself then $\{ \frac{\partial}{\partial x^i}\big|_0, \frac{\partial}{\partial y^i}\big|_0 \}_{1 \le i \le n}$ is basis for $\mathbb{R}^{2n}$. A| complex basis for $\mathbb{C}^n \cong (\mathbb{R}^{2n}, J_0)$ is, for instance, $\{ \frac{\partial}{\partial x^i}\big|_0 \}_{1 \le i \le n}$. A complex basis for $\mathbb{R}^2_J \cong \mathbb{C}$ is $e_1 = \frac{\partial}{\partial x}\big|_0$ and so $\frac{\partial}{\partial x}\big|_0, J\frac{\partial}{\partial x}\big|_0$ is a basis for $\mathbb{R}^2$. This is clear anyway since $J\frac{\partial}{\partial x}\big|_0 = \frac{\partial}{\partial y}\big|_0$. Now the complexification of $\mathbb{R}^2$ is $\mathbb{R}^2_{\mathbb{C}}$ which has basis consisting of $e_1 - \mathrm{i}Je_1 = \frac{\partial}{\partial x}\big|_0 - \mathrm{i}\frac{\partial}{\partial y}\big|_0$ and $e_1 + \mathrm{i}Je_1 = \frac{\partial}{\partial x}\big|_0 + \mathrm{i}\frac{\partial}{\partial y}\big|_0$. These are usually denoted by $\frac{\partial}{\partial z}\big|_0$ and $\frac{\partial}{\partial \bar{z}}\big|_0$. More generally, we see that if $\mathbb{C}^n$ is reified to $\mathbb{R}^{2n}$ which is then complexified to $\mathbb{R}^{2n}_{\mathbb{C}} := \mathbb{C} \otimes \mathbb{R}^{2n}$ then a basis for $\mathbb{R}^{2n}_{\mathbb{C}}$ is given by

$$\{ \frac{\partial}{\partial z^1}\bigg|_0, .., \frac{\partial}{\partial z^n}\bigg|_0, \frac{\partial}{\partial \bar{z}^1}\bigg|_0 ..., \frac{\partial}{\partial \bar{z}^n}\bigg|_0 \}$$

where

$$2\,\frac{\partial}{\partial z^i}\bigg|_0 := \frac{\partial}{\partial x^i}\bigg|_0 - \mathrm{i}\,\frac{\partial}{\partial y^i}\bigg|_0$$

and

$$2\,\frac{\partial}{\partial \bar{z}^i}\bigg|_0 := \frac{\partial}{\partial x^i}\bigg|_0 + \mathrm{i}\,\frac{\partial}{\partial y^i}\bigg|_0.$$

Now if we consider the tangent bundle $U \times \mathbb{R}^{2n}$ of an open set $U \subset \mathbb{R}^{2n}$ then we have the vector fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$. We can complexify the tangent bundle of $U \times \mathbb{R}^{2n}$ to get $U \times \mathbb{R}^{2n}_{\mathbb{C}}$ and then following the ideas above we have that the fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ also span each tangent space $T_pU := \{p\} \times \mathbb{R}^{2n}_{\mathbb{C}}$. On the other hand, so do the fields $\{\frac{\partial}{\partial z^1},..,\frac{\partial}{\partial z^n},\frac{\partial}{\partial \bar{z}^1},..,\frac{\partial}{\partial \bar{z}^n}\}$. Now if $\mathbb{R}^{2n}$ had a complex vector space structure, say $\mathbb{C}^n \cong (\mathbb{R}^{2n}, J_0)$, then J_0 defines a bundle map $J_0 : T_pU \to T_pU$ given by $(p, v) \mapsto (p, J_0 v)$. This can be extended to a complex bundle map $J_0 : TU_{\mathbb{C}} = \mathbb{C}\otimes TU \to TU_{\mathbb{C}} = \mathbb{C}\otimes TU$ and we get a bundle decomposition

$$TU_{\mathbb{C}} = T^{1.0}U \oplus T^{0.1}U$$

where $\frac{\partial}{\partial z^1},..,\frac{\partial}{\partial z^n}$ spans $T^{1.0}U$ at each point and $\frac{\partial}{\partial \bar{z}^1},..,\frac{\partial}{\partial \bar{z}^n}$ spans $T^{0.1}U$.

Now the symbols $\frac{\partial}{\partial z^1}$ etc., already have a meaning a differential operators. Let us now show that this view is at least consistent with what we have done above. For a smooth complex valued function $f : U \subset \mathbb{C}^n \to \mathbb{C}$ we have for $p = (z_1, ..., z_n) \in U$

$$\begin{aligned}
\left.\frac{\partial}{\partial z^i}\right|_p f &= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p f - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p f\right) \\
&= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p u - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p u - \left.\frac{\partial}{\partial x^i}\right|_p \mathrm{i}v - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p \mathrm{i}v\right) \\
&\frac{1}{2}\left(\left.\frac{\partial u}{\partial x^i}\right|_p + \left.\frac{\partial v}{\partial y^i}\right|_p\right) + \frac{\mathrm{i}}{2}\left(\left.\frac{\partial u}{\partial y^i}\right|_p - \left.\frac{\partial v}{\partial x^i}\right|_p\right).
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f &= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p f + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p f\right) \\
&= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p u + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p u + \left.\frac{\partial}{\partial x^i}\right|_p \mathrm{i}v + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p \mathrm{i}v\right) \\
&= \frac{1}{2}\left(\left.\frac{\partial u}{\partial x^i}\right|_p - \left.\frac{\partial v}{\partial y^i}\right|_p\right) + \frac{\mathrm{i}}{2}\left(\left.\frac{\partial u}{\partial y^i}\right|_p + \left.\frac{\partial v}{\partial x^i}\right|_p\right).
\end{aligned}$$

Definition 14.2. A function $f : U \subset \mathbb{C}^n \to \mathbb{C}$ is called **holomorphic** if

$$\frac{\partial}{\partial \bar{z}^i} f \equiv 0 \qquad \text{(all } i)$$

on U. A function f is called **antiholomorphic** if

$$\frac{\partial}{\partial z^i} f \equiv 0 \qquad \text{(all } i).$$

Definition 14.3. A map $f : U \subset \mathbb{C}^n \to \mathbb{C}^m$ given by functions $f_1, ..., f_m$ is called **holomorphic** (resp. **antiholomorphic**) if each component function $f_1, ..., f_m$ is holomorphic (resp. antiholomorphic).

Now if $f : U \subset \mathbb{C}^n \to \mathbb{C}$ is holomorphic then by definition $\frac{\partial}{\partial \bar{z}^i}\big|_p f \equiv 0$ for all $p \in U$ and so we have the **Cauchy-Riemann equations**

(Cauchy-Riemann)
$$\frac{\partial u}{\partial x^i} = \frac{\partial v}{\partial y^i}$$
$$\frac{\partial v}{\partial x^i} = -\frac{\partial u}{\partial y^i}$$

and from this we see that for holomorphic f

$$\frac{\partial f}{\partial z^i} = \frac{\partial u}{\partial x^i} + \mathrm{i}\frac{\partial v}{\partial x^i} = \frac{\partial f}{\partial x^i}$$

which means that as derivations on the sheaf $\mathcal{O}$ of locally defined holomorphic functions on $\mathbb{C}^n$, the operators $\frac{\partial}{\partial z^i}$ and $\frac{\partial}{\partial x^i}$ are equal. This corresponds to the complex isomorphism $T^{1.0}U \cong TU, J_0$ which comes from the isomorphism in lemma **??**. In fact, if one looks at a function $f : \mathbb{R}^{2n} \to \mathbb{C}$ as a differentiable map of real manifolds then with J_0 given the isomorphism $\mathbb{R}^{2n} \cong \mathbb{C}^n$, our map f is holomorphic iff

$$Tf \circ J_0 = J_0 \circ Tf$$

or in other words

$$\begin{pmatrix} \frac{\partial u}{\partial x^1} & \frac{\partial u}{\partial y^1} & \\ \frac{\partial v}{\partial x^1} & \frac{\partial v}{\partial y^1} & \\ & & \ddots \end{pmatrix} \begin{pmatrix} 0 & -1 & \\ 1 & 0 & \\ & & \ddots \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x^1} & \frac{\partial u}{\partial y^1} & \\ \frac{\partial v}{\partial x^1} & \frac{\partial v}{\partial y^1} & \\ & & \ddots \end{pmatrix} \begin{pmatrix} 0 & -1 & \\ 1 & 0 & \\ & & \ddots \end{pmatrix}.$$

This last matrix equation is just the Cauchy-Riemann equations again.

14.2. Complex structure

Definition 14.4. A manifold M is said to be an **almost complex manifold** if there is a smooth bundle map $J : TM \to TM$, called an **almost complex structure**, having the property that $J^2 = -1$.

Definition 14.5. A **complex manifold** M is a manifold modeled on $\mathbb{C}^n$ for some n, together with an atlas for M such that the transition functions are all holomorphic maps. The charts from this atlas are called **holomorphic charts**. We also use the phrase "**holomorphic coordinates**".

Example 14.6. Let $S^2(1/2) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1/4\}$ be given coordinates $\psi^+ : (x_1, x_2, x_3) \mapsto \frac{1}{1-x_3}(x_1 + \mathrm{i}x_2) \in \mathbb{C}$ on $U^+ := \{(x_1, x_2, x_3) \in S^2 : 1 - x_3 \neq 0\}$ and $\psi^- : (x_1, x_2, x_3) \mapsto \frac{1}{1+x_3}(x_1 + \mathrm{i}x_2) \in \mathbb{C}$ on $U^- := \{(x_1, x_2, x_3) \in S^2 : 1 + x_3 \neq 0\}$. Then z and w are coordinates on

$S^2(1/2)$ with transition function $\psi^- \circ \psi^+(z) = 1/z$. Since on $\psi^+U^+ \cap \psi^- U^-$ the map $z \mapsto 1/z$ is a biholomorphism we see that $S^2(1/2)$ can be given the structure of a complex 1-manifold.

Another way to get the same complex 1-manifold is by taking two copies of the complex pane, say $\mathbb{C}_z$ with coordinate z and $\mathbb{C}_w$ with coordinate z and then identify $\mathbb{C}_z$ with $\mathbb{C}_w - \{0\}$ via the map $w = 1/z$. This complex surface is of course topologically a sphere and is also the 1 point compactification of the complex plane. As the reader will not doubt already be aware, this complex 1-manifold is called the **Riemann sphere**.

Example 14.7. Let $P_n(\mathbb{C})$ be the set of all complex lines through the origin in $\mathbb{C}^{n+1}$, which is to say, the set of all equivalence classes of nonzero elements of $\mathbb{C}^{n+1}$ under the equivalence relation

$$(z^1, ..., z^{n+1}) \sim \lambda(z^1, ..., z^{n+1}) \text{ for } \lambda \in \mathbb{C}$$

For each i with $1 \leq i \leq n+1$ define the set

$$U_i := \{[z^1, ..., z^{n+1}] \in P_n(\mathbb{C}) : z^i \neq 0\}$$

and corresponding map $\psi_i : U_i \to \mathbb{C}^n$ by

$$\psi_i([z^1, ..., z^{n+1}]) = \frac{1}{z^i}(z^1, ..., \widehat{z^i}, ..., z^{n+1}) \in \mathbb{C}^n.$$

One can check that these maps provide a holomorphic atlas for $P_n(\mathbb{C})$ which is therefore a complex manifold (**complex projective n-space**).

Example 14.8. Let $\mathbb{C}^m_n$ be the space of $m \times n$ complex matrices. This is clearly a complex manifold since we can always "line up" the entries to get a map $\mathbb{C}^m_n \to \mathbb{C}^{mn}$ and so as complex manifolds $\mathbb{C}^m_n \cong \mathbb{C}^{mn}$. A little less trivially we have the complex general linear group $\mathrm{GL}(n, \mathbb{C})$ which is an open subset of $\mathbb{C}^n_n$ and so is an n^2 dimensional complex manifold.

Example 14.9 (Grassmannian manifold)**.** To describe this important example we start with the set $(\mathbb{C}^n_k)_*$ of $n \times k$ matrices with rank $k < n$ (maximal rank). The columns of each matrix from $(\mathbb{C}^n_k)_*$ span a k-dimensional subspace of $\mathbb{C}^n$. Define two matrices from $(\mathbb{C}^n_k)_*$ to be equivalent if they span the same k-dimensional subspace. Thus the set $G(k, n)$ of equivalence classes is in one to one correspondence with the set of complex k dimensional subspaces of $\mathbb{C}^n$. Now let U be the set of all $[A] \in G(k, n)$ such that A has its first k rows linearly independent. This property is independent of the representative A of the equivalence class $[A]$ and so U is a well defined set. This last fact is easily proven by a Gaussian reduction argument. Now every element $[A] \in U \subset G(k, n)$ is an equivalence class that has a unique member A_0 of the form

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

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

Definition 14.10. A complex 1-manifold (so real dimension is 2) is called a **Riemann surface**.

If S is a subset of a complex manifold M such that near each $p_0 \in S$ there exists a holomorphic chart $U, \psi = (z^1, \dots, z^n)$ such that $0 \in S \cap U$ iff $z^{k+1}(p) = \cdots = z^n(p) = 0$ then the coordinates $z^1, \dots, z^k$ restricted to $U \cap S$ is a chart on the set S and the set of all such charts gives S the structure of a complex manifold. In this case we call S a **complex submanifold** of M.

Definition 14.11. In the same way as we defined differentiability for real manifolds we define the notion of a **holomorphic map** (resp. **antiholomorphic map**) from one complex manifold to another. Note however, that we must use holomorphic charts for the definition.

The proof of the following lemma is straightforward.

Lemma 14.12. *Let $\psi : U \to \mathbb{C}^n$ be a holomorphic chart with $p \in U$. Then writing $\psi = (z^1, \dots, z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have that the map $J_p : T_pM \to T_pM$ defined by*

$$J_p \left.\frac{\partial}{\partial x^i}\right|_p = \left.\frac{\partial}{\partial y^i}\right|_p$$
$$J_p \left.\frac{\partial}{\partial y^i}\right|_p = -\left.\frac{\partial}{\partial x^i}\right|_p$$

is well defined independent of the choice of coordinates.

The maps J_p combine to give a bundle map $J : TM \to TM$ and so an almost complex structure on M called the almost complex structure induced by the holomorphic atlas.

Definition 14.13. An almost complex structure J on M is said to be **integrable** if there it has a holomorphic atlas giving the map J as the induced almost complex structure. That is if there is an family of admissible charts $\psi_\alpha : U_\alpha \to \mathbb{R}^{2n}$ such that after identifying $\mathbb{R}^{2n}$ with $\mathbb{C}^n$ the charts form a holomorphic atlas with J the induced almost complex structure. In this case, we call J a **complex structure**.

14.3. Complex Tangent Structures

Let $\mathcal{F}_p(\mathbb{C})$ denote the algebra germs of *complex* valued smooth functions at p on a complex n-manifold M thought of as a smooth real $2n-$manifold with real tangent bundle TM. Let $\mathrm{Der}_p(\mathcal{F})$ be the space of derivations this algebra. It is not hard to see that this space is isomorphic to the complexified tangent space $T_pM_\mathbb{C} = \mathbb{C}\otimes T_pM$. The (complex) algebra of germs of holomorphic functions at a point p in a complex manifold is denoted $\mathcal{O}_p$ and the set of derivations of this algebra denoted $\mathrm{Der}_p(\mathcal{O})$. We also have the algebra of germs of antiholomorphic functions at p which is $\overline{\mathcal{O}}_p$ and also $\mathrm{Der}_p(\overline{\mathcal{O}})$, the derivations of this algebra.

