# MATH 8210 LECTURE NOTES, PART 2

## 1. Submanifolds

Throughout this section fix a smooth $m$-dimensional manifold $M$.

*Definition* 1.1. Let $N$ be a smooth manifold and let $\phi\colon N \to M$ be a smooth map. Then

- $\phi$ is called a *submersion* if, for all $x \in N$, the linearization $\phi_*\colon T_xN \to T_{\phi(x)}M$ is surjective.
- $\phi$ is called an *immersion* if, for all $x \in N$, the linearization $\phi_*\colon T_xN \to T_{\phi(x)}M$ is injective.
- $\phi$ is called an *embedding* if it is an immersion and, moreover, the map $\phi$ is a homeomorphism from $N$ to $\phi(N)$, where $\phi(N)$ is equipped with the subspace topology.

Here are some examples; if these notions are unfamiliar to you then you should check for yourself that they satisfy the respective definitions.

**Example 1.2.** (i) *The projection $\pi\colon \mathbb{R}^n \to \mathbb{R}^m$ onto the first m coordinates (assuming $m \leq n$) is a submersion; in fact this provides a local model for all submersions, as will follow from the proof of Theorem 1.8. For a more interesting global example, the projection $\pi\colon \mathbb{R}^{n+1} \setminus \{\vec{0}\} \to \mathbb{R}P^n$ is a submersion (as is the projection $S^n \to \mathbb{R}P^n$).*

(ii) *Dually, if $n \leq m$, then the inclusion $i\colon \mathbb{R}^n \to \mathbb{R}^m$ (defined by $i(\vec{x}) = (\vec{x}, \vec{0})$ where $\mathbb{R}^m$ is split as $\mathbb{R}^n \times \mathbb{R}^{m-n}$) is an example of an embedding.*

(iii) *A simple example of an immersion which is not an embedding is the map $\phi\colon \mathbb{R} \to \mathbb{C}$ given by $\phi(x) = e^{ix}$.*

(iv) *Of course the problem with (iii) was that it wasn't injective, but one can also construct examples of injective immersions which are not embeddings. For instance, take two smooth functions $f, g\colon \mathbb{R} \to \mathbb{R}$ such that for all $t < 0$ one has $f(t) = t$ and $g(t) = 0$, and such that there is no $t \in \mathbb{R}$ such that $f'(t) = g'(t) = 0$. Then the map $\phi\colon \mathbb{R} \to \mathbb{R}^2$ defined by $\phi(t) = (f(t), g(t))$ will be an immersion. If one chooses $f$ and $g$ so that $\lim_{t\to\infty} f(t) = -1$ and $\lim_{t\to\infty} g(t) = 0$ and so that $\phi$ is injective (as can easily be done—you might draw a picture if this isn't obvious to you), then $\phi$ won't be an embedding, since by looking at neighborhoods of $(-1, 0)$ in $\psi(\mathbb{R})$ one sees that the image isn't a topological manifold when equipped with the subspace topology.*

Declare two embeddings $\phi_1\colon N_1 \to M$ and $\phi_2\colon N_2 \to M$ to be equivalent if there is a diffeomorphism $\psi\colon N_1 \to N_2$ such that $\phi_1 = \phi_2 \circ \psi$. An overly formal definition of a *submanifold* is that a submanifold is an equivalence class of embeddings under this equivalence relation. Of course, part of the point of the above equivalence relation is that if $\phi_1 \sim \phi_2$ then $\phi_1(N_1) = \phi_2(N_2)$; when one thinks of a submanifold one should think of the subset of $M$ formed as the image of any representative embedding. If one has a subset $N \subset M$, it inherits a subspace topology, and one can ask whether or not this subspace topology makes $N$ a topological manifold. One can then ask whether the topological space $N$ admits smooth structures (this is now an intrinsic question about $N$), and how these are related to the ambient space $M$. Accordingly I prefer the following definition:

*Definition* 1.3. A *submanifold* of $M$ is a subset $N$ which is a topological manifold with respect to its subspace topology, equipped moreover with a smooth structure such that the inclusion $i: N \to M$ is an embedding.

This is equivalent to the definition using equivalence classes of embeddings: if $\phi_1: N_1 \to M$ and $\phi_2: N_2 \to M$ are equivalent embeddings (with common image $N \subset M$) then one can get a smooth atlas on $N$ by constructing charts by precomposition with either $\phi_1^{-1}: N \to N_1$ or $\phi_2^{-1}: N \to N_2$ (since $\phi_1 \sim \phi_2$ the atlases so obtained will be equivalent), and with this atlas the inclusion $i: N \to M$ will be a distinguished member of the equivalence class of $\phi_1$ and $\phi_2$. So we can (and do) identify the equivalence class with this distinguished member.

Accordingly let $N \subset M$ be a submanifold, with $i: N \to M$ the inclusion. In particular $i$ is an immersion, so for each $x \in N$ we have an induced injective linear map $i_*: T_xN \to T_xM$. We can then identify $T_xN$ with its image under this map—in other words, for every $x \in N$ we have a natural identification of $T_xN$ with a subspace of $T_xM$.

**Theorem 1.4.** *If $N \subset M$ is a submanifold where $\dim N = n$ and $\dim M = m$ and $x_0 \in N$, there exists a coordinate chart $\phi: U \to \mathbb{R}^m$ for $M$ such that $x_0 \in U$ and $\phi^{-1}(\mathbb{R}^n \times \{\vec{0}\}) = N \cap U$.*

*Proof.* As can easily be seen from the definition of the subspace topology, there is a neighborhood $U_0 \subset M$ of $x_0$ which is the domain of a coordinate chart $\phi_0: U_0 \to \mathbb{R}^m$ for $M$, such that where $V_0 = U_0 \cap N$, $V_0$ is the domain of some coordinate chart $\psi_0: V_0 \to \mathbb{R}^n$ for $N$. By replacing $\phi_0$ and $\psi_0$ by their compositions with translations we may as well assume that $\phi_0(x_0)$ and $\psi_0(x_0)$ are the origins of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Also, by composing $\phi_0$ with an appropriate linear map, we may as well assume that the composition $\phi_0 \circ \psi_0^{-1}: \psi_0(V_0) \to \phi_0(U_0)$ (which is a smooth map with injective linearization from a neighborhood of the origin in $\mathbb{R}^n$ to a neighborhood of the origin in $\mathbb{R}^m$) has the property that its linearization at 0 is given by $\vec{v} \mapsto (\vec{v}, 0)$ where we split $\mathbb{R}^m$ as $\mathbb{R}^n \times \mathbb{R}^{m-n}$.

Now define a map $\alpha: \psi_0(V_0) \times \mathbb{R}^{m-n} \to \mathbb{R}^n$ by $\alpha(x, y) = (\phi_0 \circ \psi_0^{-1})(x) + (\vec{0}, y)$. This map $\alpha$ is $C^\infty$, and its linearization at $(\vec{0}, \vec{0})$ is the identity. The inverse function theorem from multivariable calculus then asserts that $\alpha$ is a local diffeomorphism near $\vec{0}$, *i.e.* that there is a neighborhood $W$ of the origin in $\mathbb{R}^m$ and a smooth map $\beta: \alpha(W) \to W$ so that $\beta \circ \alpha: W \to W$ and $\alpha \circ \beta: \alpha(W) \to \alpha(W)$ are the respective identities.

Now set $U = \phi_0^{-1}(\alpha(W)) \subset M$ and $\phi = \beta \circ \phi_0$. $\phi$ is a composition of two maps which are diffeomorphisms to their images in $\mathbb{R}^m$, so $\phi$ is a coordinate chart in $M$ (in the maximal atlas for $M$). Moreover since by construction we have $\alpha(\psi_0(V_0) \times \{0\}) = \phi_0(N \cap U_0)$, we see that $\beta(\phi_0(N \cap U)) = W \cap (\psi_0(V_0) \times \{0\})$. In other words, $\phi$ maps the points of its domain which lie in $N$ precisely to the points of its range (namely $W$) which lie in $\mathbb{R}^n \times \{0\}$, as desired. □

*Remark* 1.5. Conversely, suppose that $N \subset M$ is a subset such that every point $x_0 \in N$ is contained in a coordinate chart for $M$ as in Theorem 1.4, so there is an $M$-neighborhood $U$ for $x_0$ and a coordinate chart $\phi: U \to \mathbb{R}^m$ so that $\phi^{-1}(\mathbb{R}^n \times \{0\}) = N \cap U$. For any such coordinate chart $\phi$, the restriction $\phi|_{N \cap U}$ is a homeomorphism to an open subset of $\mathbb{R}^n \times \{0\} \cong \mathbb{R}^n$; this shows that $N$ is a topological manifold. Moreover if $\phi_\alpha: U_\alpha \to \mathbb{R}^m$ and $\phi_\beta: U_\alpha: U_\beta \to \mathbb{R}^n$ are two such coordinate charts, so that by restricting $\phi_\alpha, \phi_\beta$ to $U_\alpha \cap N$ and $U_\beta \cap N$ we obtain homeomorphisms $\psi_\alpha, \psi_\beta$ from open subsets of $N$ to open sets in $\mathbb{R}^n$, then the transition map $\psi_\beta \circ \psi_\alpha^{-1}$ is just the restriction to $\phi_\alpha(U_\alpha \cap (\mathbb{R}^n \times \{0\}))$ of $\phi_\beta \circ \phi_\alpha^{-1}$, which is smooth. This proves that such charts $\psi_\alpha$ give $N$ the structure of a smooth manifold. Since in terms of the charts $\psi$ and $\phi$ the inclusion $i: N \to M$ is just given my the inclusion of $\mathbb{R}^n$ into $\mathbb{R}^m$ as the first $n$ coordinates, $i$ is an immersion. Thus, as a converse to Theorem 1.4, a subset of $M$ which can be covered by charts of the type described there is a submanifold of $M$.