If $\psi : U \to \mathbb{C}^n$ is a holomorphic chart then writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have the differential operators at $p \in U$:

$$\left\{ \left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\}$$

(now transferred to the manifold). To be pedantic about it, we now denote the coordinates on $\mathbb{C}^n$ by $w_i = u_i + \mathrm{i}v_i$ and then

$$\left.\frac{\partial}{\partial z^i}\right|_p f := \left.\frac{\partial f \circ \psi^{-1}}{\partial w^i}\right|_{\psi(p)}$$
$$\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f := \left.\frac{\partial f \circ \psi^{-1}}{\partial \bar{w}^i}\right|_{\psi(p)}$$

Thought of derivations these span $\mathrm{Der}_p(\mathcal{F})$ but we have also seen that they span the complexified tangent space at p. In fact, we have the following:

$$\begin{aligned}
T_pM_{\mathbb{C}} &= \mathrm{span}_{\mathbb{C}}\left\{ \left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\} = \mathrm{Der}_p(\mathcal{F}) \\
T_pM^{1,0} &= \mathrm{span}_{\mathbb{C}}\left\{ \left.\frac{\partial}{\partial z^i}\right|_p \right\} \\
&= \{v \in \mathrm{Der}_p(\mathcal{F}) : vf = 0 \text{ for all } f \in \overline{\mathcal{O}}_p\} \\
T_pM^{0,1} &= \mathrm{span}_{\mathbb{C}}\left\{ \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\} \\
&= \{v \in \mathrm{Der}_p(\mathcal{F}) : vf = 0 \text{ for all } f \in \mathcal{O}_p\}
\end{aligned}$$

and of course

$$T_pM = \mathrm{span}_{\mathbb{R}}\left\{ \left.\frac{\partial}{\partial x^i}\right|_p, \left.\frac{\partial}{\partial y^i}\right|_p \right\}.$$

The reader should go back and check that the above statements are consistent with our definitions as long as we view the $\left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p$ not only as the algebraic objects constructed above but also as derivations. Also, the definitions of $T_pM^{1,0}$ and $T_pM^{0,1}$ are independent of the holomorphic coordinates since we also have

$$T_pM^{1,0} = \ker\{J_p : T_pM \to T_pM\}$$

14.4. The holomorphic tangent map.

We leave it to the reader to verify that the construction that we have at each tangent space globalize to give natural vector bundles $TM_{\mathbb{C}}, TM^{1,0}$ and $TM^{0,1}$ (all with M as base space).

Let M and N be complex manifolds and let $f : M \to N$ be a smooth map. The tangent map extend to a map of the complexified bundles $Tf : TM_{\mathbb{C}} \to TN_{\mathbb{C}}$. Now $TM_{\mathbb{C}} = TM^{1,0} \oplus TM^{0,1}$ and similarly $TM_{\mathbb{C}} = TN^{1,0} \oplus TN^{0,1}$. If f is holomorphic then $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$. In fact since it is easily verified that $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$ equivalent to the Cauchy-Riemann equations being satisfied by the local representative on F in any holomorphic chart we obtain the following

Proposition 14.14. *$Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$ if and only if f is a holomorphic map.*

The map given by the restriction $T_pf : T_pM^{1,0} \to T_{fp}N^{1,0}$ is called the **holomorphic tangent** map at p. Of course, these maps concatenate to give a bundle map

14.5. Dual spaces

Let M, J be a complex manifold. The dual of $T_pM_{\mathbb{C}}$ is $T_p^*M_{\mathbb{C}} = \mathbb{C} \otimes T_p^*M$. Now the map J has a dual bundle map $J^* : T^*M_{\mathbb{C}} \to T^*M_{\mathbb{C}}$ that must also satisfy $J^* \circ J^* = -1$ and so we have the at each $p \in M$ the decomposition by eigenspaces

$$T_p^*M_{\mathbb{C}} = T_p^*M^{1,0} \oplus T_p^*M^{0,1}$$

corresponding to the eigenvalues $\pm\mathrm{i}$.

Definition 14.15. The space $T_p^*M^{1,0}$ is called the space of holomorphic co-vectors at p while $T_p^*M^{0,1}$ is the space of antiholomorphic covector at p.

We now choose a holomorphic chart $\psi : U \to \mathbb{C}^n$ at p. Writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have the 1-forms

$$dz^k = dx^k + \mathrm{i}dy^k$$

and

$$d\bar{z}^k = dx^k - \mathrm{i}dy^k.$$

Equivalently, the pointwise definitions are $dz^k\big|_p = dx^k\big|_p + \mathrm{i}\ dy^k\big|_p$ and $d\bar{z}^k\big|_p = dx^k\big|_p - \mathrm{i}\ dy^k\big|_p$. Notice that we have the expected relations:

$$\begin{aligned} dz^k(\frac{\partial}{\partial z^i}) &= (dx^k + \mathrm{i}dy^k)(\frac{1}{2}\frac{\partial}{\partial x^i} - \mathrm{i}\frac{1}{2}\frac{\partial}{\partial y^i}) \\ &= \frac{1}{2}\delta_j^k + \frac{1}{2}\delta_j^k = \delta_j^k \\ dz^k(\frac{\partial}{\partial \bar{z}^i}) &= (dx^k + \mathrm{i}dy^k)(\frac{1}{2}\frac{\partial}{\partial x^i} + \mathrm{i}\frac{1}{2}\frac{\partial}{\partial y^i}) \\ &= 0 \end{aligned}$$

and similarly

$$d\bar{z}^k(\frac{\partial}{\partial \bar{z}^i}) = \delta_j^k \text{ and } d\bar{z}^k(\frac{\partial}{\partial z^i}) = \delta_j^k.$$

Let us check the action of J^* on these forms:

$$\begin{aligned} J^*(dz^k)(\frac{\partial}{\partial z^i}) &= J^*(dx^k + \mathrm{i}dy^k)(\frac{\partial}{\partial z^i}) \\ &= (dx^k + \mathrm{i}dy^k)(J\frac{\partial}{\partial z^i}) \\ &= \mathrm{i}(dx^k + \mathrm{i}dy^k)\frac{\partial}{\partial z^i} \\ &= \mathrm{i}dz^k(\frac{\partial}{\partial z^i}) \end{aligned}$$

and

$$\begin{aligned} J^*(dz^k)(\frac{\partial}{\partial \bar{z}^i}) &= dz^k(J\frac{\partial}{\partial \bar{z}^i}) \\ &= -\mathrm{i}dz^k(\frac{\partial}{\partial \bar{z}^i}) = 0 = \\ &= \mathrm{i}dz^k(\frac{\partial}{\partial \bar{z}^i}). \end{aligned}$$

Thus we conclude that $dz^k\big|_p \in T_p^*M^{1,0}$. A similar calculation shows that $d\bar{z}^k\big|_p \in T_p^*M^{0,1}$ and in fact

$$T_p^*M^{1,0} = \mathrm{span}\left\{ dz^k\Big|_p : k = 1, ..., n \right\}$$
$$T_p^*M^{0,1} = \mathrm{span}\left\{ d\bar{z}^k\Big|_p : k = 1, ..., n \right\}$$

and $\{ dz^1\big|_p, ..., dz^n\big|_p, d\bar{z}^1\big|_p, ..., d\bar{z}^n\big|_p \}$ is a basis for $T_p^*M_{\mathbb{C}}$.

Remark 14.16. If we don't specify base points then we are talking about fields (over some open set) that form a basis for each fiber separately. These are called frame fields (e.g. $\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^i}$) or coframe fields (e.g. $dz^k, d\bar{z}^k$).

14.6. The holomorphic inverse and implicit functions theorems.

Let $(z^1, ..., z^n)$ and $(w^1, ..., w^m)$ be local coordinates on complex manifolds M and N respectively. Consider a smooth map $f : M \to N$. We suppose that $p \in M$ is in the domain of $(z^1, ..., z^n)$ and that $q = f(p)$ is in the domain of the coordinates $(w^1, ..., w^m)$. Writing $z^i = x^i + \mathrm{i}y^i$ and $w^i = u^i + \mathrm{i}v^i$ we have the following Jacobian matrices:

(1) In we consider the underlying real structures then we have the Jacobian given in terms of the frame $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ and $\frac{\partial}{\partial u^i}, \frac{\partial}{\partial v^i}$

$$J_p(f) = \begin{bmatrix} \frac{\partial u^1}{\partial x^1}(p) & \frac{\partial u^1}{\partial y^1}(p) & \frac{\partial u^1}{\partial x^2}(p) & \frac{\partial u^1}{\partial y^2}(p) & \cdots \\ \frac{\partial v^1}{\partial x^1}(p) & \frac{\partial v^1}{\partial y^1}(p) & \frac{\partial v^1}{\partial x^2}(p) & \frac{\partial v^1}{\partial y^2}(p) & \cdots \\ \frac{\partial u^2}{\partial x^1}(p) & \frac{\partial u^2}{\partial y^1}(p) & \frac{\partial u^2}{\partial x^2}(p) & \frac{\partial u^2}{\partial y^2}(p) & \cdots \\ \frac{\partial v^2}{\partial x^1}(p) & \frac{\partial v^2}{\partial y^1}(p) & \frac{\partial v^2}{\partial x^2}(p) & \frac{\partial v^2}{\partial y^2}(p) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

(2) With respect to the bases $\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^i}$ and $\frac{\partial}{\partial w^i}, \frac{\partial}{\partial \bar{w}^i}$ we have

$$J_{p,\mathbb{C}}(f) = \begin{bmatrix} J_{11} & J_{12} & \cdots \\ J_{12} & J_{22} & \\ \vdots & & \end{bmatrix}$$

where the J_{ij} are blocks of the form

$$\begin{bmatrix} \frac{\partial w^i}{\partial z^j} & \frac{\partial w^i}{\partial \bar{z}^j} \\ \frac{\partial \bar{w}^i}{\partial z^j} & \frac{\partial \bar{w}^i}{\partial \bar{z}^j} \end{bmatrix}.$$

If f is holomorphic then these block reduce to the form

$$\begin{bmatrix} \frac{\partial w^i}{\partial z^j} & 0 \\ 0 & \frac{\partial \bar{w}^i}{\partial \bar{z}^j} \end{bmatrix}.$$

It is convenient to put the frame fields in the order $\frac{\partial}{\partial z^1}, ..., \frac{\partial}{\partial z^n}, \frac{\partial}{\partial \bar{z}^1}, ..., \frac{\partial}{\partial \bar{z}^n}$ and similarly for the $\frac{\partial}{\partial w^i}, \frac{\partial}{\partial \bar{w}^i}$. In this case we have for holomorphic f

$$\mathcal{J}_{p,\mathbb{C}}(f) = \begin{bmatrix} J^{1,0} & 0 \\ 0 & \overline{J^{1,0}} \end{bmatrix}$$

where

$$J^{1,0}(f) = \left[\frac{\partial w^i}{\partial z^j}\right]$$
$$\overline{J^{1,0}}(f) = \left[\frac{\partial \bar{w}^i}{\partial \bar{z}^j}\right].$$

We shall call a basis arising from a holomorphic coordinate system "separated" when arranged this way. Note that $J^{1,0}$ is just the Jacobian of the holomorphic tangent map $T^{1,0}f : T^{1,0}M \to T^{1,0}N$ with respect to this the **holomorphic frame** $\frac{\partial}{\partial z^1}, ..., \frac{\partial}{\partial z^n}$.

We can now formulate the following version of the inverse mapping theorem:

Theorem 14.17. *(1) Let U and V be open set in $\mathbb{C}^n$ and suppose that the map $f : U \to V$ is holomorphic with $J^{1,0}(f)$ nonsingular at $p \in U$. Then there exists an open set $U_0 \subset U$ containing p such that $f|_{U_0} : U_0 \to f(U_0)$ is a 1-1 holomorphic map with holomorphic inverse. That is, $f|_{U_0}$ is* ***biholomorphic****.*

(2) Similarly, if $f : U \to V$ is holomorphic map between open sets of complex manifolds M and N then if $T_p^{1,0}f : T_p^{1,0}M \to T_{fp}^{1,0}N$ is a linear isomorphism then f is a biholomorphic map when restricted to a possibly smaller open set containing p.

We also have a holomorphic version of the implicit mapping theorem.

Theorem 14.18. *(1) Let $f : U \subset \mathbb{C}^n \to V \subset \mathbb{C}^k$ and let the component functions of f be $f_1, ..., f_k$. If $J_p^{1,0}(f)$ has rank k then there are holomorphic* *functions* $g^1, g^2, ..., g^k$ *defined near* $0 \in \mathbb{C}^{n-k}$ *such that*

$$f(z^1, ..., z^n) = p$$
$$\Leftrightarrow$$
$$z^j = g^j(z^{k+1}, ..., z^n) \text{ for } j = 1, .., k$$

(2) If $f : M \to N$ *is a holomorphic map of complex manifolds and if for fixed* $q \in N$ *we have that each* $p \in f^{-1}(p)$ *is regular in the sense that* $T_p^{1,0} f : T_p^{1,0} M \to T_{fp}^{1,0} N$ *is surjective, then* $S := f^{-1}(p)$ *is a complex submanifold of (complex) dimension* $n - k$.

Example 14.19. The map $\varphi : \mathbb{C}^{n+1} \to \mathbb{C}$ given by $(z^1, ..., z^{n+1}) \mapsto (z^1)^2 + \cdots + (z^{n+1})^2$ has Jacobian at any $(z^1, ..., z^{n+1})$ given by

$$\begin{bmatrix} 2z^1 & 2z^2 & \cdots & 2z^{n+1} \end{bmatrix}$$

which has rank 1 as long as $(z^1, ..., z^{n+1}) \neq 0$. Thus $\varphi^{-1}(1)$ is a complex submanifold of $\mathbb{C}^{n+1}$ having (complex) dimension n. **Warning**: This is not the same as the sphere given by $\left|z^1\right|^2 + \cdots + \left|z^{n+1}\right|^2 = 1$ which is a real submanifold of $\mathbb{C}^{n+1} \cong \mathbb{R}^{2n+2}$ of real dimension $2n + 1$.

Chapter 15

Symplectic Geometry

Equations are more important to me, because politics is for the present, but an equation is something for eternity

-Einstein

15.1. Symplectic Linear Algebra

A (real) **symplectic vector space** is a pair V, α where V is a (real) vector space and α is a nondegenerate alternating (skew-symmetric) bilinear form $\alpha : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$. The basic example is $\mathbb{R}^{2n}$ with

$$\alpha_0(x, y) = x^t J_n y$$

where

$$J_n = \begin{pmatrix} 0 & I_{n\times n} \\ -I_{n\times n} & 0 \end{pmatrix}.$$

The standard symplectic form on α_0 is typical. It is a standard fact from linear algebra that for any N dimensional symplectic vector space V, α there is a basis $e_1, ..., e_n, f^1, ..., f^n$ called a **symplectic basis** such that the matrix that represents α with respect to this basis is the matrix J_n. Thus we may write

$$\alpha = e^1 \wedge f_1 + ... + e^n \wedge f_n$$

where $e^1, ..., e^n, f_1, ..., f_n$ is the dual basis to $e_1, ..., e_n, f^1, ..., f^n$. If V, η is a vector space with a not necessarily nondegenerate alternating form η then we can define the null space

$$N_\eta = \{v \in \mathrm{V} : \eta(v, w) = 0 \text{ for all } w \in \mathrm{V}\}.$$

On the quotient space $\overline{\mathrm{V}} = \mathrm{V}/N_\eta$ we may define $\overline{\eta}(\overline{v}, \overline{w}) = \eta(v, w)$ where v and w represent the elements $\overline{v}, \overline{w} \in \overline{\mathrm{V}}$. Then $\overline{\mathrm{V}}, \overline{\eta}$ is a symplectic vector space called the symplectic reduction of V, η.