If $N$ is a submanifold of $M$ and $\phi\colon U \to \mathbb{R}^m$ is a chart as in Theorem 1.4, note that if we define a map $f\colon U \to \mathbb{R}^{m-n}$ by taking the last $m-n$ coordinates of $\phi$ (*i.e.*, $f = (x_{m+1}, \dots, x_n)$), then $f\colon U \to \mathbb{R}^{m-n}$ is a submersion and $f^{-1}(\{\vec{0}\}) = N \cap U$. Moreover, within $U$, the tangent space to $N$ is given by the kernel of the linearization of $f$. This is a sort of converse to an important method of constructing submanifolds.

To prepare for this, we make the following definitions:

*Definition* 1.6. Let $f\colon M \to P$ be a smooth map between two smooth manifolds.

- A *critical point* of $f$ is a point $x \in M$ such that the linearization $f_*\colon T_xM \to T_{f(x)}P$ is *not* surjective.
- A *critical value* of $f$ is a point $y \in P$ such that $y = f(x)$ for some critical point $x$ of $f$.
- A *regular value* of $f$ is any point $y \in P$ which is not a critical value.

Note in particular that a point $y \in P$ which is not in the image of $f$ is still a regular value.

An important fact, which we will not prove, is the following:

**Theorem 1.7** (Sard's Theorem). *If $f\colon M \to P$ is a smooth map between two smooth manifolds then the set of critical values of $f$ has measure zero in $P$.*

(To make sense of this statement one has to know what "measure zero" means for a subset of a smooth manifold—to interpret this, note that a diffeomorphism between two open sets in Euclidean space preserves the class of sets of measure zero (even though it generally isn't measure preserving), so we can define a set of measure zero in a smooth manifold to be one whose intersection with the domain of every coordinate chart is mapped by that coordinate chart to a set of measure zero. If you prefer a statement that does not appeal to measure theory, it is also true that the set of regular values is residual in the sense of Baire—*i.e.*, it contains a countable intersection of open dense sets.)

Note that if $\dim M < \dim P$ and $f\colon M \to P$ is smooth, then since the linearization of $f$ is never surjective every point of $f(M) \subset P$ is a critical value. So in this case Sard's theorem amounts to the statement that $f(M)$ has measure zero in $P$, *i.e.* that the image of $f$ misses almost every point of $P$.

Whether we find a regular value of $f$ by appealing to Sard's theorem or by directly examining the map, the following gives a useful way of producing submanifolds:

**Theorem 1.8.** *Let $f\colon M \to P$ be a smooth map between two smooth manifolds and let $y_0 \in P$ be a regular value. Then $N = f^{-1}(\{y_0\})$ is a submanifold of $M$. Moreover if $n_0 \in N$ then $T_{n_0}N = \ker(f_*\colon T_{n_0}M \to T_{f(n_0)}P)$. (In particular, $\dim N = \dim M - \dim P$.)*

*Proof.* Let $m = \dim M$, and $p = \dim P$, and $n_0 \in N$. Let $(y_1, \dots, y_p)\colon V \to \mathbb{R}^p$ be a coordinate chart for $P$ around $f(n_0)$ which sends $f(n_0)$ to the origin For $i = 1, \dots, p$ define $z_i\colon f^{-1}(V) \to \mathbb{R}$ by $z_i = y_i \circ f$. By the surjectivity of $f_*$ at $n_0$, we may choose tangent vectors $v_1, \dots, v_p \in T_{x_0}M$ so that $dz_i(v_j) = dy_i(f_*v_j) = \delta_{ij}$. Let $S \leq T_{n_0}M$ be the span of $v_1, \dots, v_p$. We may then choose linearly independent cotangent vectors $\alpha_{p+1}, \dots, \alpha_m \in T^*_{n_0}M$ so that each $\alpha_i|_S = 0$. Let $\phi = (x_1, \dots, x_m)\colon U \to \mathbb{R}^m$ be a coordinate chart around $n_0$. In terms of these coordinates, the cotangent vectors $\alpha_i$ at $n_0$ can be written as $\sum_{k=1}^m \alpha_{ki}dx_k$ for some real numbers $\alpha_{ki}$. Define functions $z_{p+1}, \dots, z_m\colon U \to \mathbb{R}$ by $z_i = \sum_{k=1}^m \alpha_{ki}x_k$.

We now claim that the functions $z_1, \dots, z_m$ together provide a coordinate chart on a neighborhood of $n_0$. First note that the covectors $(dz_1)|_{n_0}, \dots, (dz_m)|_{n_0}$ are linearly independent elements of $T^*_{n_0}M$. For if $\sum c_i dz_i$ were to vanish at $n_0$, then by evaluating both sides on $v_j$ for $j = 1, \dots, p$ we obtain that $c_1 = \cdots = c_p = 0$, from which it also follows that $c_{p+1} = \cdots = c_m = 0$ since we chose the $\alpha_i = (dz_i)|_{n_0}$ for $i \geq p+1$ to be linearly independent. Since the $dz_i$ are linearly independent at $n_0$, they form a basis for $T^*_{n_0}M$ by a dimension count, and in particular there is a unique, bijective linear map of $T^*_{n_0}M$ which sends $(dz_i)_{n_0}$ to $(dx_i)_{n_0}$ for $i = 1, \dots, m$. But the this linear map is the transpose of the Jacobian at $\phi(n_0)$ of the map which sends $(x_1, \dots, x_m) \in \phi(U)$ to $(z_1, \dots, z_m)$, and so the Jacobian of this map $F: (x_1, \dots, x_n) \mapsto (z_1, \dots, z_n)$ is invertible at $\phi(n_0)$. So by the inverse function theorem there is an open set $W$ around $\phi(n_0)$ so that $F|_W$ is a diffeomorphism to its image. Recalling that $\phi = (x_1, \dots, x_m)$ was a coordinate chart for $M$, it follows from this that $(z_1, \dots, z_m): \phi^{-1}(W) \to \mathbb{R}^n$ is a diffeomorphism to its image, and so it contained in the maximal atlas defining the smooth structure on $M$.

So given a point $n_0 \in N \subset M$, we have constructed coordinate charts $\tilde{\phi}: \phi^{-1}(W) \to \mathbb{R}^m$ around $n_0$ and $(y_1, \dots, y_p): V \to \mathbb{R}^p$ around $f(n_0)$ in terms of which the map $f$ is given by $(z_1, \dots, z_m) \mapsto (z_1, \dots, z_p)$. In particular for any such coordinate chart $N \cap W$ is, in local coordinates, given by the preimage under the coordinate chart of $\{0\} \times \mathbb{R}^{m-p}$. By Remark 1.5, this suffices to establish that $N$ is a submanifold of $M$. $\square$

**Theorem 1.9.** *Let $K$ be a compact subset of a smooth manifold $M$. Then there exists an open set $V \subset M$ with $K \subset V$, a positive number $q$, and an embedding $\psi: V \to \mathbb{R}^q$. In particular if $M$ is a compact manifold then there is an embedding $\psi: M \to \mathbb{R}^q$ for some $q$.*

*Remark* 1.10. In fact the compactness assumption is not necessary—any smooth manifold $M$ embeds into Euclidean space of some dimension $q$, and indeed a result called the Whitney Embedding Theorem implies that one can take $q = 2\dim M$ (Whitney also showed that $\mathbb{R}P^{2^k}$ does not embed in any Euclidean space of dimension less than $2 \cdot 2^k$, so this is generally the best one can do).

*Proof.* Write $m = \dim M$. Any point $x \in K$ is contained in the image of a surjective coordinate chart $\phi^{(x)}: U^{(x)} \to B^m(2)$ with $\phi^{(x)}(x) = \vec{0}$ (where $B^m(2)$ denotes the open ball of radius 2 around the origin in $\mathbb{R}^m$). If we write $V^{(x)} = (\phi^{(x)})^{-1}(B^m(1))$, then the $V^{(x)}$ still cover $K$, and so by compactness they have a finite subcover $\{V^{(x_1)}, \dots, V^{(x_n)}\}$. Rename the $V^{(x_i)}, U^{(x_i)}$, and $\phi^{(x_i)}$ as $V_i, U_i, \phi_i$. For each $i$ let $\chi_i: M \to [0,1]$ be a smooth function such that $\chi_i^{-1}(1) = \bar{V}_i$ and which is supported in $U_i$. Also, define $\psi_i: M \to \mathbb{R}^n$ by $\psi_i(x) = \chi_i(x)\phi_i(x)$ if $x \in U_i$ and $\psi_i(x) = 0$ otherwise; of course this is smooth since $\chi_i$ is supported in $U_i$. Now define

$$\psi: M \to \mathbb{R}^{n(m+1)}$$

by

$$\psi(x) = \big(\chi_1(x), \psi_1(x), \chi_2(x), \psi_2(x), \dots, \chi_n(x), \psi_n(x)\big)$$

I claim that the restriction of $\psi$ to the open subset $V = \cup_{i=1}^n V_i$ is an embedding.

First, $\psi|_V$ is an immersion. For if $x \in V$ then $V \in V_i$ for some $i$, and since $m$ of the coordinates of $\psi(x)$ are given by $\psi_i(x)$ and $\psi_i: V_i \to B^m(1)$ is (the restriction of) a coordinate chart, if we had $\psi_* v = 0$ for some $v \in T_x M$ then it would hold that $(\psi_i)_* v = 0$ and so $v = 0$. We must now show that $\psi|_V$ is a homeomorphism to its image. Of course $\psi$ is continuous since all of its coordinates are. To see that $\psi|_V$ is injective, suppose that $\psi(x) = \psi(y)$. For some $i$ we have $x \in V_i$, so $\chi_i(x) = 1$, and so $\chi_i(y) = 1$. We chose $\chi_i$ to be 1 precisely on $\bar{V}_i$, so this forces $y \in \bar{V}_i$. But then since $\psi(x) = \psi(y)$ and $x, y \in \bar{V}_i \subset U_i$ we have $\phi_i(x) = \psi_i(x) = \psi_i(y) = \phi_i(y)$, forcing $x = y$ since $\phi_i$ is a coordinate chart on $U_i$.

Finally we must show that the inverse of $\psi|_V: V \to \psi(V)$ is continuous. Let $x \in V$; we should show that for any neighborhood $W$ of $x$ there is an open set in $\mathbb{R}^{n(m+1)}$, containing $\psi(x)$, whose preimage under $\psi$ is contained in $W$. To do this, let $i$ be such that $x \in V_i$, and let $\epsilon > 0$ be small enough that the preimage under $\phi_i$ of the ball of radius $2\epsilon$ around $\phi_i(x)$ is contained in $W \cap V_i$.

If $\delta > 0$, there is an open set $W' \subset \mathbb{R}^{n(m+1)}$ so that

$$(\psi|_V)^{-1}(W') = \{y \in V | \chi_i(y) > 1-\delta, |\psi_i(y) - \psi_i(x)| < \epsilon\}.$$

If we take $\delta = \frac{2}{2+\epsilon}$, any $y \in (\psi|_V)^{-1}(W')$ belonging to this latter set will obey

$$\begin{aligned}|\phi_i(y) - \phi_i(x)| &= \left|\frac{1}{\chi_i(y)}\phi_i(y) - \phi_i(x)\right| \\ &\leq \left|\frac{1}{\chi_i(y)} - 1\right| |\phi_i(y)| + |\phi_i(y) - \phi_i(x)| < \frac{\epsilon}{2}2 + \epsilon = 2\epsilon\end{aligned}$$

and so $y$ will belong to $W$, as desired.

□

## 2. Vector bundles and tubular neighborhoods

A vector bundle $E$ of rank $k$ over a smooth manifold $M$ is, to be brief (and to leave out some important details), a family of vector spaces $E_x$ parametrized by the points $x \in M$.

More precisely:

*Definition* 2.1. Let $M$ be a smooth manifold, and $k$ a positive integer. A (smooth, real) *vector bundle of rank $k$ over $M$* is a smooth map $\pi\colon E \to M$ where $E$ is a smooth manifold, with the following additional structure

- For all $x \in M$, the preimage $\pi^{-1}(\{x\})$ (also denoted $E_x$) has the structure of a real vector space of dimension $k$.
- There is an open cover $\cup_\alpha U_\alpha$ of $M$ and, for each $\alpha$, a diffeomorphism $\Phi_\alpha\colon \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k$ such that, for each $x \in U_\alpha$, $\Phi_\alpha$ restricts to $E_x$ as a linear isomorphism to the vector space $\{x\} \times \mathbb{R}^k$.

For a definition more closely analogous to our definition of a smooth manifold, and in order to resolve concerns about uniqueness, one could insist that the collection of transition functions is maximal; just as in the smooth manifold case any collection of local trivializations as in the definition can be enlarged in a unique way to a maximal such collection.

The vector space $E_x$ is called the *fiber of $E$ at $x$*, and the $\Phi_\alpha$ are called *local trivializations* of $E$ over $U_\alpha$. Note that the map $\pi\colon E \to M$ is automatically a surjective submersion.

The smooth manifold $E$ carries a distinguished copy of $M$ embedded inside it as the *zero section* $0_M$, whose intersection with each $E_x$ consists of just the zero element of the vector space $E_x$ (in terms of the local trivializations, $0_M = \cup_\alpha \Phi_\alpha^{-1}(U_\alpha \times \{0\})$).

**Example 2.2.** *If $M$ is a smooth $m$-dimensional manifold then the* tangent bundle $\pi\colon TM \to M$ *is a vector bundle of rank $m$. For all intents and purposes we showed this in the first part of the course: $M$ is covered by coordinate charts $(x_1^\alpha, \ldots, x_n^\alpha)\colon U_\alpha \to \mathbb{R}^m$, and for a local trivialization of $TM$ over $U_\alpha$ we can take the inverse of the map $U_\alpha \times \mathbb{R}^k \to \pi^{-1}(U_\alpha)$ which sends $(x, v_1, \ldots, v_m)$ to the tangent vector $\sum_{i=1}^m v_i \frac{\partial}{\partial x_i}$ at $x$.*

**Example 2.3.** *Let $f : N \to M$ be a smooth map between two smooth manifolds, and let $\pi\colon E \to M$ be a vector bundle. We can then form the* pullback bundle $\Pi\colon f^*E \to N$ *as follows. Set theoretically, define*

$$f^*E = \left\{(n, e) \in N \times E | e \in E_{f(n)}\right\}.$$

*We have local trivializations $\Phi_\alpha\colon \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k$; for $x \in U_\alpha$ define $\phi_{\alpha x}\colon E_x \to \mathbb{R}^k$ to be the linear isomorphism such that for $e \in E_x$ we have $\Phi_\alpha(e) = (x, \phi_{\alpha x}(e))$. Now for each $\alpha$ define $\Psi_\alpha\colon \Pi^{-1}(f^{-1}(U_\alpha)) \to f^{-1}(U_\alpha)$ by, for $e \in \Pi^{-1}(n)$ where $n \in f^{-1}(U_\alpha)$, setting $\Psi_\alpha(e) =$*

$(n, \phi_{\alpha f(n)}(e))$. *Now the various $f^{-1}(U_\alpha)$ cover $N$, and it's not hard to check that there is a unique smooth structure imposed on $f^*E$ by requiring that the maps $\Psi_\alpha$ be diffeomorphisms. This gives the map $\Pi\colon f^*E \to N$ the structure of a vector bundle, and we have a commutative diagram*

$$\begin{tikzcd} f^*E \arrow[r] \arrow[d, "\Pi"] & E \arrow[d, "\pi"] \\ N \arrow[r, "f"] & M \end{tikzcd}$$

*where the upper map just sends $(n, e) \in f^*E \subset N \times E$ to $e$, and maps fibers of $f^*E$ isomorphically to fibers of $E$.*

*As an important special case, we can let $f$ be the* inclusion of a submanifold $N \subset M$. *In this case $f^*E$ is more often just denoted by $E|_N$. In particular, we have $TM|_N$, the restriction of the tangent bundle of the ambient manifold $M$ to the submanifold $N$; its fiber over $n \in N$ consists of the whole tangent space $T_nM$.*