Proposition 15.1. For any $\eta \in \bigwedge \mathrm{V}^*$ (regarded as a bilinear form) there is linearly independent set of elements $e^1, ..., e^k, f_1, ..., f_k$ from V^* such that

$$\eta = e^1 \wedge f_1 + ... + e^k \wedge f_k$$

where $\dim(\mathrm{V}) - 2k \geq 0$ is the dimension of N_η.

Definition 15.2. Note: The number k is called the **rank** of η. The matrix that represents η actually has rank $2k$ and so some might call k the **half rank** of η.

Proof. Consider the symplectic reduction $\overline{\mathrm{V}}, \overline{\eta}$ of V, η and choose set of elements $e^1, ..., e^k, f_1, ..., f_k$ such that $\bar{e}^1, ..., \bar{e}^k, \bar{f}_1, ..., \bar{f}_k$ form a symplectic basis of $\overline{\mathrm{V}}, \overline{\eta}$. Add to this set a basis $b_1, ..., b_l$ a basis for N_η and verify that $e^1, ..., e^k, f_1, ..., f_k, b_1, ..., b_l$ must be a basis for V. Taking the dual basis one can check that

$$\eta = e^1 \wedge f_1 + ... + e^k \wedge f_k$$

by testing on the basis $e^1, ..., e^k, f_1, ..., f_k, b_1, ..., b_l$. □

Now if W is a subspace of a symplectic vector space then we may define

$$\mathrm{W}^\perp = \{v \in \mathrm{V} : \eta(v, w) = 0 \text{ for all } w \in \mathrm{W}\}$$

and it is true that $\dim(\mathrm{W}) + \dim(\mathrm{W}^\perp) = \dim(\mathrm{V})$ but it is **not** necessarily the case that $\mathrm{W} \cap \mathrm{W}^\perp = 0$. In fact,we classify subspaces W by two numbers: $d = \dim(\mathrm{W})$ and $\nu = \dim(\mathrm{W} \cap \mathrm{W}^\perp)$. If $\nu = 0$ then $\eta|_{\mathrm{W}}, \mathrm{W}$ is a symplectic space and so we call W a **symplectic subspace** . At the opposite extreme, if $\nu = d$ then W is called a **Lagrangian subspace** . If $\mathrm{W} \subset \mathrm{W}^\perp$ we say that W is an **isotropic subspace**.

A linear transformation between symplectic vector spaces $\ell : \mathrm{V}_1, \eta_1 \to \mathrm{V}_2, \eta_2$ is called a **symplectic linear map** if $\eta_2(\ell(v), \ell(w)) = \eta_1(v, w)$ for all $v, w \in \mathrm{V}_1$; In other words, if $\ell^*\eta_2 = \eta_1$. The set of all symplectic linear isomorphisms from V, η to itself is called the **symplectic group** and denoted $Sp(\mathrm{V},\eta)$. With respect to a symplectic basis $\mathcal{B}$ a symplectic linear isomorphism ℓ is represented by a matrix $A = [\ell]_{\mathcal{B}}$ that satisfies

$$A^t J\, A = J$$

where $J = J_n$ is the matrix defined above and where $2n = \dim(\mathrm{V})$. Such a matrix is called a symplectic matrix and the group of all such is called the **symplectic matrix group** and denoted $Sp(n, \mathbb{R})$. Of course if $\dim(\mathrm{V}) = 2n$ then $Sp(\mathrm{V},\eta) \cong Sp(n, \mathbb{R})$ the isomorphism depending a choice of basis.

If η is a symplectic from on V with $\dim(\mathrm{V}) = 2n$ then $\eta^n \in \wedge^{2n}\mathrm{V}$ is nonzero and so orients the vector space V.

Lemma 15.3. *If $A \in Sp(n, \mathbb{R})$ then* $\det(A) = 1$.

Proof. If we use A as a linear transformation $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ then $A^*\alpha_0 = \alpha_0$ and $A^*\alpha_0^n = \alpha_0^n$ where α_0 is the standard symplectic form on $\mathbb{R}^{2n}$ and $\alpha_0^n \in \wedge^{2n}\mathbb{R}^{2n}$ is top form. Thus $\det A = 1$. □

Theorem 15.4 (Symplectic eigenvalue theorem)**.** *If λ is a (complex) eigenvalue of a symplectic matrix A then so is $1/\lambda$, $\bar{\lambda}$ and $1/\bar{\lambda}$.*

Proof. Let $p(\lambda) = \det(A - \lambda I)$ be the characteristic polynomial. It is easy to see that $J^t = -J$ and $JAJ^{-1} = (A^{-1})^t$. Using these facts we have

$$\begin{aligned} p(\lambda) &= \det(J(A - \lambda I)J^{-1}) = \det(A^{-1} - \lambda I) \\ &= \det(A^{-1}(I - \lambda A)) = \det(I - \lambda A) \\ &= \lambda^{2n}\det(\frac{1}{\lambda}I - A)) = \lambda^{2n}p(1/\lambda). \end{aligned}$$

So we have $p(\lambda) = \lambda^{2n}p(1/\lambda)$. Using this and remembering that 0 is not an eigenvalue one concludes that $1/\lambda$ and $\bar{\lambda}$ are eigenvalues of A. □

Exercise 15.5. With respect to the last theorem, show that λ and $1/\lambda$ have the same multiplicity.

15.2. Canonical Form (Linear case)

Suppose one has a vector space W with dual W*. We denote the pairing between W and W* by $\langle ., . \rangle$. There is a simple way to produce a symplectic form on the space $\mathrm{Z} = \mathrm{W} \times \mathrm{W}^*$ which we will call the **canonical symplectic form**. This is defined by

$$\Omega((v_1, \alpha_1), (v_2, \alpha_2)) := \langle \alpha_2, v_1 \rangle - \langle \alpha_1, v_2 \rangle.$$

If W is an inner product space with inner product $\langle ., . \rangle$ then we may form the canonical symplectic from on $\mathrm{Z} = \mathrm{W} \times \mathrm{W}$ by the same formula. As a special case we get the standard symplectic form on $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ given by

$$\Omega((x, y), (\widetilde{x}, \widetilde{y})) = \widetilde{y} \cdot x - y \cdot \widetilde{x}.$$

15.3. Symplectic manifolds

We have already defined symplectic manifold in the process of giving example. We now take up the proper study of symplectic manifold. We give the definition again:

Definition 15.6. A **symplectic form** on a manifold M is a *nondegenerate closed* 2-form $\omega \in \Omega^2(M) = \Gamma(M, T^*M)$. A **symplectic manifold** is a pair (M, ω) where ω is a symplectic form on M. If there exists a symplectic form on M we say that M has a symplectic structure or admits a symplectic structure.

A map of symplectic manifolds, say $f : (M, \omega) \to (N, \varpi)$ is called a **symplectic map** if and only if $f^*\varpi = \omega$. We will reserve the term **symplectomorphism** to refer to diffeomorphisms that are symplectic maps. Notice that since a symplectic form such as ω is nondegenerate, the $2n$ form $\omega^n = \omega \wedge \cdots \wedge \omega$ is nonzero and global. Hence a symplectic manifold is orientable (more precisely, it is oriented).

Definition 15.7. The form $\Omega_\omega = \frac{(-1)^n}{(2n)!}\omega^n$ is called the **canonical volume form** or **Liouville volume**.

We immediately have that if $f : (M, \omega) \to (M, \omega)$ is a symplectic diffeomorphism then $f^*\Omega_\omega = \Omega_\omega$.

Not every manifold admits a symplectic structure. Of course if M does admit a symplectic structure then it must have even dimension but there are other more subtle obstructions. For example, the fact that $H^2(S^4) = 0$ can be used to show that S^4 does not admit ant symplectic structure. To see this, suppose to the contrary that ω is a closed nondegenerate 2-form on S^4. Then since $H^2(S^4) = 0$ there would be a 1-form θ with $d\theta = \omega$. But then since $d(\omega \wedge \theta) = \omega \wedge \omega$ the 4-form $\omega \wedge \omega$ would be exact also and Stokes' theorem would give $\int_{S^4} \omega \wedge \omega = \int_{S^4} d(\omega \wedge \theta) = \int_{\partial S^4 = \emptyset} \omega \wedge \theta = 0$. But as we have seen $\omega^2 = \omega \wedge \omega$ is a nonzero top form so we must really have $\int_{S^4} \omega \wedge \omega \neq 0$. So in fact, S^4 does not admit a symplectic structure. We will give a more careful examination to the question of obstructions to symplectic structures but let us now list some positive examples.

Example 15.8 (surfaces)**.** Any orientable surface with volume form (area form) qualifies since in this case the volume ω itself is a closed nondegenerate two form.

Example 15.9 (standard)**.** The form $\omega_{can} = \sum_{i=1}^n dx^i \wedge dx^{i+n}$ on $\mathbb{R}^{2n}$ is the prototypical symplectic form for the theory and makes $\mathbb{R}^n$ a symplectic manifold. (See Darboux's theorem 15.22 below)

Example 15.10 (cotangent bundle)**.** We will see in detail below that the cotangent bundle of any smooth manifold has a natural symplectic structure. The symplectic form in a natural bundle chart (q, p) has the form $\omega = \sum_{i=1}^n dq^i \wedge dp_i$. (warning: some authors use $-\sum_{i=1}^n dq^i \wedge dp_i = \sum_{i=1}^n dp_i \wedge dq^i$ instead).

Example 15.11 (complex submanifolds)**.** The symplectic $\mathbb{R}^{2n}$ may be considered the realification of $\mathbb{C}^n$ and then multiplication by i is thought of as a map $J : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$. We have that $\omega_{can}(v, Jv) = -|v|^2$ so that ω_{can} is nondegenerate on any complex submanifold M of $\mathbb{R}^{2n}$ and so $M, \omega_{can}|_M$ is a symplectic manifold.

Example 15.12 (coadjoint orbit)**.** Let G be a Lie group. Define the coadjoint map $\mathrm{Ad}^{\dagger} : G \to GL(\mathfrak{g}^*)$, which takes g to $\mathrm{Ad}^{\dagger}_g$, by

$$\mathrm{Ad}^{\dagger}_g(\xi)(x) = \xi(\mathrm{Ad}_{g^{-1}}(x)).$$

The action defined by $\mathrm{Ad}^{\dagger}$,

$$g \to g \cdot \xi = \mathrm{Ad}^{\dagger}_g(\xi),$$

is called the **coadjoint action**. Then we have an induced map $\mathrm{ad}^{\dagger} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}^*)$ at the Lie algebra level;

$$\mathrm{ad}^{\dagger}(x)(\xi)(y) = -\xi([x, y]).$$

The orbits of the action given by Ad^* are called coadjoint orbits and we will show in theorem below that each orbit is a symplectic manifold in a natural way.

15.4. Complex Structure and Kähler Manifolds

Recall that a complex manifold is a manifold modeled on $\mathbb{C}^n$ and such that the chart overlap functions are all biholomorphic. Every (real) tangent space T_pM of a complex manifold M has a complex structure $J_p : T_pM \to T_pM$ given in biholomorphic coordinates $z = x + iy$ by

$$J_p(\left.\frac{\partial}{\partial x^i}\right|_p) = \left.\frac{\partial}{\partial y^i}\right|_p$$
$$J_p(\left.\frac{\partial}{\partial y^i}\right|_p) = -\left.\frac{\partial}{\partial x^i}\right|_p$$

and for any (biholomorphic) overlap function $\Delta = \varphi \circ \psi^{-1}$ we have $T\Delta \circ J = J \circ T\Delta$.

Definition 15.13. An **almost complex structure** on a smooth manifold M is a bundle map $J : TM \to TM$ covering the identity map such that $J^2 = -\mathrm{id}$. If one can choose an atlas for M such that all the coordinate change functions (overlap functions) Δ satisfy $T\Delta \circ J = J \circ T\Delta$ then J is called a **complex structure** on M.

Definition 15.14. An **almost symplectic structure** on a manifold M is a nondegenerate smooth 2-form ω that is not necessarily closed.

Theorem 15.15. *A smooth manifold M admits an almost complex structure if and only if it admits an almost symplectic structure.*

Proof. First suppose that M has an almost complex structure J and let g be any Riemannian metric on M. Define a quadratic form q_p on each tangent space by

$$q_p(v) = \mathrm{g}_p(v, v) + \mathrm{g}_p(Jv, Jv).$$

Then we have $q_p(Jv) = q_p(v)$. Now let h be the metric obtained from the quadratic form q by polarization. It follows that $\mathrm{h}(v, w) = \mathrm{h}(Jv, Jw)$ for all $v, w \in TM$. Now define a two form ω by

$$\omega(v, w) = \mathrm{h}(v, Jw).$$

This really is skew-symmetric since $\omega(v, w) = \mathrm{h}(v, Jw) = \mathrm{h}(Jv, J^2 w) = -\mathrm{h}(Jv, w) = \omega(w, v)$. Also, ω is nondegenerate since if $v \neq 0$ then $\omega(v, Jv) = \mathrm{h}(v, v) > 0$.

Conversely, let ω be a nondegenerate two form on a manifold M. Once again choose a Riemannian metric g for M. There must be a vector bundle map $\Omega : TM \to TM$ such that

$$\omega(v, w) = \mathrm{g}(\Omega v, w) \text{ for all } v, w \in TM.$$

Since ω is nondegenerate the map Ω must be invertible. Furthermore, since Ω is clearly anti-symmetric with respect to g the map $-\Omega \circ \Omega = -\Omega^2$ must be symmetric and positive definite. From linear algebra applied fiberwise we know that there must be a positive symmetric square root for $-\Omega^2$. Denote this by $P = \sqrt{-\Omega^2}$. Finite dimensional spectral theory also tell us that $P\Omega = \Omega P$. Now let $J = \Omega P^{-1}$ and notice that

$$J^2 = (\Omega P^{-1})(\Omega P^{-1}) = \Omega^2 P^{-2} = -\Omega^2 \Omega^{-2} = -\operatorname{id}.$$

$\square$

One consequence of this result is that there must be characteristic class obstructions to the existence of a symplectic structure on a manifolds. In fact, if (M, ω) is a symplectic manifold then it is certainly almost symplectic and so there is an almost complex structure J on M. The tangent bundle is then a complex vector bundle with J giving the action of multiplication by $\sqrt{-1}$ on each fiber T_pM. Denote the resulting complex vector bundle by TM^J and then consider the total Chern class

$$c(TM^J) = c_n(TM^J) + ... + c_1(TM^J) + 1.$$

Here $c_i(TM^J) \in H^{2i}(M, \mathbb{Z})$. Recall that with the orientation given by ω^n the top class $c_n(TM^J)$ is the Euler class $e(TM)$ of TM. Now for the real bundle TM we have the total Pontrijagin class

$$p(TM) = p_n(TM) + ... + p_1(TM) + 1$$

which are related to the Chern classes by the Whitney sum

$$\begin{aligned} p(TM) &= c(TM^J) \oplus c(TM^{-J}) \\ &= (c_n(TM^J) + ... + c_1(TM^J) + 1)((-1)^n c_n(TM^J) - +... + c_1(TM^J) + 1) \end{aligned}$$

where TM^{-J} is the complex bundle with $-J$ giving the multiplication by $\sqrt{-1}$. We have used the fact that

$$c_i(TM^{-J}) = (-1)^i c_i(TM^J).$$

Now the classes $p_k(TM)$ are invariants of the diffeomorphism class of M an so can be considered constant over all possible choices of J. In fact, from the above relations one can deduce a quadratic relation that must be satisfied:

$$p_k(TM) = c_k(TM^J)^2 - 2c_{k-1}(TM^J)c_{k+1}(TM^J) + \cdots + (-1)^k 2c_{2k}(TM^J).$$

Now this places a restriction on what manifolds might have almost complex structures and hence a restriction on having an almost symplectic structure. Of course some manifolds might have an almost symplectic structure but still have no symplectic structure.