**Example 2.4.** *If $N \subset M$ is a submanifold we have the vector bundles $TN$ and $TM|_N$; these give rise to a third vector bundle over $N$, the* normal bundle $\nu_{N,M} \to N$, *whose fiber over a point $n \in N$ is naturally identified with $\frac{T_nM}{T_nN}$. Perhaps the easiest way of constructing this bundle is to make use of the adapted coordinate charts from Theorem 1.4. We cover a neighborhood of $N$ in $M$ by charts $\phi_\alpha\colon U_\alpha \to \mathbb{R}^m$, such that for each $\alpha$ we have $U_\alpha \cap N = \phi_\alpha^{-1}(\{0\} \times \mathbb{R}^n)$. So the $\psi_\alpha := \phi_\alpha|_{U_\alpha \cap N}$ form an atlas for $N$. Let $\pi_1\colon \mathbb{R}^m \to \mathbb{R}^{m-n}$ be the projection onto the first $m-n$ coordinates. So if $v \in T_nN$ where $n \in V_\alpha$, then since $\phi_\alpha$ sends $N$ to $\{0\} \times \mathbb{R}^n$, we will have (borrowing the notation of the previous example) $\pi_1 \circ \phi_{\alpha n} v = 0$. Thus $\pi_1 \circ \phi_{\alpha n}$ descends to a linear isomorphism from $(\nu_{N,M})_n$ to $\mathbb{R}^{m-n}$. Consequently the $\pi_1 \circ \phi_{\alpha n}$ give rise to local trivializations over $V_\alpha$ for $\nu_{N,M}$, confirming that $\nu_{N,M}$ is a vector bundle (again, to get the smooth manifold structure on $\nu_{N,M}$ one can just require that these local trivializations are diffeomorphisms).*

*Note that, if* $\dim M = m$ *and* $\dim N = n$*, then the rank of $\nu_{N,M}$ is $m-n$, and so the dimension of $\nu_{N,M}$ as a smooth manifold is $n+(m-n) = m$, the same as the dimension of the ambient manifold. The tubular neighborhood theorem (Theorem 2.11) will show that, in fact, $\nu_{N,M}$ is diffeomorphic to an open neighborhood of $N$ in $M$.*

*Remark* 2.5. It is possible to formulate the notion of a subbundle of a vector bundle, and then show quite generally that if $F \leq E$ is a subbundle then one can form the quotient bundle $E/F$ (with fiber over $x$ canonically identified with $E_x/F_x$). In the case of a submanifold $N \subset M$, one can show that $TN$ is a subbundle of $TM|_N$, and so the normal bundle can be identified with the quotient bundle of the latter by the former.

*Definition* 2.6. An *orthogonal structure* on a vector bundle $\pi\colon E \to M$ is a map

$$\langle \cdot, \cdot \rangle\colon \cup_{x \in M} (E_x \times E_x) \to \mathbb{R}$$

whose restriction to each $E_x \times E_x$ defines an inner product on $E_x$, and such that whenever $s_1, s_2\colon M \to E$ are two smooth sections (i.e. smooth maps so that $\pi \circ s_i = 1_M$), the map $x \mapsto \langle s_1(x), s_2(x) \rangle$ is smooth.

In other words, an orthogonal structure is a smoothly varying family of inner products on the fibers of $E$. In the case that $E$ is the tangent bundle $TM$ of $M$ an orthogonal structure on $TM$ is called a Riemannian metric on $M$. (Indeed, sometimes one uses the term "Riemannian metric" to refer to an orthogonal structure on any vector bundle.)

**Proposition 2.7.** *If $\pi\colon E \to M$ is a vector bundle there exists a orthogonal structure on $E$.*

*Proof.* Recall that the vector bundle structure on $E$ gives us an open cover $\{U_\alpha\}$ of $M$ and diffeomorphisms $\Phi_\alpha\colon \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k$ which commute with the projections to $U_\alpha$ and restrict to the fibers $E_x$ as a linear isomorphism $\phi_{\alpha x}$ to $\mathbb{R}^k$. So if we denote the standard inner product on $\mathbb{R}^k$ by $(\cdot,\cdot)_0$ we can define $\langle,\cdot,\cdot\rangle_\alpha\colon \cup_{x\in U_\alpha} E_x \times E_x \to \mathbb{R}$ by, for $e_1, e_2 \in E_x$, setting $\langle e_1, e_2\rangle_\alpha = \big(\phi_{\alpha x}e_1, \phi_{\alpha x}e_2\big)_0$.

Now let $\{\chi_\alpha\}$ be a partition of unity subordinate to the cover $\{U_\alpha\}$ and define $\langle\cdot,\cdot\rangle\colon \cup_{x\in E} E_x \times E_x \to \mathbb{R}$ to be equal to $\sum_\alpha \chi_\alpha(x)\langle\cdot,\cdot\rangle_\alpha$, where we have extended $\chi_\alpha(x)\langle\cdot,\cdot\rangle_\alpha$ by zero outside of $U_\alpha$. Since convex combinations of inner products on vector spaces are still inner products, it's easy to see that this satisfies the requirements. $\square$

Note that an orthogonal structure $\langle\cdot,\cdot\rangle$ on $E$ gives rise to a smooth function $\|\cdot\|^2\colon E \to [0,\infty)$ defined by $\|e\|^2 = \langle e, e\rangle$. The square root of this function, $\|\cdot\|\colon E \to \mathbb{R}$, is smooth on the complement of the zero section.

Using orthogonal structures one can show:

**Proposition 2.8.** *Let $\pi\colon E \to M$ and let $U$ be any neighborhood of the zero section $0_M$. Then there is an open set $V$ with $0_M \subset V \subset U$ and a diffeomorphism $\psi\colon E \to V$ which restricts to the identity on $0_M$.*

Thus the entire total space of a vector bundle can be shrunk by a diffeomorphism to an arbitrarily small neighborhood of the zero-section. This is basically a parametrized version of the statement that $\mathbb{R}^k$ is diffeomorphic to an arbitrarily small ball around the origin.

*Proof of Proposition 2.8.* This is easier if $M$ is compact, since then there is $r > 0$ such that the open set $V = E_r = \{e \in E | \|e\|^2 < r^2\}$ is contained in $U$. (Proof: The subset $\bar{E}_1 := \{\|e\|^2 \le 1\}$ is in this case also compact, as one can see by writing it as a union of compact sets obtained from a finite cover by local trivializations, so $\bar{E}_1 \setminus U$ is also compact. If the statement were false then one could find a sequence $e_i \in \bar{E}_1 \setminus U$ with $\|e_i\|^2 \to 0$. But since $\bar{E}_1 \setminus U$ is compact a subsequence of the $e_i$ would converge to some $e$, which would have the contradictory properties that $\|e\| = 0$ and $e \notin U$.) In this case one can choose a diffeomorphism $f\colon [0,\infty) \to [0,1)$ such that $f$ is equal the identity on a neighborhood of the origin (for instance, take $f(t) = \int_0^t g(s)ds$ where $g(s) = 1$ for small $s$, $g(s) > 0$ for all $s$ and $\int_0^\infty g(s) = 1$). Then define $\psi\colon E \to V$ by

$$\psi(e) = rf(\|e\|)\frac{e}{\|e\|}$$

(so $\psi$ rescales each fiber in such a way that a point with norm $n$ now has norm $rf(n)$). This is easily seen to satisfy the required properties (the only place where smoothness either of $\psi$ or $\psi^{-1}$ might seem to be an issue is at the zero section, but in fact we have arranged for $\psi$ to be just scalar multiplication by $r$ on a neighborhood of the zero section).

If $M$ is noncompact then there might not be a single number $r > 0$ as above. However, we shall construct below a smooth function $r\colon M \to (0,\infty)$ so that $V := \{e \in E | e \in E_x \Rightarrow \|e\|^2 < r(x)^2\}$ is contained in the given open set $U$. If we can do this, then a simple modification of the $\psi$ constructed above works: just define $\psi(e) = r(x)f(\|e\|)\frac{e}{\|e\|}$ for $e \in E_x$; since $r$ is smooth and positive (as, therefore, is $\frac{1}{r}$) this $\psi$ will be a diffeomorphism from $E$ to $V$ just as before.

To construct the desired $r\colon M \to (0,\infty)$ we can just proceed as follows. Cover $M$ by open sets $O_\beta$ such that $\bar{O}_\beta$ is compact. Then for each $\beta$ there will, as earlier, be a number $r_\beta > 0$ so that if $x \in \bar{O}_\beta$ and $e \in E_x$ has $\|e\| < r_\beta$ then $e \in U$. Now let $\{\chi_\beta\}$ be a partition of unity subordinate to $\{O_\beta\}$ and let $r(x) = \sum_\beta r_\beta \chi_\beta(x)$. For each $x$, $r(x)$ will then be a convex combination of those $r_\beta$ with $x \in O_\beta$, and hence will be less than or equal to one of them, so any $e \in E_x$ with $\|e\| < r(x)$ will lie in $U$. □

*Definition* 2.9. Let $N \subset M$ be a submanifold. A *tubular neighborhood of $N$ in $M$* consists of an open subset $U \subset M$ with $N \subset U$, and a diffemorphism $\Phi\colon\ \nu_{N,M} \to U$, where $\nu_{N,M}$ is the normal bundle of $N$ in $M$, such that the restriction of $\Phi$ to the zero section $N \cong 0_N \subset \nu_{N,M}$ is the identity map to $N$.