Definition 15.16. A positive definite real bilinear form h on an almost complex manifold (M, J) is will be called Hermitian metric or J-metric if h is J invariant. In this case h is the real part of a Hermitian form on the complex vector bundle TM, J given by

$$\langle v, w\rangle = h(v, w) + ih(Jv, w)$$

Definition 15.17. A diffeomorphism $\phi : (M, J, h) \to (M, J, h)$ is called a Hermitian isometry if and only if $T\phi \circ J = J \circ T\phi$ and

$$h(T\phi v, T\phi w) = h(v, w).$$

A group action $\rho : G \times M \to M$ is called a Hermitian action if $\rho(g, .)$ is a Hermitian isometry for all g. In this case, we have for every $p \in M$ a the representation $d\rho_p : H_p \to \mathrm{Aut}(T_pM, J_p)$ of the isotropy subgroup H_p given by

$$d\rho_p(g)v = T_p\rho_g \cdot v.$$

Definition 15.18. Let M, J be a complex manifold and ω a symplectic structure on M. The manifold is called a **Kähler manifold** if $h(v, w) := \omega(v, Jw)$ is positive definite.

Equivalently we can define a **Kähler manifold as** a complex manifold M, J with Hermitian metric h with the property that the nondegenerate 2-form $\omega(v, w) := h(v, Jw)$ is closed.

Thus we have the following for a Kähler manifold:

(1) A complex structure J,

(2) A J-invariant positive definite bilinear form b,

(3) A Hermitian form $\langle v, w\rangle = h(v, w) + ih(Jv, w)$.

(4) A symplectic form ω with the property that $\omega(v, w) = h(v, Jw)$.

Of course if M, J is a complex manifold with Hermitian metric h then $\omega(v, w) := h(v, Jw)$ automatically gives a nondegenerate 2-form; the question is whether it is closed or not. Mumford's criterion is useful for this purpose:

Theorem 15.19 (Mumford)**.** *Let $\rho : G \times M \to M$ be a smooth Lie group action by Hermitian isometries. For $p \in M$ let H_p be the isometry subgroup of the point p. If $J_p \in d\rho_p(H_p)$ for every p then we have that ω defined by $\omega(v, w) := h(v, Jw)$ is closed.*

Proof. It is easy to see that since ρ preserves both h and J it also preserves ω and $d\omega$. Thus for any given $p \in M$, we have

$$d\omega(d\rho_p(g)u, d\rho_p(g)v, d\rho_p(g)w) = d\omega(u, v, w)$$

for all $g \in H_p$ and all $u, v, w \in T_pM$. By assumption there is a $g_p \in H_p$ with $J_p = d\rho_p(g_p)$. Thus with this choice the previous equation applied twice gives

$$\begin{aligned} d\omega(u, v, w) &= d\omega(J_pu, J_pv, J_pw) \\ &= d\omega(J_p^2u, J_p^2v, J_p^2w) \\ &= d\omega(-u, -v, -w) = -d\omega(u, v, w) \end{aligned}$$

so $d\omega = 0$ at p which was an arbitrary point so $d\omega = 0$. □

Since a Kähler manifold is a posteriori a Riemannian manifold it has associated with it the Levi-Civita connection ∇. In the following we view J as an element of $\mathfrak{X}(M)$.

Theorem 15.20. *For a Kähler manifold M, J, h with associated symplectic form ω we have that*

$$d\omega = 0 \quad \textit{if and only if} \quad \nabla J = 0.$$

15.5. Symplectic musical isomorphisms

Since a symplectic form ω on a manifold M is nondegenerate we have a map

$$\omega_\flat : TM \to T^*M$$

given by $\omega_\flat(X_p)(v_p) = \omega(X_p, v_p)$ and the inverse $\omega^\sharp$ is such that

$$\iota_{\omega^\sharp(\alpha)}\omega = \alpha$$

or

$$\omega(\omega^\sharp(\alpha_p), v_p) = \alpha_p(v_p)$$

Let check that $\omega^\sharp$ really is the inverse. (one could easily be off by a sign in this business.) We have

$$\begin{aligned}\omega_\flat(\omega^\sharp(\alpha_p))(v_p) &= \omega(\omega^\sharp(\alpha_p), v_p) = \alpha_p(v_p) \text{ for all } v_p \\ &\Longrightarrow \omega_\flat(\omega^\sharp(\alpha_p)) = \alpha_p.\end{aligned}$$

Notice that $\omega^\sharp$ induces a map on sections also denoted by $\omega^\sharp$ with inverse $\omega_\flat : \mathfrak{X}(M) \to \mathfrak{X}^*(M)$.

Notation 15.21. Let us abbreviate $\omega^\sharp(\alpha)$ to $\sharp\alpha$ and $\omega_\flat(v)$ to $\flat v$.

15.6. Darboux's Theorem

Lemma 15.22 (Darboux's theorem)**.** *On a* $2n$*- manifold* (M,ω) *with a closed 2-form* ω *with* $\omega^n \neq 0$ *(for instance if* (M,ω) *is symplectic) there exists a sub-atlas consisting of charts called symplectic charts (canonical coordinates) characterized by the property that the expression for* ω *in such a chart is*

$$\omega_U = \sum_{i=1}^n dx^i \wedge dx^{i+n}$$

and so in particular M *must have even dimension* $2n$.

Remark 15.23. Let us agree that the canonical coordinates can be written (x^i, y_i) instead of (x^i, x^{i+n}) when convenient.

Remark 15.24. It should be noticed that if x^i, y_i is a symplectic chart then $\sharp dx^i$ must be such that

$$\sum_{r=1}^n dx^r \wedge dy^r(\sharp dx^i, \frac{\partial}{\partial x^j}) = \delta^i_j$$

but also

$$\begin{aligned}\sum_{r=1}^n dx^r \wedge dy^r(\sharp dx^i, \frac{\partial}{\partial x^j}) &= \sum_{r=1}^n \left(dx^r(\sharp dx) dy^r(\frac{\partial}{\partial x^j}) - dy^r(\sharp dx^i) dx^r(\frac{\partial}{\partial x^j}) \right) \\ &= -dy^j(\sharp dx^i)\end{aligned}$$

and so we conclude that $\sharp dx^i = -\frac{\partial}{\partial y^i}$ and similarly $\sharp dy^i = \frac{\partial}{\partial x^i}$.

Proof. We will use induction and follow closely the presentation in [**?**]. Assume the theorem is true for symplectic manifolds of dimension $2(n-1)$. Let $p \in M$. Choose a function y^1 on some open neighborhood of p such that $dy_1(p) \neq 0$. Let $X = \sharp dy_1$ and then X will not vanish at p. We can then choose another function x^1 such that $Xx^1 = 1$ and we let $Y = -\sharp dx^1$. Now since $d\omega = 0$ we can use Cartan's formula to get

$$\mathcal{L}_X \omega = \mathcal{L}_Y \omega = 0.$$

In the following we use the notation $\langle X, \omega\rangle = \iota_X \omega$ (see notation **??**). Contract ω with the bracket of X and Y :

$$\begin{aligned}\langle [X,Y], \omega\rangle &= \langle \mathcal{L}_X Y, \omega\rangle = \mathcal{L}_X \langle Y, \omega\rangle - \langle Y, \mathcal{L}_X \omega\rangle \\ &= \mathcal{L}_X(-dx^1) = -d(X(x^1)) = -d1 = 0.\end{aligned}$$

Now since ω is nondegenerate this implies that $[X,Y] = 0$ and so there must be a local coordinate system $(x^1, y_1, w^1, ..., w^{2n-2})$ with

$$\begin{aligned}\frac{\partial}{\partial y_1} &= Y \\ \frac{\partial}{\partial x^1} &= X.\end{aligned}$$

In particular, the theorem is true if $n = 1$. Assume the theorem is true for symplectic manifolds of dimension $2(n-1)$. If we let $\omega' = \omega - dx^1 \wedge dy_1$ then since $d\omega' = 0$ and hence

$$\langle X, \omega'\rangle = \mathcal{L}_X \omega' = \langle Y, \omega'\rangle = \mathcal{L}_Y \omega' = 0$$

we conclude that ω' can be expressed as a 2-form in the $w^1, ..., w^{2n-2}$ variables alone. Furthermore,

$$\begin{aligned}0 \neq \omega^n &= (\omega - dx^1 \wedge dy_1)^n \\ &= \pm n dx^1 \wedge dy_1 \wedge (\omega')^n\end{aligned}$$

from which it follows that ω' is the pull-back of a form nondegenerate form ϖ on $\mathbb{R}^{2n-2}$. To be exact if we let the coordinate chart given by $(x^1, y_1, w^1, ..., w^{2n-2})$ by denoted by ψ and let pr be the projection $\mathbb{R}^{2n} = \mathbb{R}^2 \times \mathbb{R}^{2n-1} \to \mathbb{R}^{2n-1}$ then $\omega' = (pr \circ \psi)^* \varpi$. Thus the induction hypothesis says that ω' has the form $\omega' = \sum_{i=2}^n dx^i \wedge dy_i$ for some functions x^i, y_i with $i = 2, ..., n$. It is easy to see that the construction implies that in some neighborhood of p the full set of functions x^i, y_i with $i = 1, ..., n$ form the desired symplectic chart. □

An atlas $\mathcal{A}$ of symplectic charts is called a symplectic atlas. A chart (U, φ) is called compatible with the symplectic atlas $\mathcal{A}$ if for every $(\psi_\alpha, U_\alpha) \in \mathcal{A}$ we have

$$(\varphi \circ \psi^{-1})^* \omega_0 = \omega_0$$

for the canonical symplectic $\omega_{can} = \sum_{i=1}^n du^i \wedge du^{i+n}$ defined on $\psi_\alpha(U \cap U_\alpha) \subset \mathbb{R}^{2n}$ using standard rectangular coordinates u^i.

15.7. Poisson Brackets and Hamiltonian vector fields

Definition 15.25 (on forms)**.** The **Poisson bracket** of two 1-forms is defined to be

$$\{\alpha, \beta\}_\pm = \mp \flat[\sharp\alpha, \sharp\beta]$$

where the musical symbols refer to the maps $\omega^\sharp$ and $\omega_\flat$. This puts a Lie algebra structure on the space of 1-forms $\Omega^1(M) = \mathfrak{X}^*(M)$.

Definition 15.26 (on functions)**.** The **Poisson bracket** of two smooth functions is defined to be

$$\{f, g\}_\pm = \pm\omega(\sharp df, \sharp dg) = \pm\omega(X_f, X_g)$$

This puts a Lie algebra structure on the space $\mathcal{F}(M)$ of smooth function on the symplectic M. It is easily seen (using $dg = \iota_{X_g}\omega$) that $\{f, g\}_\pm = \pm L_{X_g} f = \mp L_{X_f} g$ which shows that $f \mapsto \{f, g\}$ is a derivation for fixed g. The connection between the two Poisson brackets is

$$d\{f, g\}_\pm = \{df, dg\}_\pm.$$

Let us take canonical coordinates so that $\omega = \sum_{i=1}^n dx^i \wedge dy_i$. If $X_p = \sum_{i=1}^n dx^i(X)\frac{\partial}{\partial x^i} + \sum_{i=1}^n dy_i(X)\frac{\partial}{\partial y_i}$ and $v_p = dx^i(v_p)\frac{\partial}{\partial x^i} + dy_i(v_p)\frac{\partial}{\partial y_i}$ then using the Einstein summation convention we have

$$\begin{aligned}
&\omega_\flat(X)(v_p) \\
&= \omega(dx^i(X)\frac{\partial}{\partial x^i} + dy_i(X)\frac{\partial}{\partial y_i}, dx^i(v_p)\frac{\partial}{\partial x^i} + dy_i(v_p)\frac{\partial}{\partial y_i}) \\
&= (dx^i(X)dy_i - dy_i(X)dx^i)(v_p)
\end{aligned}$$

so we have

Lemma 15.27. $\omega_\flat(X_p) = \sum_{i=1}^n dx^i(X)dy_i - dy_i(X)dx^i = \sum_{i=1}^n(-dy_i(X)dx^i + dx^i(X)dy_i)$

Corollary 15.28. If $\alpha = \sum_{i=1}^n \alpha(\frac{\partial}{\partial x^i})dx^i + \sum_{i=1}^n \alpha(\frac{\partial}{\partial y_i})dy^i$ then $\omega^\sharp(\alpha) = \sum_{i=1}^n \alpha(\frac{\partial}{\partial y_i})\frac{\partial}{\partial x^i} - \sum_{i=1}^n \alpha(\frac{\partial}{\partial x^i})\frac{\partial}{\partial y_i}$

An now for the local formula:

Corollary 15.29. $\{f, g\} = \sum_{i=1}^n(\frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial x^i})$

Proof. $df = \frac{\partial f}{\partial x^i}dx^i + \frac{\partial f}{\partial y_i}dy_i$ and $dg = \frac{\partial g}{\partial x^j}dx^j + \frac{\partial g}{\partial y_i}dy_i$ so $\sharp df = \frac{\partial f}{\partial y_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y_i}$ and similarly for dg. Thus (using the summation convention again);

$$\begin{aligned}
\{f, g\} &= \omega(\sharp df, \sharp dg) \\
&= \omega(\frac{\partial f}{\partial y_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y_i}, \frac{\partial g}{\partial y_i}\frac{\partial}{\partial x^j} - \frac{\partial g}{\partial x^j}\frac{\partial}{\partial y_i}) \\
&= \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial x^i}
\end{aligned}$$

$\square$

A main point about Poison Brackets is

Theorem 15.30. *f is constant along the orbits of X_g if and only if $\{f,g\} = 0$. In fact,*

$$\tfrac{d}{dt} g \circ \varphi_t^{X_f} = 0 \iff \{f,g\} = 0 \iff \tfrac{d}{dt} f \circ \varphi_t^{X_g} = 0$$

Proof. $\frac{d}{dt} g \circ \varphi_t^{X_f} = (\varphi_t^{X_f})^* L_{X_f} g = (\varphi_t^{X_f})^* \{f,g\}$. Also use $\{f,g\} = -\{g,f\}$. □

The equations of motion for a Hamiltonian H are

$$\frac{d}{dt} f \circ \varphi_t^{X_H} = \pm\{f \circ \varphi_t^{X_H}, H\}_\pm = \mp\{H, f \circ \varphi_t^{X_H}\}_\pm$$

which is true by the following simple computation

$$\begin{aligned} \frac{d}{dt} f \circ \varphi_t^{X_H} &= \frac{d}{dt} (\varphi_t^{X_H})^* f = (\varphi_t^{X_H})^* L_{X_H} f \\ &= L_{X_H}(f \circ \varphi_t^{X_H}) = \{f \circ \varphi_t^{X_H}, H\}_\pm. \end{aligned}$$

Notation 15.31. ¿From now on we will use only $\{.,.\}_+$ unless otherwise indicated and shall write $\{.,.\}$ for $\{.,.\}_+$.

Definition 15.32. A Hamiltonian system is a triple (M, ω, H) where M is a smooth manifold, ω is a symplectic form and H is a smooth function $H : M \to \mathbb{R}$.

The main example, at least from the point of view of mechanics, is the cotangent bundle of a manifold which is discussed below. From a mechanical point of view the Hamiltonian function controls the dynamics and so is special.

Let us return to the general case of a symplectic manifold M, ω

Definition 15.33. Now if $H : M \to \mathbb{R}$ is smooth then we define the **Hamiltonian vector field** X_H with energy function H to be $\omega^\sharp dH$ so that by definition $\iota_{X_H} \omega = dH$.

Definition 15.34. A vector field X on M, ω is called a **locally Hamiltonian vector field** or a **symplectic vector field** if and only if $L_X \omega = 0$.

If a symplectic vector field is complete then we have that $(\varphi_t^X)^* \omega$ is defined for all $t \in \mathbb{R}$. Otherwise, for any relatively compact open set U the restriction φ_t^X to U is well defined for all $t \leq b(U)$ for some number depending only on U. Thus $(\varphi_t^X)^* \omega$ is defined on U for $t \leq b(U)$. Since U can be chosen to contain any point of interest and since M can be covered by relatively compact sets, it will be of little harm to write $(\varphi_t^X)^* \omega$ even in the case that X is not complete.

Lemma 15.35. *The following are equivalent:*

(1) *X is* ***symplectic vector field****, i.e.* $L_X\omega = 0$
(2) $\iota_X\omega$ *is closed*
(3) $(\varphi_t^X)^*\omega = \omega$
(4) *X is locally a Hamiltonian vector field.*

Proof. (1)$\Longleftrightarrow$ (4) by the Poincaré lemma. Next, notice that $L_X\omega = d \circ \iota_X\omega + \iota_X \circ d\omega = d \circ \iota_X\omega$ so we have (2)$\Longleftrightarrow$(1). The implication (2)$\Longleftrightarrow$(3) follows from Theorem **??**. $\square$

Proposition 15.36. We have the following easily deduced facts concerning Hamiltonian vector fields:

(1) The H is constant along integral curves of X_H
(2) The flow of X_H is a local symplectomorphism. That is $\varphi_t^{X_H}{}^*\omega = \omega$

Notation 15.37. Denote the set of all Hamiltonian vector fields on M, ω by $\mathcal{H}(\omega)$ and the set of all symplectic vector fields by $\mathcal{SP}(\omega)$

Proposition 15.38. The set $\mathcal{SP}(\omega)$ is a Lie subalgebra of $\mathfrak{X}(M)$. In fact, we have $[\mathcal{SP}(\omega), \mathcal{SP}(\omega)] \subset \mathcal{H}(\omega) \subset \mathfrak{X}(M)$.

Proof. Let $X, Y \in \mathcal{SP}(\omega)$. Then

$$\begin{aligned}
[X,Y] \lrcorner \omega &= \mathcal{L}_X Y \lrcorner \omega = \mathcal{L}_X(Y \lrcorner \omega) - Y \lrcorner \mathcal{L}_X\omega \\
&= d(X \lrcorner Y \lrcorner \omega) + X \lrcorner d(Y \lrcorner \omega) - 0 \\
&= d(X \lrcorner Y \lrcorner \omega) + 0 + 0 \\
&= -d(\omega(X,Y)) = -X_{\omega(X,Y)} \lrcorner \omega
\end{aligned}$$

and since ω in nondegenerate we have $[X,Y] = X_{-\omega(X,Y)} \in \mathcal{H}(\omega)$. $\square$

15.8. Configuration space and Phase space

Consider the cotangent bundle of a manifold Q with projection map

$$\pi : T^*Q \to Q$$

and define the **canonical 1-form** $\theta \in T^*(T^*Q)$ by

$$\theta : v_{\alpha_p} \mapsto \alpha_p(T\pi \cdot v_{\alpha_p})$$

where $\alpha_p \in T_p^*Q$ and $v_{\alpha_p} \in T_{\alpha_p}(T_p^*Q)$. In local coordinates this reads

$$\theta_0 = \sum p_i dq^i.$$

Then $\omega_{T^*Q} = -d\theta$ is a symplectic form that in natural coordinates reads

$$\omega_{T^*Q} = \sum dq^i \wedge dp_i$$

Lemma 15.39. *θ is the unique 1-form such that for any $\beta \in \Omega^1(Q)$ we have*

$$\beta^*\theta = \beta$$

where we view β as $\beta : Q \to T^*Q$.

Proof: $\beta^*\theta(v_q) = \theta|_{\beta(q)}(T\beta \cdot v_q) = \beta(q)(T\pi \circ T\beta \cdot v_q) = \beta(q)(v_q)$ since $T\pi \circ T\beta = T(\pi \circ \beta) = T(\mathrm{id}) = \mathrm{id}$.

The cotangent lift T^*f of a diffeomorphism $f : Q_1 \to Q_2$ is defined by the commutative diagram

$$\begin{array}{ccc} T^*Q_1 & \overset{T^*f}{\longleftarrow} & T^*Q_2 \\ \downarrow & & \downarrow \\ Q_1 & \overset{f}{\rightarrow} & Q_2 \end{array}$$

and is a symplectic map; i.e. $(T^*f)^*\,\omega_0 = \omega_0$. In fact, we even have $(T^*f)^*\,\theta_0 = \theta_0$.

The triple (T^*Q, ω_{T^*Q}, H) is a Hamiltonian system for any choice of smooth function. The most common form for H in this case is $\frac{1}{2}K+V$ where K is a Riemannian metric that is constructed using the mass distribution of the bodies modeled by the system and V is a smooth potential function which, in a conservative system, depends only on $\mathbf{q}$ when viewed in natural cotangent bundle coordinates q^i, p_i.

Now we have $\sharp dg = \frac{\partial g}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial}{\partial p_i}$ and introducing the $\pm$ notation one more time we have

$$\begin{aligned}\{f,g\}_\pm &= \pm\omega_{T^*Q}(\sharp df, \sharp dg) = \pm df(\sharp dg) = \pm df(\frac{\partial g}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial}{\partial p_i}) \\ &= \pm(\frac{\partial g}{\partial p_i}\frac{\partial f}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial f}{\partial p_i}) \\ &= \pm(\frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i})\end{aligned}$$

Thus letting

$$\varphi_t^{X_H}(q_0^1, ..., q_0^n, p_0^1, ..., p_0^n) = (q^1(t), ..., q^n(t), p^1(t), ..., p^n(t))$$

the equations of motions read

$$\begin{aligned}\frac{d}{dt}f(q(t), p(t)) &= \frac{d}{dt}f \circ \varphi_t^{X_H} = \{f \circ \varphi_t^{X_H}, H\} \\ &= \frac{\partial f}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q^i}.\end{aligned}$$

Where we have abbreviated $f \circ \varphi_t^{X_H}$ to just f. In particular, if $f = q^i$ and $f = p_i$ then

$$\dot{q}^i(t) = \frac{\partial H}{\partial p_i}$$
$$\dot{p}_i(t) = -\frac{\partial H}{\partial q^i}$$

which should be familiar.

15.9. Transfer of symplectic structure to the Tangent bundle

15.9.0.1. *Case I: a (pseudo) Riemannian manifold.* If Q, g is a (pseudo) Riemannian manifold then we have a map $\mathrm{g}^\flat : TQ \to T^*Q$ defined by

$$\mathrm{g}^\flat(v)(w) = \mathrm{g}(v, w)$$

and using this we can define a symplectic form ϖ_0 on TQ by

$$\varpi_0 = \left(\mathrm{g}^\flat\right)^* \omega$$

(Note that $d\varpi_0 = d(\mathrm{g}^{\flat *}\omega) = \mathrm{g}^{\flat *} d\omega = 0$.) In fact, ϖ_0 is exact since ω is exact:

$$\begin{aligned} \varpi_0 &= \left(\mathrm{g}^\flat\right)^* \omega \\ &= \left(\mathrm{g}^\flat\right)^* d\theta = d\left(\mathrm{g}^{\flat *}\theta\right). \end{aligned}$$

Let us write $\Theta_0 = \mathrm{g}^{\flat *}\theta$. Locally we have

$$\begin{aligned} \Theta_0(x, v)(v_1, v_2) &= \mathrm{g}_x(v, v_1) \text{ or} \\ \Theta_0 &= \sum \mathrm{g}_{ij} \dot{q}^i dq^j \end{aligned}$$

and also

$$\begin{aligned} &\varpi_0(x, v)((v_1, v_2), ((w_1, w_2))) \\ &= \mathrm{g}_x(w_2, v_1) - \mathrm{g}_x(v_2, w_1) + D_x\mathrm{g}_x(v, v_1) \cdot w_1 - D_x\mathrm{g}_x(v, w_1) \cdot v_1 \end{aligned}$$

which in classical notation (and for finite dimensions) looks like

$$\varpi_h = \mathrm{g}_{ij} dq^i \wedge d\dot{q}^j + \sum \frac{\partial \mathrm{g}_{ij}}{\partial q^k} \dot{q}^i dq^j \wedge dq^k$$

15.9.0.2. *Case II: Transfer of symplectic structure by a Lagrangian function.*

Definition 15.40. Let $L : TQ \to Q$ be a Lagrangian on a manifold Q. We say that L is **regular** or **non-degenerate** at $\xi \in TQ$ if in any canonical coordinate system $(q, \dot{q})$ whose domain contains ξ, the matrix

$$\left[\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}(q(\xi), \dot{q}(\xi))\right]$$

is non-degenerate. L is called **regular** or **nondegenerate** if it is regular at all points in TQ.

We will need the following general concept:

Definition 15.41. Let $\pi_E : E \to M$ and $\pi_F : F \to M$ be two vector bundles. A map $L : E \to F$ is called a **fiber preserving map** if the following diagram commutes:

$$\begin{array}{ccccc} E & & \stackrel{L}{\to} & & F \\ & & & & \\ \pi_E & \searrow & & \swarrow & \pi_F \\ & & M & & \end{array}$$

We do not require that the map L be linear on the fibers and so in general L is not a vector bundle morphism.

Definition 15.42. If $L : E \to F$ is a fiber preserving map then if we denote the restriction of L to a fiber E_p by L_p define the **fiber derivative**

$$\mathbf{F}L : E \to \mathrm{Hom}(E, F)$$

by $\mathbf{F}L : e_p \mapsto Df|_p(e_p)$ for $e_p \in E_p$.

In our application of this concept, we take F to be the trivial bundle $Q \times \mathbb{R}$ over Q so $\mathrm{Hom}(E, F) = \mathrm{Hom}(E, \mathbb{R}) = T^*Q$.

Lemma 15.43. *A Lagrangian function $L : TQ \to \mathbb{R}$ gives rise to a fiber derivative $\mathbf{F}L : TQ \to T^*Q$. The Lagrangian is nondegenerate if and only if $\mathbf{F}L$ is a diffeomorphism.*

Definition 15.44. The form ϖ_L is defined by

$$\varpi_L = (\mathbf{F}L)^* \omega$$

Lemma 15.45. *ω_L is a symplectic form on TQ if and only if L is nondegenerate (i.e. if $\mathbf{F}L$ is a diffeomorphism).*

Observe that we can also define $\theta_L = (\mathbf{F}L)^* \theta$ so that $d\theta_L = d(\mathbf{F}L)^* \theta = (\mathbf{F}L)^* d\theta = (\mathbf{F}L)^* \omega = \varpi_L$ so we see that ω_L is exact (and hence closed a required for a symplectic form).

Now in natural coordinates we have

$$\varpi_L = \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} dq^i \wedge dq^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} dq^i \wedge d\dot{q}^j$$

as can be verified using direct calculation.

The following connection between the transferred forms ϖ_L and ϖ_0 and occasionally not pointed out in some texts.

Theorem 15.46. *Let V be a smooth function on a Riemannian manifold M, h. If we define a Lagrangian by $L = \frac{1}{2}h - V$ then the Legendre transformation $\mathbf{F}L :: TQ \to T^*Q$ is just the map $g^\flat$ and hence $\varpi_L = \varpi_h$.*

Proof. We work locally. Then the Legendre transformation is given by

$$\begin{aligned} q^i &\mapsto q^i \\ \dot{q}^i &\mapsto \tfrac{\partial L}{\partial \dot{q}^i} \end{aligned} .$$

But since $L(\dot{\mathbf{q}}, \dot{\mathbf{q}}) = \frac{1}{2}\mathrm{g}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) - V(q)$ we have $\frac{\partial L}{\partial \dot{q}^i} = \frac{\partial}{\partial \dot{q}^i}\frac{1}{2}\mathrm{g}_{kl}\dot{q}^l\dot{q}^k = \mathrm{g}_{il}\dot{q}^l$ which together with $q^i \mapsto q^i$ is the coordinate expression for $\mathrm{g}^\flat$:

$$\begin{aligned} q^i &\mapsto q^i \\ \dot{q}^i &\mapsto \mathrm{g}_{il}\dot{q}^l \end{aligned}$$

□

15.10. Coadjoint Orbits

Let G be a Lie group and consider $\mathrm{Ad}^\dagger : G \to GL(\mathfrak{g}^*)$ and the corresponding coadjoint action as in example 15.12. For every $\xi \in \mathfrak{g}^*$ we have a Left invariant 1-form on G defined by

$$\theta^\xi = \xi \circ \omega_G$$

where ω_G is the canonical $\mathfrak{g}$-valued 1-form (the Maurer-Cartan form). Let the G_ξ be the isotropy subgroup of G for a point $\xi \in \mathfrak{g}^*$ under the coadjoint action. Then it is standard that orbit $G \cdot \xi$ is canonically diffeomorphic to the orbit space G/G_ξ and the map $\phi_\xi : g \mapsto g \cdot \xi$ is a submersion onto . Then we have

Theorem 15.47. *There is a unique symplectic form Ω^ξ on $G/G_\xi \cong G \cdot \xi$ such that $\phi_\xi^* \Omega^\xi = d\theta^\xi$.*

Proof: If such a form as Ω^ξ exists as stated then we must have

$$\Omega^\xi(T\phi_\xi.v, T\phi_\xi.w) = d\theta^\xi(v, w) \text{ for all } v, w \in T_g G$$

We will show that this in fact *defines* Ω^ξ as a symplectic form on the orbit $G \cdot \xi$. First of all notice that by the structure equations for the Maurer-Cartan form we have for $v, w \in T_eG = \mathfrak{g}$

$$\begin{aligned} d\theta^\xi(v, w) &= \xi(d\omega_G(v, w)) = \xi(\omega_G([v, w])) \\ &= \xi(-[v, w]) = \mathrm{ad}^\dagger(v)(\xi)(w) \end{aligned}$$

From this we see that

$$\mathrm{ad}^\dagger(v)(\xi) = 0 \Longleftrightarrow v \in \mathrm{Null}(\left. d\theta^\xi \right|_e)$$

where $\mathrm{Null}(\left. d\theta^\xi \right|_e) = \{v \in \mathfrak{g} : \left. d\theta^\xi \right|_e (v, w) \text{ for all } w \in \mathfrak{g}\}$. On the other hand, $G_\xi = \ker\{g \longmapsto \mathrm{Ad}_g^\dagger(\xi)\}$ so $\mathrm{ad}^\dagger(v)(\xi) = 0$ if and only if $v \in T_eG_\xi = \mathfrak{g}_\xi$.