*Remark* 2.10. In view of Proposition 2.8, to construct a tubular neighborhood it is enough to construct a diffeomorphism $\Phi'\colon\ U' \to U$ restricting as the identity on $N$, where $U' \subset \nu_{N,M}$ is some neighborhood of the zero section. For then we can find a subneighborhood $V \subset U'$ and a diffeomorphism $\psi\colon\ \nu_{N,M} \to V$ as in Proposition 2.8, and then $\Phi' \circ \psi$ will give a tubular neighborhood (with image $\Phi'(U')$, which will still be an open neighborhood of $N$ in $M$).

The rest of this section will be concerned with proving the following:

**Theorem 2.11.** *If $N \subset M$ is any compact submanifold then there exists a tubular neighborhood of $N$ in $M$.*

(In fact, the compactness assumption is not strictly necessary–its main role in the proof given here will be to allow us to embed a neighborhood of $N$ in $M$ into $\mathbb{R}^q$ for some $q$, and as mentioned after Theorem 1.9 this can be done without the compactness assumption. Near the end of the proof we will also use the compactness of $N$ to find the limit of a sequence, but one can get around this as long as one arranges for the embedding of $M$ into $\mathbb{R}^q$ to be proper, as can be arranged in the Whitney embedding theorem.) To construct the tubular neighborhood, one needs some systematic way of “moving in directions normal to $N$ in $M$.” There are two common ways of doing this—either by choosing a Riemannian metric on $M$ and using the theory of geodesics, or by embedding a neighborhood of $N$ in $M$ into Euclidean space and using the special structure of Euclidean space. To avoid a digression into Riemannian geometry, we'll take the latter approach.

Throughout the following discussion, for $x \in \mathbb{R}^q$ we will make the standard identification of $T_x\mathbb{R}^q$ with $\mathbb{R}^q$ (and so if $X \subset \mathbb{R}^q$ is a submanifold and $x \in X$ then $T_xX$ is identified with a subspace of $\mathbb{R}^q$).

To begin the proof of the theorem note that we may as well replace $M$ by a small neighborhood of the compact submanifold $N$, and then by Theorem 1.9 (applied with $K = N$) we can assume that $M$ is embedded in $\mathbb{R}^q$. We can then define

$$\tilde{\nu}_{M,\mathbb{R}^q} = \{(x,v) \in M \times \mathbb{R}^q | v \in T_xM^\perp\}$$

and

$$\tilde{\nu}_{N,M} = \{(x,v) \in N \times \mathbb{R}^q | v \in (T_xM) \cap (T_xN)^\perp\}.$$

Note that there is a bijection $\alpha_{M,\mathbb{R}^q}\colon\ \tilde{\nu}_{M,\mathbb{R}^q} \to \nu_{M,\mathbb{R}^q}$, taking $(x,v)$ to the equivalence class of $v$ in $T_x\mathbb{R}^q/T_xM$. Similarly there is a bijection $\alpha_{N,M}\colon\ \tilde{\nu}_{N,M} \to \nu_{N,M}$ sending $(x,v)$ to its equivalence class in $T_xM/T_xN$. These bijections commute with the projections to $M$ (in the case of $\alpha_{M,\mathbb{R}^q}$) or $N$ (in the case of $\alpha_{N,M}$). It should at least appear that $\tilde{\nu}_{M,\mathbb{R}^q}$ is a vector bundle over $M$, and likewise that $\tilde{\nu}_{N,M}$ is a vector bundle over $N$, and that these are isomorphic to the respective normal bundles. This is indeed true, but so far we have not even shown that $\tilde{\nu}_{M,\mathbb{R}^q}$ and $\tilde{\nu}_{N,M}$ are smooth manifolds. We now remedy this:

**Lemma 2.12.** *Let $M \subset \mathbb{R}^q$ be a submanifold, and $N \subset M$ a submanifold. Then $\tilde{\nu}_{M,\mathbb{R}^q}$ and $\tilde{\nu}_{N,M}$ are smooth manifolds, and the bijections $\alpha_{M,\mathbb{R}^q}\colon\ \tilde{\nu}_{M,\mathbb{R}^q} \to \nu_{M,\mathbb{R}^q}$ and $\alpha_{N,M}\colon\ \tilde{\nu}_{N,M} \to \nu_{N,M}$ are diffeomorphisms.*

*Proof.* Actually the statements about $\tilde{\nu}_{M,\mathbb{R}^q}$ are, after renaming, just special cases of those about $\tilde{\nu}_{N,M}$, but for clarity's sake we prove the results about $\tilde{\nu}_{M,\mathbb{R}^q}$ first. To show that $\tilde{\nu}_{M,\mathbb{R}^q}$ is a smooth manifold it suffices to show that for any point $m_0 \in M$ there is a neighborhood $U \subset M$ of $m_0$ so that $\tilde{\nu}_{M,\mathbb{R}^q} \cap (U \times \mathbb{R}^q)$ is a submanifold of $U \times \mathbb{R}^q$.

To do this, note that on a sufficiently small neighborhood $U \subset M$ we there will be smooth functions $a_{ij}\colon U \to \mathbb{R}$ $(1 \le i \le m, q \le j \le q)$ so that, for all $X \in U$,

$$T_x M = span \left\{ \sum_{j=1}^{q} a_{ij}(m) \frac{\partial}{\partial x_j} : 1 \le i \le m \right\}$$

(for instance, one could take an adapted coordinate chart as in Theorem 1.4 and use for $\sum a_{ij} \frac{\partial}{\partial x_j}$ the vector fields that are mapped by the coordinate chart to the standard coordinate vector fields on $\mathbb{R}^m \times \{0\}$) So at each $x \in U$ the matrix $A(x) = \{a_{ij}(x)\}$ has full rank $m$. Define

$$\begin{aligned} F\colon\ & U \times \mathbb{R}^q \to \mathbb{R}^m \\ (x, v_1, \ldots, v_q) &\mapsto \left( \sum_{j=1}^{k} a_{1j}(x) v_j, \ldots, \sum_{j=1}^{k} a_{mj}(x) v_j \right). \end{aligned}$$

Identifying $T_{(x,v)}(M \times \mathbb{R}^q)$ with $T_x M \oplus \mathbb{R}^q$, we see that the linearization $F_*\colon T_{(x,v)}(M \times \mathbb{R}^q) \to \mathbb{R}^m$ has $F_*(0, w) = A(x)w$. In particular since $A(x)$ has full rank, $F_*$ is surjective. By Theorem 1.8 this proves that $\tilde{\nu}_{M,\mathbb{R}^q} \cap (U \times \mathbb{R}^q)$ is a submanifold of $U \times \mathbb{R}^q$ for each member $U$ of an open cover of a neighborhood of $M$ in $\mathbb{R}^q$, and hence that $\tilde{\nu}_{M,\mathbb{R}^q}$ is a smooth manifold.

Moreover, inspection of the coordinate charts constructed in the proof of Theorem 1.8 shows that the smooth structure on $\tilde{\nu}_{M,\mathbb{R}^q}$ is consistent with that of $\nu_{M,\mathbb{R}^q}$ under the obvious bijection between them. Indeed, in the intersection of $\tilde{\nu}_{M,\mathbb{R}^q}$ with $U \times \mathbb{R}^q$ where $U$ is a sufficiently small open set as in the previous paragraph, we can define a coordinate system whose first $m$ coordinates (parametrizing $M$) are the same as those of an adapted coordinate chart for $M \subset \mathbb{R}^q$, and whose last $q - m$ coordinates depend only on the $\mathbb{R}^q$ factor. It's not hard to see that such a coordinate chart is diffeomorphic via $\alpha_{M,\mathbb{R}^q}$ to a corresponding local trivialization for $\nu_{M,\mathbb{R}^q}$ as described in Example 2.4. So since the bijection $\alpha_{M,\mathbb{R}^q}$ restricts to each member of an open cover as a diffeomorphism it is a diffeomorphism.

Now we turn to the slightly more complicated case of $\tilde{\nu}_{N,M}$. In this case, for any $n_0 \in N$ we can find a neighborhood of $n_0$ in $\mathbb{R}^q$ and smooth functions $a_{ij}$ $(1 \le n + q - m,\ 1 \le j \le q)$ on $U$ so that, for each $x \in U$, $T_x N$ is spanned by the $\sum_{j=1}^{q} a_{ij}(x) \frac{\partial}{\partial x_j}$ for $1 \le i \le n$, and $T_x M^\perp$ is spanned by the $\sum_{j=1}^{q} a_{ij}(x) \frac{\partial}{\partial x_j}$ for $n+1 \le i \le n+q-m$. Namely, as before we can use an adapted coordinate chart for the vector fields spanning $TN|_U$, while for the vector fields spanning $TM^\perp|_U$ we can start with a similar such basis of vector fields spanning $T_x M$ at every $x \in U$, extend this to a basis for $\mathbb{R}^q$ (say using vector fields with constant coefficients, and perhaps shrinking $U$ in the process), and then modify this basis using the Gram-Schmidt procedure to get a basis for all of $T_x \mathbb{R}^q$ at every point of $U$ consisting of smooth vector fields, the last $q - m$ of which span $T_x M^\perp$ at every $x \in U$.