Now notice that since $d\theta^\xi$ is left invariant we have that $\mathrm{Null}(\left. d\theta^\xi \right|_g) = TL_g(\mathfrak{g}_\xi)$ which is the tangent space to the coset gG_ξ and which is also $\ker \left. T\phi_\xi \right|_g$. Thus we conclude that

$$\mathrm{Null}(\left. d\theta^\xi \right|_g) = \ker \left. T\phi_\xi \right|_g .$$

It follows that we have a natural isomorphism

$$T_{g\cdot\xi}(G \cdot \xi) = \left. T\phi_\xi \right|_g (T_gG) \approx T_gG/(TL_g(\mathfrak{g}_\xi))$$

Another view: Let the vector field on $G \cdot \xi$ corresponding to $v, w \in \mathfrak{g}$ generated by the action be denoted by $v^\dagger$and $w^\dagger$. Then we have $\Omega^\xi(\xi)(v^\dagger, w^\dagger) := \xi(-[v, w])$ at $\xi \in G \cdot \xi$ and then extend to the rest of the points of the orbit by equivariance:

$$\Omega^\xi(g \cdot \xi)(v^\dagger, w^\dagger) = \overset{\dagger}{\underset{g}{\mathrm{Ad}}}(\xi(-[v, w]))$$

15.11. The Rigid Body

In what follows we will describe the rigid body rotating about one of its points in three different versions. The basic idea is that we can represent the configuration space as a subset of $\mathbb{R}^{3N}$ with a very natural kinetic energy function. But this space is also isomorphic to the rotation group $SO(3)$ and we can transfer the kinetic energy metric over to $SO(3)$ and then the evolution of the system is given by geodesics in $SO(3)$ with respect to this metric. Next we take advantage of the fact that the tangent bundle of $SO(3)$ is trivial to transfer the setup over to a trivial bundle. But there are two natural ways to do this and we explore the relation between the two.

15.11.1. The configuration in $\mathbb{R}^{3N}$. Let us consider a rigid body to consist of a set of point masses located in $\mathbb{R}^3$ at points with position vectors $\mathbf{r}_1(t), ...\mathbf{r}_N(t)$ at time t. Thus $\mathbf{r}_i = (x_1, x_2, x_3)$ is the coordinates of the i-th point mass. Let $m_1, ..., m_N$ denote the masses of the particles. To say that this set of point masses is rigid is to say that the distances $|\mathbf{r}_i - \mathbf{r}_j|$ are constant for each choice of i and j. Let us assume for simplicity that the body is in a state of uniform rectilinear motion so that by re-choosing our coordinate axes if necessary we can assume that the there is one of the point masses at the origin of our coordinate system at all times. Now the set of all possible configurations is some submanifold of $\mathbb{R}^{3N}$ which we denote by M. Let us also assume that at least 3 of the masses, say those located at $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2$ are situated so that the position vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2$ form a basis of $\mathbb{R}^3$. For convenience let $\mathbf{r}$ and $\dot{\mathbf{r}}$ be abbreviations for $(\mathbf{r}_1(t), ..., \mathbf{r}_N(t))$ and $(\dot{\mathbf{r}}_1(t), ..., \dot{\mathbf{r}}_N(t))$.The correct kinetic energy for the system of particles forming the rigid body is $\frac{1}{2}K(\dot{\mathbf{r}}, \dot{\mathbf{r}})$ where the kinetic energy metric K is

$$K(\mathbf{v}, \mathbf{w}) = m_1 \mathbf{v}_1 \cdot \mathbf{w}_1 + \cdots + m_N \mathbf{v}_N \cdot \mathbf{w}_N.$$

Since there are no other forces on the body other than those that constrain the body to be rigid the Lagrangian for M is just $\frac{1}{2}K(\dot{\mathbf{r}}, \dot{\mathbf{r}})$ and the evolution of the point in M representing the body is a geodesic when we use as Hamiltonian K and the symplectic form pulled over from T^*M as described previously.

15.11.2. Modelling the rigid body on $S\mathrm{O}(3)$. Let $\mathbf{r}_1(0), ...\mathbf{r}(0)_N$ denote the initial positions of our point masses. Under these condition there is a unique matrix valued function $g(t)$ with values in $S\mathrm{O}(3)$ such that $\mathbf{r}_i(t) = g(t)\mathbf{r}_i(0)$. Thus the motion of the body is determined by the curve in $S\mathrm{O}(3)$ given by $t \mapsto g(t)$. In fact, we can map $S\mathrm{O}(3)$ to the set of all possible configurations of the points making up the body in a 1-1 manner by letting $\mathbf{r}_1(0) = \xi_1, ...\mathbf{r}(0)_N = \xi_N$ and mapping $\Phi : g \mapsto (g\xi_1, ..., g\xi_N) \in M \subset \mathbb{R}^{3N}$. If we use the map Φ to transfer this over to $TS\mathrm{O}(3)$ we get

$$k(\xi, \upsilon) = K(T\Phi \cdot \xi, T\Phi \cdot \upsilon)$$

for $\xi, \upsilon \in TS\mathrm{O}(3)$. Now k is a Riemannian metric on $S\mathrm{O}(3)$ and in fact, k is a left invariant metric:

$$k(\xi, \upsilon) = k(TL_g\xi, TL_g\upsilon) \text{ for all } \xi, \upsilon \in TS\mathrm{O}(3).$$

Exercise 15.48. Show that k really is left invariant. Hint: Consider the map $\mu_{g_0} : (\mathbf{v_1}, \cdots, \mathbf{v_N}) \mapsto (g_0\mathbf{v_1}, \cdots, g_0\mathbf{v_N})$ for $g_0 \in S\mathrm{O}(3)$ and notice that $\mu_{g_0} \circ \Phi = \Phi \circ L_{g_0}$ and hence $T\mu_{g_0} \circ T\Phi = T\Phi \circ TL_{g_0}$.

Now by construction, the Riemannian manifolds M, K and $S\mathrm{O}(3), \mathrm{k}$ are isometric. Thus the corresponding path $g(t)$ in $S\mathrm{O}(3)$ is a geodesic with respect to the left invariant metric k. Our Hamiltonian system is now $(TSO(3), \Omega_k, k)$ where Ω_k is the Legendre transformation of the canonical symplectic form Ω on $T^*SO(3)$

15.11.3. The trivial bundle picture. Recall that we the Lie algebra of $SO(3)$ is the vector space of skew-symmetric matrices $\mathfrak{su}(3)$. We have the two trivializations of the tangent bundle $T\mathrm{SO}(3)$ given by

$$\mathrm{triv}_L(v_g) = (g, \omega_G(v_g)) = (g, g^{-1}v_g)$$
$$\mathrm{triv}_R(v_g) = (g, \omega_G(v_g)) = (g, v_g g^{-1})$$

with inverse maps $\mathrm{SO}(3) \times \mathfrak{so}(3) \to T\mathrm{SO}(3)$ given by

$$(g, B) \mapsto TL_g B$$
$$(g, B) \mapsto TR_g B$$

Now we should be able to represent the system in the trivial bundle $\mathrm{SO}(3) \times \mathfrak{so}(3)$ via the map $\mathrm{triv}_L(v_g) = (g, \omega_G(v_g)) = (g, g^{-1}v_g)$. Thus we let k_0 be the metric on $\mathrm{SO}(3) \times \mathfrak{so}(3)$ coming from the metric k. Thus by definition

$$\mathrm{k}_0((g, v), (g, w)) = \mathrm{k}(TL_g v, TL_g w) = \mathrm{k}_e(v, w)$$

where $v, w \in \mathfrak{so}(3)$ are skew-symmetric matrices.

15.12. The momentum map and Hamiltonian actions

Remark 15.49. In this section all Lie groups will be assumed to be connected.

Suppose that (a connected Lie group) G acts on M, ω as a group of symplectomorphisms.

$$\sigma : G \times M \to M$$

Then we say that σ is a **symplectic G-action** . Since G acts on M we have for every $v \in \mathfrak{g}$ the fundamental vector field $X^v = v^\sigma$. The fundamental vector field will be symplectic (locally Hamiltonian). Thus every one-parameter group g^t of G induces a symplectic vector field on M. Actually, it is only the infinitesimal action that matters at first so we define

Definition 15.50. Let M be a smooth manifold and let $\mathfrak{g}$ be the Lie algebra of a connected Lie group G. A linear map $\sigma' : v \mapsto X^v$ from $\mathfrak{g}$ into $\mathfrak{X}(M)$ is called a **$\mathfrak{g}$-action** if

$$[X^v, X^w] = -X^{[v,w]} \text{ or}$$
$$[\sigma'(v), \sigma'(w)] = -\sigma'([v, w]).$$

If M, ω is symplectic and the $\mathfrak{g}$-action is such that $\mathcal{L}_{X^v}\omega = 0$ for all $v \in \mathfrak{g}$ we say that the action is a **symplectic g-action**.

Definition 15.51. Every symplectic action $\sigma : G \times M \to M$ induces a $\mathfrak{g}$-action $d\sigma$ via

$$d\sigma : v \mapsto X^v$$

$$\text{where } X^v(x) = \left.\frac{d}{dt}\right|_0 \sigma(\exp(tv), x).$$

In some cases, we may be able to show that for all v the symplectic field X^v is a full fledged Hamiltonian vector field. In this case associated to each $v \in \mathfrak{g}$ there is a Hamiltonian function $J_v = J_{X^v}$ with corresponding Hamiltonian vector field equal to X^v and J_v is determined up to a constant by $X^v = \sharp dJ_{X^v}$. Now $\iota_{X^v}\omega$ is always closed since $d\iota_{X^v}\omega = \mathcal{L}_{X^v}\omega$. When is it possible to define J_v for every $v \in \mathfrak{g}$?

Lemma 15.52. *Given a symplectic $\mathfrak{g}$-action $\sigma' : v \mapsto X^v$ as above, there is a linear map $v \mapsto J_v$ such that $X^v = \sharp dJ_v$ for every $v \in \mathfrak{g}$ if and only if $\iota_{X^v}\omega$ is exact for all $v \in \mathfrak{g}$.*

Proof. If $H_v = H_{X^v}$ exists for all v then $dJ_{X^v} = \omega(X^v, .) = \iota_{X^v}\omega$ for all v so $\iota_{X^v}\omega$ is exact for all $v \in \mathfrak{g}$. Conversely, if for every $v \in \mathfrak{g}$ there is a smooth function h_v with $dh_v = \iota_{X^v}\omega$ then $X^v = \sharp dh_v$ so h_v is Hamiltonian for X^v. Now let $v_1, ..., v_n$ be a basis for $\mathfrak{g}$ and define $J_{v_i} = h_{v_i}$ and extend linearly. □

Notice that the property that $v \mapsto J_v$ is linear means that we can define a map $J : M \to \mathfrak{g}^*$ by

$$J(x)(v) = J_v(x)$$

and this is called a **momentum map** .

Definition 15.53. A symplectic G-action σ (resp. $\mathfrak{g}$-action σ') on M such that for every $v \in \mathfrak{g}$ the vector field X^v is a Hamiltonian vector field on M is called a **Hamiltonian G-action** (resp. Hamiltonian $\mathfrak{g}$-action).

We can thus associate to every Hamiltonian action at least one momentum map-this being unique up to an additive constant.

Example 15.54. If G acts on a manifold Q by diffeomorphisms then G lifts to an action on the cotangent bundle T^*M which is automatically symplectic. In fact, because $\omega_0 = d\theta_0$ is exact the action is also a Hamiltonian action. The Hamiltonian function associated to an element $v \in \mathfrak{g}$ is given by

$$J_v(x) = \theta_0\left(\left.\frac{d}{dt}\right|_0 \exp(tv) \cdot x\right).$$

Definition 15.55. If G (resp. $\mathfrak{g}$) acts on M in a symplectic manner as above such that the action is Hamiltonian and such that we may choose a momentum map J such that

$$J_{[v,w]} = \{J_v, J_w\}$$

where $J_v(x) = J(x)(v)$ then we say that the action is a **strongly Hamiltonian** G-action (resp. $\mathfrak{g}$-action).

Example 15.56. The action of example 15.54 is strongly Hamiltonian.

We would like to have a way to measure of whether a Hamiltonian action is strong or not. Essentially we are just going to be using the difference $J_{[v,w]} - \{J_v, J_w\}$ but it will be convenient to introduce another view which we postpone until the next section where we study "Poisson structures".

PUT IN THEOREM ABOUT MOMENTUM CONSERVATION!!!!

What is a momentum map in the cotangent case? Pick a fixed point $\alpha \in T^*Q$ and consider the map $\Phi_\alpha : G \to T^*Q$ given by $\Phi_\alpha(g) = g \cdot \alpha = g^{-1*}\alpha$. Now consider the pull-back of the canonical 1-form $\Phi_\alpha^*\theta_0$.

Lemma 15.57. *The restriction* $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}$ *is an element of* $\mathfrak{g}^*$ *and the map* $\alpha \mapsto \Phi_\alpha^*\theta_0|_{\mathfrak{g}}$ *is the momentum map.*

Proof. We must show that $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v) = H_v(\alpha)$ for all $v \in \mathfrak{g}$. Does $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v)$ live in the right place? Let $g_v^t = \exp(vt)$. Then

$$\begin{aligned}(T_e\Phi_\alpha v) &= \frac{d}{dt}\Big|_0 \Phi_\alpha(\exp(vt)) \\ &= \frac{d}{dt}\Big|_0 (\exp(-vt))^*\alpha \\ &\quad \frac{d}{dt}\Big|_0 \exp(vt)\cdot\alpha\end{aligned}$$

We have

$$\begin{aligned}\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v) &= \theta_0|_{\mathfrak{g}}(T_e\Phi_\alpha v) \\ = \theta_0(\frac{d}{dt}\Big|_0 \exp(vt)\cdot\alpha) &= J_v(\alpha)\end{aligned}$$

□

Definition 15.58. Let G act on a symplectic manifold M, ω and suppose that the action is Hamiltonian. A momentum map J for the action is said to be equivariant with respect to the coadjoint action if $J(g \cdot x) = \mathrm{Ad}_{g^{-1}}^* J(x)$.

Chapter 16

Poisson Geometry

Life is good for only two things, discovering mathematics and teaching mathematics

–Siméon Poisson

16.1. Poisson Manifolds

In this chapter we generalize our study of symplectic geometry by approaching things from the side of a Poisson bracket.