Now since $T_x N \cap T_x M^\perp = \{0\}$ for all $x \in M$ (as $TN \subset TM$), our entire set of vector fields $\left\{ \sum_{j=1}^{q} a_{ij}(x) \frac{\partial}{\partial x_j} : 1 \le i \le n + q - m \right\}$ is linearly independent at each $x \in U$. So just as before we can define

$$G\colon U\times\mathbb{R}^q\to\mathbb{R}^{n+m-q}$$
$$(x,v_1,\ldots,v_q)\mapsto\left(\sum_{j=1}^k a_{1j}(x)v_j,\ldots,\sum_{j=1}^k a_{mj}(x)v_j\right)$$

The preimage of 0 under this map consists of those pairs $(x,v)$ where $v$ is perpendicular both to the subspace $T_xN$ and to the subspace $T_xM^\perp$, *i.e.* where $v\in T_xM\cap T_xN^\perp$, so $G^{-1}(\{0\})=\{(x,v)\in\tilde{\nu}_{N,M}|x\in U\}$. As in the case of $\tilde{\nu}_{M,\mathbb{R}^q}$, the linear independence of the vector fields that we have chosen implies that $G_*$ is surjective, so $G^{-1}(\{0\})$ is a submanifold, and indeed following the proof of Theorem 1.8 we can take a coordinate system on $G^{-1}(\{0\})$ so that the first $n$ coordinates depend only on the $N$ factor and the last $m-n$ depend only on the $\mathbb{R}^q$ factor, so that in this coordinate system the projection $\tilde{\nu}_{N,M}\to N$ appears as the projection onto the first $n$ coordinates. Allowing $U$ to vary through sufficiently small open neighborhoods in $\mathbb{R}^q$ of points of $N$ produces an atlas for $\tilde{\nu}_{N,M}$ each member of which can be seen as a local trivialization for the bundle $\tilde{\nu}_{N,M}\to N$. Once again, these trivializations are compatible under the bijection $\alpha_{N,M}$ with the standard normal bundle trivializations as given in Example 2.4, completing the proof. □

The following (when combined with Proposition 2.8) proves the tubular neighborhood theorem for submanifolds of $\mathbb{R}^q$, and will also be used in the more general case. To prepare for the statement, note that the space

$$\tilde{\nu}_{M,\mathbb{R}^q}=\{(x,v)|x\in M,v\in T_xM^\perp\}$$

of the previous lemma contains a distinguished "zero section" consisting of points of form $(x,0)$.

**Lemma 2.13.** *If $M\subset\mathbb{R}^q$ is a submanifold, define*

$$\epsilon_{M,\mathbb{R}^q}\colon\ \tilde{\nu}_{M,\mathbb{R}^q}\to\mathbb{R}^q$$

*by*

$$\epsilon_{M,\mathbb{R}^q}(x,v)=x+v.$$

*Then there is a neighborhood $V$ of the zero section of $\tilde{\nu}_{M,\mathbb{R}^q}$ such that $\epsilon_{M,\mathbb{R}^q}$ restricts to $V$ as a diffeomorphism to its image, which is an open neighborhood of $M$ in $\mathbb{R}^q$.*

*Proof.* Writing $0_M=\{(m,0)\}$ for the zero section of $\tilde{\nu}_{M,\mathbb{R}^q}$, for any $(x,0)\in 0_M$ the tangent space $T_{(x,0)}\tilde{\nu}_{M,\mathbb{R}^q}$ splits naturally as $T_xM\oplus(T_xM)^\perp$, where the first factor is tangent to $0_M$ and the second is tangent to the fibers of the bundle projection $\tilde{\nu}_{M,\mathbb{R}^q}$. Of course, since $T_xM$ is identified via the embedding as a subspace of $\mathbb{R}^q$, $T_xM\oplus(T_xM)^\perp$ in turn may be identified with all of $\mathbb{R}^q$. As should be clear from the definition of $\epsilon_{M,\mathbb{R}^q}$, with respect to these identifications the linearization of $\epsilon_{M,\mathbb{R}^q}$ at $(x,0)$ is just the identity from $\mathbb{R}^q$ to itself and in particular is invertible. So by the inverse function theorem every point $0_M$ has a neighborhood to which $\epsilon_{M,\mathbb{R}^q}$ restricts as a diffeomorphism to its image.

If $x\in M$ and $\delta>0$, define

$$V_{x,\delta}=\{(y,v)\in\tilde{\nu}_{M,\mathbb{R}^q}||y-x|+|v|<\delta\},$$

where $|\cdot|$ refers to the standard distance in Euclidean space $\mathbb{R}^q$. By the previous paragraph, for any $x\in M$ there is $\delta>0$ so that $\epsilon_{M,\mathbb{R}^q}|_{V_{x,\delta}}$ is a diffeomorphism to its image. So define a function $\delta\colon M\to\mathbb{R}$ by setting $\delta(x)$ equal to the supremum of all numbers $\delta$ such that $\epsilon_{M,\mathbb{R}^q}|_{V_{x,\delta}}$ is a diffeomorphism to its image. So evidently $\delta(x)>0$ for all $x$.

I now claim that $\delta\colon M \to \mathbb{R}$ is continuous. Indeed, one has a relationship

$$\delta(y) \geq \delta(x) - |x-y|,$$

resulting from the fact that $V_{y,\delta-|x-y|} \subset V_{x,\delta}$ for all $\delta > 0$. Combining this relationship with the same one where $x$ and $y$ are reversed shows that

$$|\delta(x) - \delta(y)| \leq |x-y|,$$

so $\delta$ is indeed continuous. Now define

$$V = \left\{ (x,v) \in \tilde{\nu}_{M,\mathbb{R}^q} \,\middle|\, |v| < \frac{1}{3}\delta(x) \right\}$$

$V$ is then an open subset (since $(x,v) \mapsto |v| - \frac{1}{3}\delta(x)$ is continuous and $V$ is the preimage of an open set under this map), and we will show that it has the property stated in the lemma.

The main issue is to show that $\epsilon_{M,\mathbb{R}^q}|_V$ is injective. So we must show that if $(x,v),(y,w) \in \tilde{\nu}_{M,\mathbb{R}^q}$ with $|v| < \frac{\delta(x)}{3}$, $|w| < \frac{\delta(y)}{3}$, and $x+v = y+w$, then $(x,v) = (y,w)$ Without loss of generality assume that $\delta(x) \leq \delta(y)$. Now the assumed relation $x+v = y+w$ is equivalent to $x - y = w - v$. But $|w-v| \leq |w| + |v| < 2\delta(x)/3$, and so

$$|x-y| + |w| = |w-v| + |w| < \delta(x).$$

So for some $\delta < \delta(x)$ we have $(x,v),(y,w) \in V_{x,\delta}$. But by the definition of $\delta(x)$, $\epsilon_{M,\mathbb{R}^q}$ restricts injectively to $V_{x,\delta}$ for all $\delta < \delta(x)$. So indeed $(x,v) = (y,w)$.

So we have shown that $\epsilon_{M,\mathbb{R}^q}$ restricts injectively to $V$. By the construction of $V$ and by what was done at the start of the proof, $V$ is covered by open sets on which $\epsilon_{M,\mathbb{R}^n}$ is a local diffeomorphism, and so $\epsilon_{M,\mathbb{R}^n}|_V$ is also continuous and open. Thus $\epsilon_{M,\mathbb{R}^q}|_V$ is a diffeomorphism to its image, which is open in $\mathbb{R}^q$. □

**Corollary 2.14.** *Let $M \subset \mathbb{R}^q$ be a submanifold. Then there is an open neighborhood $W$ of $M$ and a smooth map $r\colon W \to M$ so that $r|_M$ is the identity.*

(Indeed, $r$ can be taken to be a deformation retraction, as you can check.)