Definition 16.1. A **Poisson structure** on an associative algebra $\mathcal{A}$ is a Lie algebra structure with bracket denoted by $\{.,.\}$ such for a fixed $a \in \mathcal{A}$ that the map $x \mapsto \{a, x\}$ is a derivation of the algebra. An associative algebra with a Poisson structure is called a **Poisson algebra** and the bracket is called a **Poisson bracket** .

We have already seen an example of a Poisson structure on the algebra $\mathfrak{F}(M)$ of smooth functions on a symplectic manifold. Namely,

$$\{f, g\} = \omega(\omega^\sharp df, \omega^\sharp dg).$$

By the Darboux theorem we know that we can choose local coordinates $(q^1, ..., q^n, p_1, ..., p_n)$ on a neighborhood of any given point in the manifold. Recall also that in such coordinates we have

$$\omega^\sharp df = \sum_{i=1}^{n} \frac{\partial f}{\partial p_i} \frac{\partial}{\partial q^i} - \sum_{i=1}^{n} \frac{\partial f}{\partial q^i} \frac{\partial}{\partial p_i}$$

sometimes called the symplectic gradient. It follows that

$$\sum_{i=1}^{n} (\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i})$$

Definition 16.2. A smooth manifold with a Poisson structure on is algebra of smooth functions is called a **Poisson manifold**.

So every symplectic n-manifold gives rise to a Poisson structure. On the other hand, there are Poisson manifolds that are not so by virtue of being a symplectic manifold.

Now if our manifold is finite dimensional then every derivation of $\mathfrak{F}(M)$ is given by a vector field and since $g \mapsto \{f, g\}$ is a derivation there is a corresponding vector field X_f. Since the bracket is determined by these vector field and since vector fields can be defined locally (recall the presheaf $\mathfrak{X}_M$) we see that a Poisson structure is also a locally defined structure. In fact, $U \mapsto \mathfrak{F}_M(U)$ is a presheaf of Poisson algebras.

Now if we consider the map $w : \mathfrak{F}_M \to \mathfrak{X}_M$ defined by $\{f, g\} = w(f) \cdot g$ we see that $\{f, g\} = w(f) \cdot g = -w(g) \cdot f$ and so $\{f, g\}(p)$ depends only on the differentials df, dg of f and g. Thus we have a tensor $B(.,.) \in \Gamma \bigwedge^2 TM$ such that $B(df, dg) = \{f, g\}$. In other words, $B_p(.,.)$ is a symmetric bilinear map $T_p^*M \times T_p^*M \to \mathbb{R}$. Now any such tensor gives a bundle map $B^\sharp : T^*M \mapsto T^{**}M = TM$ by the rule $B^\sharp(\alpha)(\beta) = B(\beta, \alpha)$ for $\beta, \alpha \in T_p^*M$ and any $p \in M$. In other words, $B(\beta, \alpha) = \beta(B^\sharp(\alpha))$ for all $\beta \in T_p^*M$ and arbitrary $p \in M$. The 2-vector B is called the Poisson tensor for the given Poisson structure. B is also sometimes called a cosymplectic **structure** for reasons that we will now explain.

If M, ω is a symplectic manifold then the map $\omega_\flat : TM \to T^*M$ can be inverted to give a map $\omega^\sharp : T^*M \to TM$ and then a form $W \in \bigwedge^2 TM$ defined by $\omega^\sharp(\alpha)(\beta) = W(\beta, \alpha)$ (here again β, α must be in the same fiber). Now this form can be used to define a Poisson bracket by setting $\{f, g\} = W(df, dg)$ and so W is the corresponding Poisson tensor. But notice that

$$\begin{aligned}\{f, g\} &= W(df, dg) = \omega^\sharp(dg)(df) = df(\omega^\sharp(dg)) \\ &= \omega(\omega^\sharp df, \omega^\sharp dg)\end{aligned}$$

which is just the original Poisson bracket defined in the symplectic manifold M, ω.

Given a Poisson manifold $M, \{.,.\}$ we can always define $\{.,.\}_-$ by $\{f, g\}_- = \{g, f\}$. Since we some times refer to a Poisson manifold $M, \{.,.\}$ by referring just to the space we will denote M with the opposite Poisson structure by M^-.

A Poisson map is map $\phi : M, \{.,.\}_1 \to N, \{.,.\}_2$ is a smooth map such that $\phi^*\{f, g\} = \{\phi^* f, \phi^* g\}$ for all $f, g \in \mathfrak{F}(M)$.

For any subset S of a Poisson manifold let S_0 be the set of functions from $\mathfrak{F}(M)$ that vanish on S. A submanifold S of a Poisson manifold $M, \{.,.\}$ is called **coisotropic** if S_0 closed under the Poisson bracket. A Poisson manifold is called symplectic if the Poisson tensor B is non-degenerate since in this case we can use $B^\sharp$ to define a symplectic form on M. A Poisson manifold admits a (singular) foliation such that the leaves are symplectic. By a theorem of A. Weinstien we can locally in a neighborhood of a point p find a coordinate system (q^i, p_i, w^i) centered at p and such that

$$B = \sum_{i=1}^{k} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i} + \frac{1}{2} \sum_{i,j} a^{ij}() \frac{\partial}{\partial w^i} \wedge \frac{\partial}{\partial w^j}$$

where the smooth functions depend only on the w's. vanish at p. Here k is the dimension of the leave through p. The rank of the map $B^\sharp$ on T_p^*M is k.

Now to give a typical example let $\mathfrak{g}$ be a Lie algebra with bracket $[.,.]$ and $\mathfrak{g}^*$ its dual. Choose a basis $e_1, ..., e_n$ of $\mathfrak{g}$ and the corresponding dual basis $\epsilon^1, ..., \epsilon^n$ for $\mathfrak{g}^*$. With respect to the basis $e_1, ..., e_n$ we have

$$[e_i, e_j] = \sum C_{ij}^k e_k$$

where C_{ij}^k are the structure constants.

For any functions $f, g \in \mathfrak{F}(\mathfrak{g}^*)$ we have that df_α, dg_α are linear maps $\mathfrak{g}^* \to \mathbb{R}$ where we identify $T_\alpha \mathfrak{g}^*$ with $\mathfrak{g}^*$. This means that df_α, dg_α can be considered to be in $\mathfrak{g}$ by the identification $\mathfrak{g}^{**} = \mathfrak{g}$. Now define the $\pm$ Poisson structure on $\mathfrak{g}^*$ by

$$\{f, g\}_\pm(\alpha) = \pm\alpha([df_\alpha, dg_\alpha])$$

Now the basis $e_1, ..., e_n$ is a coordinate system y on $\mathfrak{g}^*$ by $y_i(\alpha) = \alpha(e_i)$.

Proposition 16.3. In terms of this coordinate system the Poisson bracket just defined is

$$\{f, g\}_\pm = \pm \sum_{i=1}^{n} B_{ij} \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j}$$

where $B_{ij} = \sum C_{ij}^k y_k$.

Proof. We suppress the $\pm$ and compute:

$$\begin{aligned}
\{f, g\} = [df, dg] &= [\sum \frac{\partial f}{\partial y_i} dy_i, \sum \frac{\partial g}{\partial y_j} dy_j] \\
&= \sum \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} [dy_i, dy_j] = \sum \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} \sum C_{ij}^k y_k \\
&= \sum_{i=1}^{n} B_{ij} \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j}
\end{aligned}$$

□

Chapter 17

Geometries

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

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

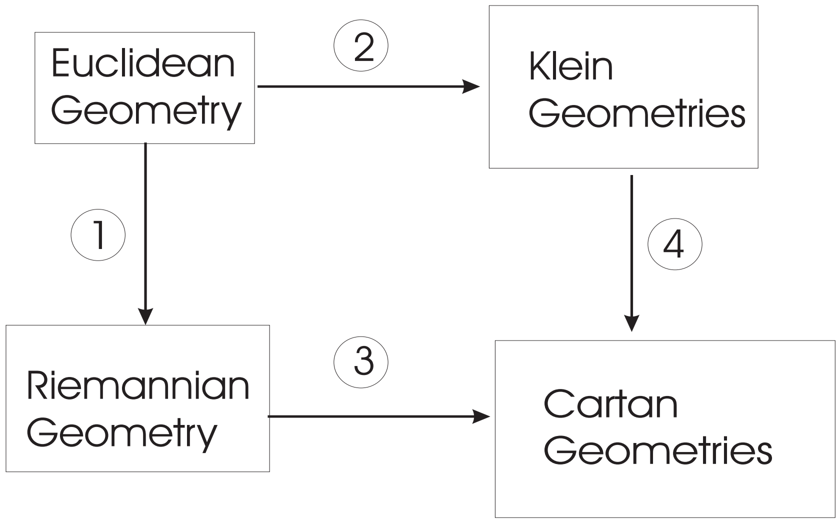
geometry. A major effect of this was to make the subject easier (thank you, Descarte!). Coordinates paved the way for the calculus of several variables and then the modern theory of differentiable manifolds which will be a major topic of the sequel.

It was Felix Klein's program (Klein's Erlangen Programm) that made groups 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 17.

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 **??**).

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



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 connection which takes values in the Lie algebra of $\mathrm{Euc}(n)$ 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 [**Shrp**]. 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 17.1. 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 17.1. 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 17.2. 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.

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

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

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".

17.1. 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 17.3. 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 17.4. 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 17.5. If G acts on M the we call M a G-**space**.

Exercise 17.6. 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 17.7. 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 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.

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

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

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 17.8. 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 17.9. 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 17.10. 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 17.11. 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 17.12. 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	$\mathrm{O}(\mathsf{E})$	?

17.2. Some Klein Geometries

Definition 17.13. 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 17.14. 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 17.15. 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.

17.2.1. 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 17.16. Show that if $\mathbf{A}$ an is a n-dimensional affine space with difference space V then $\mathrm{Aff}(\mathbf{A}) \cong \mathrm{Aff}(\mathbf{A}^n)$ and the $\mathrm{Aff}(\mathbf{A})$-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 17.17. 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.

17.2.2. 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 17.18. 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.

17.2.3. 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 17.19. 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 17.20. 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 17.21. 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

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

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 17.22. 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 17.23. 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

$$FAF^t + FA^tF^t = 0$$
$$A + A^t = 0$$

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

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

Exercise 17.24. 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) $\mathbf{T}_\gamma(t) = \gamma' / \|\gamma'\|$
(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 17.25. Show that $\gamma'' = \frac{dv}{dt}\mathbf{T}_\gamma + v^2 \kappa_\gamma \mathbf{N}_\gamma$ where $v = \|\gamma'\|$.

Example 17.26. 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 17.27. 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 17.28. 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 17.21 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 17.22 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.

17.2.4. 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$\times\mathbb{R}$. 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**

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

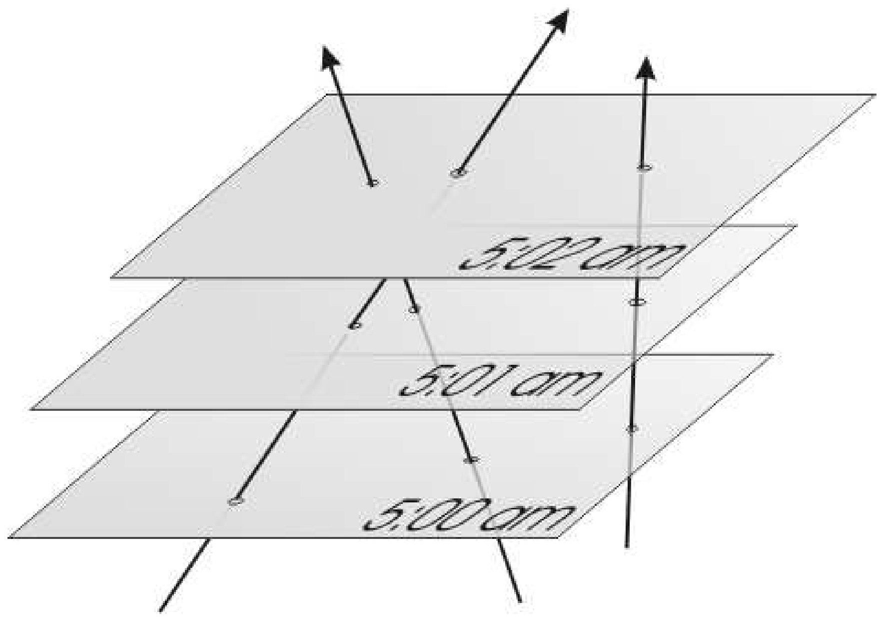
for some $(t_0, x_0, y_0, z_0) \in \mathbb{R}^4$.

(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.

17.2.5. Minkowski Spacetime. As we have seen, the vector space $\mathbb{R}^4$ may be provided with a special scalar product given by $\langle x, y\rangle := x^0y^0 - \sum_{i=1}^{3} x^iy^i$ 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 $\langle x, y\rangle := c^2x^0y^0 - \sum_{i=1}^{3} x^iy^i$ 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 17.29. 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

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

[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.

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) 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

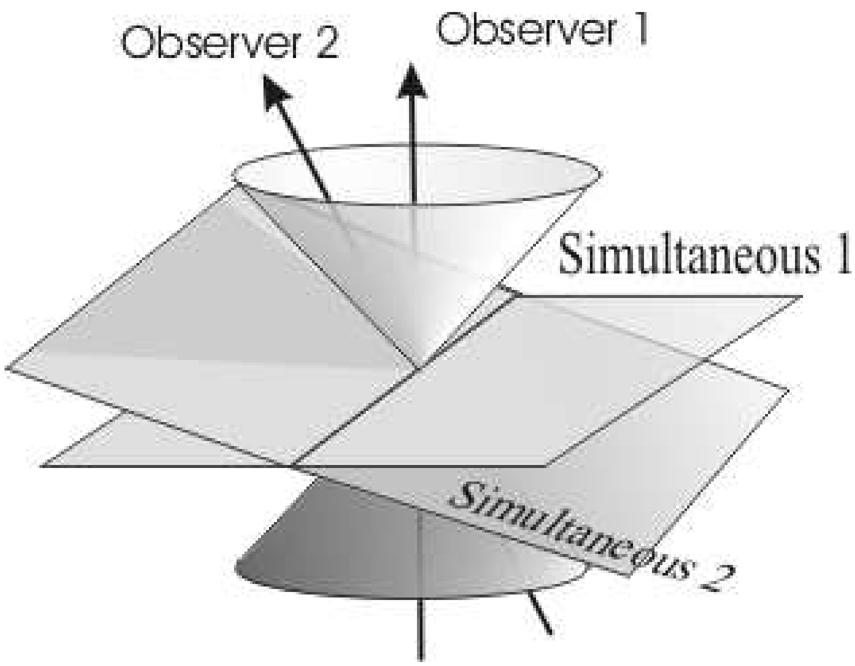
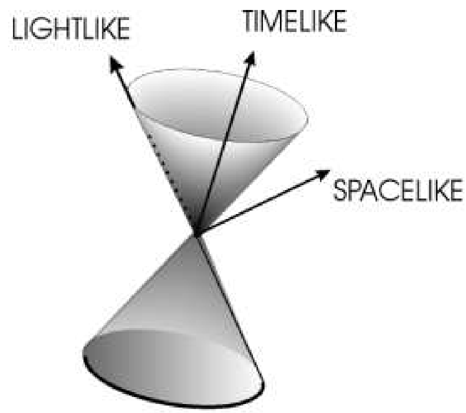


Figure 17.1. Relativity of Simultaneity

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 17.30. 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**.