*Proof.* Let $V$ be a neighborhood of $0_M \subset \tilde{\nu}_{M,\mathbb{R}^q}$ as in Lemma 2.13, so that $\epsilon_{M,\mathbb{R}^q}\colon V \to \mathbb{R}^q$ is a diffeomorphism to its image. Denote this image by $W \subset M$. Then where $\pi\colon \tilde{\nu}_{M,\mathbb{R}^q} \to M$ is the bundle projection and where we identify $M$ with $0_M$, define $r\colon W \to M$ by $r = \pi \circ (\epsilon_{M,\mathbb{R}^q}|_V)^{-1}$. Since $\epsilon_{M,\mathbb{R}^q}$ restricts to $0_M \cong M$ as the identity, $r$ is easily seen to satisfy the desired property. □

*End of the proof of Theorem 2.11.* We let $N \subset M$ be any compact submanifold, and by replacing $M$ by a sufficiently small open set containing $N$ and applying Theorem 1.9 we assume $M$ to be embedded as a submanifold of $\mathbb{R}^q$. Where again

$$\tilde{\nu}_{N,M} = \{(x,v) \in N \times \mathbb{R}^q | v \in T_xM \cap (T_xN)^\perp\}$$

define $f_0\colon \tilde{\nu}_{N,M} \to \mathbb{R}^q$ by $f(x,v) = x + v$. Where $r$ and $W$ is as in Corollary 2.14, let $U_0 = f^{-1}(W)$, and define

$$f\colon U_0 \to M \qquad \text{by } f = r \circ f_0.$$

We have the zero section $0_N = \{(x,0)\} \subset \tilde{\nu}_{N,M}$; clearly for $(x,0) \in 0_N$, $f_0(x,0) = x \in N \subset M$ and so $f(x,0) = x$ also. Moreover, for $(x,0) \in 0_N$, $T_{(x,0)}U_0$ splits (compatibly with the splitting of $N \times \mathbb{R}^q$) as a direct sum $T_xN \oplus (T_xM \cap (T_xN)^\perp)$ (which is the same as $T_xM$), and with respect to this splitting the linearization $(f_0)_*\colon T_{(x,0)}U_0 \to T_x\mathbb{R}^q \cong \mathbb{R}^q$ acts as the inclusion. Now since the map $r$ acts as the identity on $M$, and since $(f_0)_*$ sends $T_{(x,0)}U_0$ isomorphically to $T_xM \leq \mathbb{R}^q$, it follows by the chain rule that $f_* = r_* \circ (f_0)_*$ also sends $T_{(x,0)}U_0$ isomorphically to $T_xM$ for all $(x,0) \in 0_N$. Thus around any point $(x,0) \in 0_N \subset U_0$ there is a neighborhood $V_x$ to which $f: U_0 \to M$ restricts as a diffeomorphism to its image, which is open in $M$.

So much like the proof of Lemma 2.13 we should now show that, perhaps after shrinking the neighborhood $U_0$ of $0_N$ to some smaller neighborhood $U_1$, $f$ restricts injectively to $U_1$. For this we exploit the compactness of $N$. If there were no neighborhood of $0_N$ to which $f$ restricted injectively, we could find $(x_i, v_i), (y_i, w_i) \in U_0$ such that $(x_i, v_i) \neq (y_i, w_i)$ and $v_i, w_i \to 0$ but $f(x_i, v_i) = f(y_i, w_i)$ for all $i$. After passing to subsequences, the sequences $x_i, y_i$ would converge in $N$ by compactness, say to $x$ and $y$, and we would have $f(x,0) = f(y,0)$ and hence $x = y$. But then $(x_i, v_i)$ and $(y_i, w_i)$ would eventually both lie in the neighborhood $V_x$ from the previous paragraph, contradicting the fact that $f$ is injective on that neighborhood. This contradiction shows that there is some neighborhood $U_1$ of $0_N$, which we may as well take to be contained in $\cup_{x\in N} V_x$, such that $f|_{U_1}$ is injective.

Since $f: U_1 \to M$ is injective and $U_1$ is covered by sets to which $f$ restricts as a diffeomorphism to its image, it follows that $f: U_1 \to f(U_1)$ is a global diffeomorphism to its image (since smoothness of $f$ and of $f^{-1}$ can be checked on these open sets).

This shows that a neighborhood $U_1$ of $0_N$ in $\tilde{\nu}_{N,M}$ is diffeomorphic to a neighborhood of $N$ in $M$ by a diffeomorphism restricting to the identity on $M$. By Remark 2.10 and the fact that $\tilde{\nu}_{N,M}$ is diffeomorphic to $\nu_{N,M}$ by a diffeomorphism acting as the identity on the zero section, this suffices to yield a tubular neighborhood $\Phi\colon \nu_{N,M} \to M$. □

## 3. VECTOR FIELDS AND FLOWS

The following is a basic result from the theory of ordinary differential equations:

**Theorem 3.1.** *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a compactly supported smooth function. Then for any $x_0 \in \mathbb{R}^n$ there is a unique solution $\gamma_{x_0}\colon \mathbb{R}^{\to}\mathbb{R}^n$ to the initial value problem*

$$\gamma'(t) = F(\gamma(t))$$
$$\gamma(0) = x_0$$

*Moreover if $I \subset \mathbb{R}$ is any open interval, if $t_0 \in \mathbb{R}$, and if $\gamma\colon I \to \mathbb{R}$ obeys $\gamma'(t) = F(\gamma(t))$ and $\gamma(t_0) = \gamma_{x_0}(t_0)$, then $\gamma = \gamma_{x_0}|_I$. Furthermore, the map*

$$\begin{aligned} \Phi\colon \mathbb{R} \times \mathbb{R}^n &\to \mathbb{R}^n \\ (t,x) &\mapsto \gamma_x(t) \end{aligned}$$

*is a smooth map.*

*Sketch of proof.* (See [Lee, Chapter 17] for details.) First of all, suppose that we can show that there is $\epsilon > 0$ such that for every $x_0 \in \mathbb{R}^n$ and every $t_0 \in \mathbb{R}$ there is a solution $\gamma_{x_0}\colon (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{R}^n$ to $\gamma'(t) = F(\gamma(t))$ with $\gamma'(t_0) = x_0$, and such that any other solution $\gamma$ on some subinterval of $(t_0 - \epsilon, t_0 + \epsilon)$ such that $\gamma(t_1) = \gamma_{x_0}(t_1)$ for some $t_1$ coincides is equal to $\gamma_{x_0}$ everywhere. From this the existence of the *all-time* solution $\gamma_{x_0}\colon \mathbb{R} \to \mathbb{R}^n$ would follow. Indeed, we could initially apply the result to get a solution $\gamma_0$ on $(-\epsilon, \epsilon)$ with $\gamma_0(0) = x_0$. But we could then also get a solution $\gamma_1$ on $(-ep/2, 3\epsilon/2)$ with $\gamma_1(\epsilon/2) = \gamma_0(\epsilon/2)$. The uniqueness statement would then force $\gamma_1$ and $\gamma_0$ to be equal everywhere that they are both defined; hence they would combine to give a solution (still denoted $\gamma_0$) on $(-\epsilon, 3\epsilon/2)$. But there would also be a solution $\gamma_2$ on $(0, 2\epsilon)$ with $\gamma_2(\epsilon) = \gamma_0(\epsilon)$, and by uniqueness this then coincides with $\gamma_0$ everywhere, allowing the domain of $\gamma_0$ to be extended to $(-\epsilon, 2\epsilon)$. This can be repeated indefinitely, and the union of all of the solutions so obtained gives a map $\gamma_{x_0} \to \mathbb{R} \to \mathbb{R}^n$.

In other words, existence and uniqueness for all time (*i.e.*, all of the theorem except the last sentence) will follow if we can prove existence and uniqueness on all time intervals $I$ of length at most $2\epsilon$ for some fixed $\epsilon$. We will do this by converting the differential equation to a fixed point problem and applying the contractive mapping principle. Namely, observe that the fundamental theorem of calculus implies that the following two statements about a map $\gamma\colon I \to \mathbb{R}^n$, where $I$ is an interval containing a point $t_0$ are equivalent:

$$\gamma \text{ is a differentiable map such that } \gamma'(t) = X(\gamma(t)) \text{ and } \gamma(t_0) = x_0$$

and

$$\gamma \text{ is a continuous map such that } \gamma(t) = x_0 + \int_{t_0}^{t} F(\gamma(s))ds \text{ for all } t \in I$$

Let $C(I, \mathbb{R}^n)$ denote the space of continuous functions from $I$ to $\mathbb{R}^n$, endowed with the uniform ("sup") norm. A standard fact in analysis is that $C(I, \mathbb{R}^n)$ is a Banach space (basically this is because a uniform limit of continuous functions is continuous). Define

$$\mathscr{A}\colon C(I, \mathbb{R}^n) \to C(I, \mathbb{R}^n)$$

by

$$\big(\mathscr{A}\gamma\big)(t) = x_0 + \int_{t_0}^{t} F(\gamma(s))ds$$

Now $F$ was assumed compactly supported and smooth—in particular $F$ is Lipschitz (actually $F$ being Lipschitz is all that is needed for the conclusion of the theorem), *i.e.*, there is $C$ such that $|F(x) - F(y)| \le C|x - y|$ for all $x, y \in \mathbb{R}^n$ ($C$ can be taken to be the maximum norm of the gradient of $F$). An easy computation shows that, *provided the length of $I$ is less than $\frac{1}{C}$*, the above map $\mathscr{A}$ is *contractive*, *i.e.*, there is $r < 1$ such that $\|\mathscr{A}\gamma - \mathscr{A}\eta\| \le r\|\gamma - \eta\|$. But the contractive mapping mapping principle asserts that a contractive map from a Banach space to itself always has a unique fixed point (if $\gamma_0$ is chosen arbitrarily and we define $\gamma_i = \mathscr{A}\gamma_{i-1}$, the fixed point is $\lim_{i=1}^{\infty} \mathscr{A}\gamma_i$). This precisely gives the desired existence and uniqueness of solutions on sufficiently short time intervals, and hence by the first paragraph proves all of the theorem except the last sentence.