[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.



Remark 17.31 (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 17.32. 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 17.33. 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 $\langle \gamma(t+\epsilon) - \gamma(t), \gamma(t+\epsilon) - \gamma(t)\rangle$ 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 $t \to (t, 0, 0, 0)$. 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:

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

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.

17.2.6. 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

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

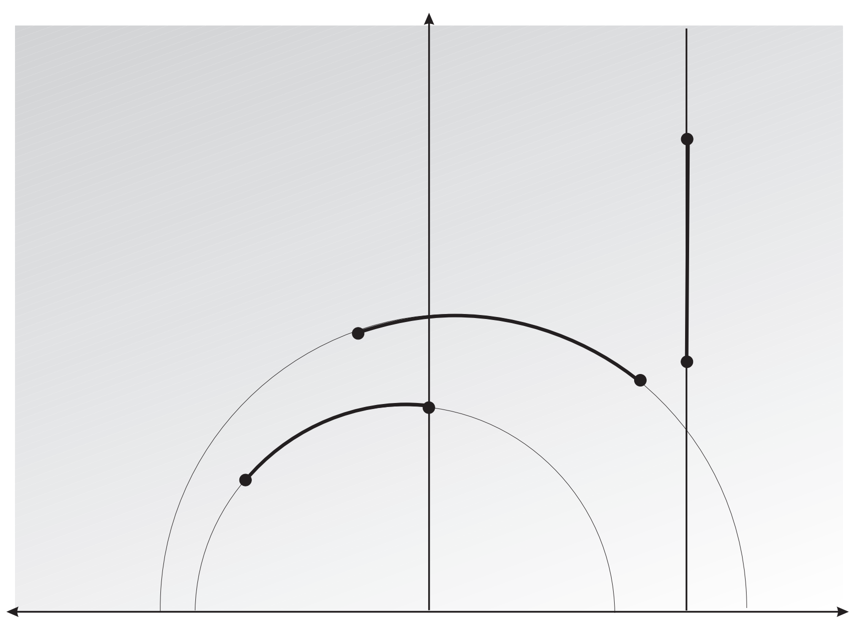
where $p = (x, y) \sim x + iy$ and $v = (v^1, v^2)$, $w = (w^1, w^2)$. In this context, the assignment of an inner product the tangent space at each point is called a metric. We are identifying $\mathbb{R}^2$ with $\mathbb{C}$. Now the length of a curve $\gamma : t \mapsto (x(t), y(t))$ defined on $[a,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

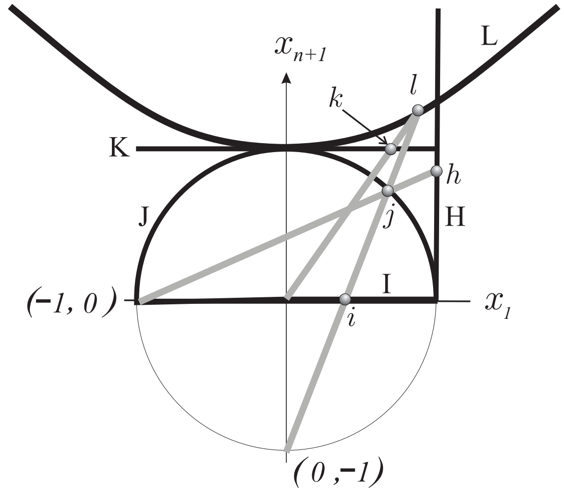
$$s := l(t) = \int_a^t \left(\frac{\dot{x}(\tau)^2 + \dot{y}(\tau)^2}{y(\tau)^2} \right)^{1/2} d\tau$$

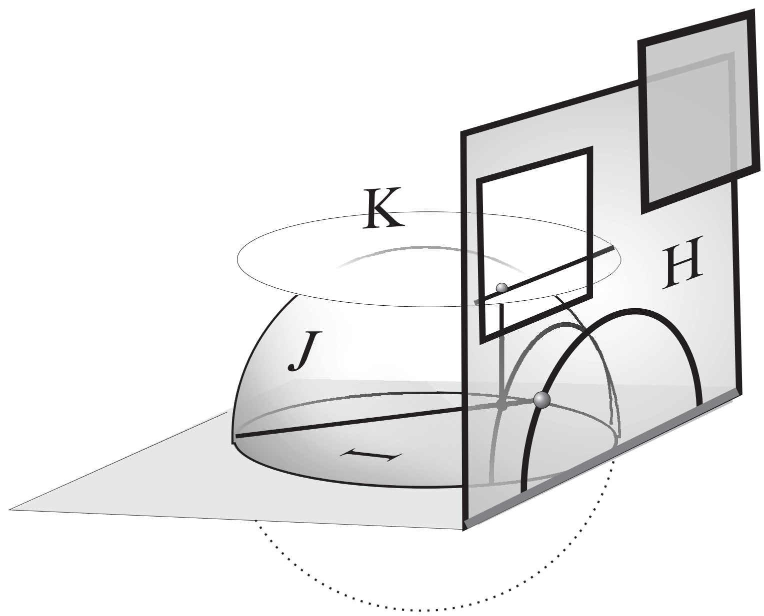
The function $l(t)$ is invertible and it is then possible to reparameterize the curve by arc length $\widetilde{\gamma}(s) := \gamma(l^{-1}(s))$.The distance between any two points in the upper half-plane is the length of the shortest curve that connects the two points. Of course, one must show that 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.

17.2.7. Models of Hyperbolic Space. Under constructionn





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

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

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.

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

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.

17.2.8. 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.

Bibliography

[Arm] M. A. Armstrong, *Basic Topology*, Springer-Verlag, (1983).

[At] M. F. Atiyah, *K-Theory*, W.A.Benjamin, (1967).

[A-M-R] R. Abraham, J.E. Marsden,, and T. Ratiu, *Manifolds, tensor analysis, and applications*, Addison Wesley, Reading, (1983).

[Arn] V.I. Arnold, *Mathematical methods of classical mechanics*, Graduate Texts in Math. **60**, Springer-Verlag, New York, 2nd edition (1989).

[BaMu] John Baez and Javier P. Muniain, *Gauge Fields, Knots and Gravity*, World Scientific, (1994).

[Be] Arthur L. Besse, *Einstein Manifolds*, Classics in Mathematics, Springer, (1987).

[BishCr] R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds*, New York:Academic Press, (1964).

[Bo-Tu] R. Bott and L. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag GTM 82, (1982).

[Bre] G. Bredon, *Topology and Geometry*, Springer-Verlag GTM 139, (1993).

[BrCl] F. Brickell and R. Clark, *Differentiable Manifolds; An Introduction*, Van Nostran Rienhold, (1970).

[Bro-Jan] Th. Bröcker and K. Jänich, *Introduction to Differential Topology*, Cambridge University Press, (1982).

[Brou] L.E.J. Brouwer, “ Uber Abbildung von Mannigfaltigkeiten,” Mathematische Annalen 71 (1912) 97-115.

[Chavel] I. Chavel, *Eigenvalues in Riemannian Geometry,* Orlando: Academic Press, (1984).

[C-E] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, Amsterdam: North-Holland, (1975).

[Cheng] S.Y. Cheng, Eigenvalue Comparison Theorems and its Geometric Applications Math. Z. **143**, 289–297, (1975).

[Clark] C.J.S. Clark, On the Global isometric embedding of pseudo-Riemannian manifolds, Proc. Roy. Soc. A314 (1970) 417-428

[Dar] R. W. R. Darling, *Differential Forms and Connections*, Cambridge University Press, (1994)

[Dieu] J. Dieudonn´e, *A History of Algebraic and Differential Topology* 1900-1960, Birkhauser,(1989).

[Dod-Pos] C.T.J. Dodson and T. Poston, *Tensor Geometry: The Geometric Viewpoint and Its Uses*, Springer-Verlag (2000).

[Dol] A. Dold, *Lectures on Algebraic Topology*, Springer-Verlag, (1980).

[Donaldson] S. Donaldson, An application of Gauge theory to the Topology of 4-manifolds, J. Diff. Geo. 18 (1983), 269-316.

[Dug] J. Dugundji, *Topology*, Allyn & Bacon, (1966).

[Eil-St] S. Eilenberg and N. Steenrod, *Foundations of Algebraic Topology*, Princeton Univ. Press, (1952).

[Fen] R. Fenn, *Techniques of Geometric Topology*, Cambridge Univ. Press, (1983).

[Freedman] M. Freedman, The Topology of Four Dimensional Manifolds, J. Diff. Geo. **17** (1982) 357-454.

[Fr-Qu] M. Freedman and F. Quinn, *Topology of 4-Manifolds*, Princeton Univ. Press, (1990).

[Fult] W. Fulton, *Algebraic Topology: A First Course*, Springer-Verlag, (1995).

[G2] V. Guillemin, , and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge Univ. Press, Cambridge, (1984).

[Gu-Hu-We] K. Guruprasad, J. Huebschmann, L. Jeffrey, and A. Weinstein, Group systems, groupoids, and moduli spaces of parabolic bundles, Duke Math. J. 89 (1997), 377-412.

[Gray] B. Gray, *Homotopy Theory*, Academic Press, (1975).

[Gre-Hrp] M. Greenberg and J. Harper, *Algebraic Topology: A First Course*, Addison-Wesley, (1981).

[Grom] M. Gromov, "Convex sets and Kähler manifolds" in *Advances in differential geometry and topology,* World Sci. Publ., Teaneck, NJ ,{1990}, 1-38.

[Helg] S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces,* Amer. Math. Soc., 2001.

[Hicks] Noel J. Hicks, *Notes on Differential Geometry*, D. Van Nostrand Company Inc, (1965).

[Hilt1] P. J. Hilton, *An Introduction to Homotopy Theory*, Cambridge University Press, (1953).

[Hilt2] P. J. Hilton and U. Stammbach, *A Course in Homological Algebra*, Springer-Verlag, (1970).

[Hus] D. Husemoller, *Fibre Bundles*, McGraw-Hill, (1966) (later editions by Springer-Verlag).

[Kirb-Seib] R. Kirby and L. Siebenmann, *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*, Ann. of Math.Studies **88**, (1977).

[Klein] Felix Klein, *Gesammelte Abhandlungen, Vol I,* Springer, Berlin, (1971).

[Kob] S. Kobayashi, *On Conjugate and Cut Loci* in Global Differential Geometry, M.A.A. Studies in Maths, Vol. 27, S.S Chern Editor, Prentice Hall (1989)

[K-N] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry.* I, II, J. Wiley Interscience, (1963), (1969).

[L1] S. Lang, *Fundamentals of Differential Geometry*, Springer-Verlag GTM vol 191

[L2] S. Lang, *Algebra,* Springer-Verlag GTM vol 211

[L-M] H. Lawson and M. Michelsohn, *Spin geometry*, Princeton University Press, Princeton, (1989).

[Lee, Jeff] Jeffrey M Lee, Online Supplement to the present text, Internet: `http://webpages.acs.ttu.edu/jlee/Supp.pdf`

[Lee, John] John Lee, *Introduction to Smooth Manifolds*, Springer-Verlag GTM vol 218 (2002).

[L-R] David Lovelock and Hanno Rund, Tensors, Differential forms, and Variational Principles, Dover Publications, New York (1989).

[Madsen] Id Madsen and Jørgen Tornehave, *From Calculus to Cohomology*, Cambridge University Press (1997).

[Matsu] Y. Matsushima, *Differentiable Manifolds*, Marcel Dekker, New York, 1972.

[M-T-W] C. Misner, J. Wheeler, and K. Thorne, *Gravitation*, Freeman, (1974)

[Mil] J. Milnor, Morse Theory, Annals of Mathematics Studies **51**, Princeton U. Press, Princeton, (1963).

[MacL] S. MacLane, *Categories for the Working Mathematician*, Springer-Verlag GTM 5, (1971).

[Mass] W. Massey, *Algebraic Topology: An Introduction*, Harcourt, Brace & World, (1967) (reprinted by Springer-Verlag).

[Mass2] W. Massey, *A Basic Course in Algebraic Topology*, Springer-Verlag, (1993).

[Maun] C. R. F. Maunder, *Algebraic Topology*, Cambridge Univ. Press, (1980) (reprinted by Dover Publications).

[Mich] P. Michor, *Topics in Differential Geometry*, Graduate Studies in Mathematics, v.93, (2008)

[Miln1] J. Milnor, *Topology from the Differentiable Viewpoint*, Univ. Press of Virginia, (1965).

[Mil-St] J. Milnor and J. Stasheff, *Characteristic Classes*, Ann. of Math. Studies 76, (1974).

[Molino] P. Molino, *Riemannian foliations,* Progress in Mathematics, Boston: Birkhauser, (1988).

[My-St] S. B. Myers, N. E. Steenrod, The group of isometries of a Riemannian manifold. Annals of Mathematics, Vol. 40, No. 2, (1939).

[ON1] B. O'Neill, *Semi-Riemannian Geometry*, Academic Press, New York, (1983).

[ON2] B. O'Neill, *Elementary differential geometry*, Academic Press, (1997).

[Pa] Richard, Palais, *Natural Operations on Differential Forms*, Trans. Amer. Math. Soc. (1959), pp 125-141

[Pa2] Richard, Palais, *A Modern Course on Curves and Surfaces*, Online Book at http://www.math.uci.edu/~cterng/NotesByPalais.pdf.

[Pen] Roger Penrose, *The Road to Reality*, Alfred A. Knopf, New York, (2005).

[Pe] Peter Peterson, *Riemannian Geometry*, Springer-Verlag (1991).

[Poor] Walter Poor, *Differential Geometric Structures*, Mcgraw Hill, (1981).

[Nash1] John Nash, C^1-isometric imbeddings", Annals of Mathematics, 60 (1954), pp 383-396.

[Nash2] John Nash, The imbedding problem for Riemannian manifolds, Annals of Mathematics, 63 (1956), pp 20-63.

[Roe] J. Roe, Elliptic Operators, *Topology and Asymptotic methods*, Longman, (1988).

[Ros] W. Rossmann, *Lie Groups: An Introduction through Linear Groups*, Oxford University Press (2002).

[Shrp] R. Sharpe, *Differential Geometry; Cartan's Generalization of Klein's Erlangen Program*, Springer-Verlag (1997).

[Span] Edwin H. Spanier, *Algebraic Topology*, Springer-Verlag, (1966).

[Spv] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, (5 volumes) Publish or Perish Press, (1979).

[St] N. Steenrod. *Topology of fiber bundles*, Princeton University Press, (1951).

[Stern] Shlomo Sternberg, *Lectures on Differential Geometry*, Prentice Hall (1964).

[Tond] Ph. Tondeur, *Foliations on Riemannian manifolds,* New York: Springer Verlag, (1988).

[Tond2] Ph. Tondeur, *Geometry of Foliations*, Monographs in Mathematics, vol. 90, Basel: Birkhäuser, (1997).

[Wal] Gerard Walschap, *Metric Structures in Differential Geometry*, Springer-Verlag, (2004).

[War] Frank Warner,. *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, (1983).

[Wh] John A. Wheeler, *A Journey into Gravity and Spacetime*, Scientific American Library, A division of HPHLP, New York (1990).

[We1] A. Weinstein, *Lectures on Symplectic Manifolds*, Regional conference series in mathematics **29**, Amer. Math. Soc.,Providence, (1977).

Index