The smoothness of $\Phi$ relies on some somewhat subtle estimates which can be found in [Lee]; I'll just prove the fact that $\Phi$ is continuous, which is a first step in the smoothness proof. First of all observe that for any smooth $u\colon \mathbb{R} \to \mathbb{R}^n$ such that $u(t)$ is nonzero for all $t$, one has, using the Cauchy–Schwarz inequality and the chain and product rules,

$$\frac{d}{dt}|u(t)| = \frac{d}{dt}\sqrt{u(t)\cdot u(t)} = \frac{2u(t)\cdot u'(t)}{2\sqrt{u(t)\cdot u(t)}} \le \frac{|u(t)||u'(t)|}{|u(t)|} = \left|\frac{du}{dt}\right|. \tag{1}$$

Now if $x, y \in \mathbb{R}^n$ are distinct points, by uniqueness of solutions we have $\gamma_x(t) \ne \gamma_y(t)$ for all $t$, so we can apply (1) with $u(t) = \gamma_x(t) - \gamma_y(t)$ to get

$$\begin{aligned}\frac{d}{dt}|\gamma_x(t) - \gamma_y(t)| &\le \left|\frac{d}{dt}\big(\gamma_x(t) - \gamma_y(t)\big)\right| = \big|F(\gamma_x(t)) - F(\gamma_y(t))\big| \\ &\le C|\gamma_x(t) - \gamma_y(t)|\end{aligned}$$

where as before $C$ is the Lipschitz constant of $F$. Dividing by $|\gamma_x(t) - \gamma_y(t)|$ (which as noted earlier is nowhere zero) and recalling the identity $\frac{d}{dt}(\ln f) = \frac{f'}{f}$ then gives

$$\frac{d}{dt}\ln|\gamma_x(t) - \gamma_y(t)| \le C.$$

The Fundamental Theorem of Calculus (and then exponentiation of both sides) then shows that

$$|\gamma_x(t) - \gamma_y(t)| \le e^{Ct}|\gamma_x(0) - \gamma_y(0)|,$$

*i.e.*

$$|\gamma_x(t)-\gamma_y(t)| \le e^{Ct}|x-y|.$$

This shows that $\Phi$ is continuous as a function of $x$ for fixed $t$. To take into account the varying of $t$ we can just note that, if $D>0$ is such that $|F(x)| \le D$ for all $x$ (such $D$ exists since we assumed $F$ was compactly supported), then $|\gamma_x(s)-\gamma_x(t)| \le D|s-t|$. So we get

$$|\Phi(s,x)-\Phi(t,y)| \le |\Phi(s,x)-\Phi(t,x)| + |\Phi(t,x)-\Phi(t,y)| \le D|s-t| + e^{Ct}|x-y|,$$

and this proves that $\Phi$ is continuous at the (arbitrary) point $(t,y)$.

As mentioned earlier, smoothness as opposed to continuity takes more work; I'll just mention that part of the idea is to differentiate the equation

$$\frac{\partial}{\partial t}(\Phi(t,x)) = F(\Phi(t,x))$$

which is satisfied by $\Phi$ with respect to $t$ and/or $x$, in order to get a differential equation satisfied by a partial derivative of $\Phi$; one can work inductively on the order of the derivative.

□

In other words, for compactly supported vector fields $F$ on $\mathbb{R}^n$, there is always a unique integral curve of a vector field passing through any given point, and this curve varies smoothly with the point. This can easily be exported to smooth manifolds to yield the following corollary:[1]

**Corollary 3.2.** *Let $M$ be a smooth manifold and let $X$ be a compactly supported vector field on $M$. Then there is a unique family of diffeomorphisms, parametrized by $t \in \mathbb{R}$,*

$$\phi_X^t \colon M \to M$$

*such that for all $m \in M$ we have*

$$\phi_X^0(m) = m \text{ and } \frac{d}{dt}\phi_X^t(m) = X(\phi_X^t(m)).$$

*These diffeomorphisms obey*

$$\phi_X^t \circ \phi_X^s = \phi_X^{t+s} \tag{2}$$

*and the map $(t,m) \mapsto \phi_X^t(m)$ is smooth.*

The equation 2 should be easy to see: both sides represent the effect of starting at a point and flowing along the flow of the vector field for a time $t+s$. This equation is part of what leads to the conclusion that the $\phi_X^t$ are diffeomorphisms rather than just being smooth maps, since evidently $\phi_X^{-t}$ is an inverse to $\phi_X^t$. The family $\{\phi_X^t\}$ is called the *flow* of the vector field $X$.

*Remark* 3.3. It is sometimes useful to allow the vector field $X$ to itself depend on $t$, *i.e.* one can have a family of vector fields $X_t$ varying with the parameter $t$. As long as this dependence is smooth and $X_t$ has a uniform Lipschitz constant for all $t$ then the proofs of Theorem 3.1 and Corollary 3.2 go through essentially without change in order to show that one still gets diffeomorphisms $\phi_X^t$ so that $\phi_X^0 = id_M$ and $\frac{d}{dt}\phi_X^t(m) = X_t(\phi_X^t(m))$. In fact, any smooth path of diffeomorphisms starting at the identity can be described as such a "time-dependent" flow—given such a path $\phi_t$ one can define $X_t(m) = \frac{d\phi_t}{dt}(\phi_t^{-1}(m))$ and, more or less tautologically, the flow of $X_t$ will recover $\phi_t$. Of course, for one of these time-dependent flows the homomorphism property (2) typically will not hold.

[1]One can approach the derivation of the corollary from Theorem 3.1 in either of a couple of different ways, either by directly working in local coordinate charts or by embedding $M$ in $\mathbb{R}^q$ for some large $q$ and using the tubular neighborhood theorem to construct a vector field on $\mathbb{R}^q$ which restricts to $M$ as the given vector field $X$; details are left to you

*Remark* 3.4. If one drops the hypothesis that $X$ is compactly supported (or, more generally, Lipschitz in a suitable sense) then Corollary 3.1 will no longer be true as stated. However a "local" statement can be made: for any $m \in M$ there will still be $\epsilon > 0$ and a neighborhood $U$ of $m$ in $M$ on which there exists a unique "partial flow"

$$(-\epsilon, \epsilon) \times U \to M$$
$$(t, y) \mapsto \phi_X^t(y)$$

so that $\frac{d}{dt}\phi_X^t(y) = X(\phi_X^t(y))$ and $\phi_X^0(y) = y$. In other words, while a long-time existence result along the lines of Theorem 3.1 will typically fail, one still has uniqueness and short-time existence, for a time $\epsilon$ which depends on the point of interest in $M$.

The classic example of failure of long-time existence comes in the case $M = \mathbb{R}$, where the vector field $X$ is given by $X(x) = x^2$. Thus the relevant differential equation is

$$\frac{dx}{dt} = x^2.$$

This equation can be solved by separation of variables to yield, where $x_0 = x(0)$,

$$x(t) = \frac{x_0}{1 - x_0 t}.$$

So for any $x_0$ we have a unique integral curve $x(t)$ through $x_0$, but this solution "blows up in finite time"—it ceases to be well-defined at time $t = \frac{1}{x_0}$ (but is a perfectly good solution until then).

3.1. **The Lie Derivative.** Given a (say compactly supported for convenience, but this is not really necessary for this section) vector field $X$ on a smooth manifold $M$, the flow of $X$ as described above provides a path of diffeomorphisms $\phi_X^t \colon M \to M$. The *Lie derivative* of a vector field or of a differential form along $X$ is meant to be a measurement of how that vector field or differential form changes as one moves along the flow of $X$.

We'll start with the definition for vector fields:

*Definition* 3.5. Let $X$ and $Y$ be vector fields on $M$ The Lie derivative of $Y$ along $X$ is the vector field $\mathscr{L}_X Y$ whose value at a point $m \in M$ is the element of $T_m M$ defined by

$$(\mathscr{L}_X Y)_m = \lim_{t \to 0} \frac{(\phi_X^{-t})_*(Y_{\phi_X^t(m)}) - Y_m}{t}$$

Note that this definition makes sense: recall that $\phi_X^{-t}$ is inverse to $\phi_X^t$, and therefore we have a map $(\phi_X^{-t})_* \colon T_{\phi_X^t(m)} M \to T_m M$. Thus the two tangent vectors in the numerator belong to the same vector space, namely $T_m M$.

There is a similar definition for differential forms, but actually it can be rewritten in a somewhat simpler way because one moves from $T^*_{\phi_X^t(m)} M$ to $T^*M$ by pullback by the map $\phi_X^t$. So if $\omega \in \Omega^p(M)$ and we wish to compare $\omega_{\phi_X^t(m)}$ to $\omega_m$ we can hit the first of these with the transpose of the linearization of $\phi_X^t$. But recall that pullback of differential forms was defined precisely to so that the differential form $((\phi_X^t)^*\omega)_m$ would be equal to the result of applying the transpose of the linearization of $\phi_X^t$ to $\omega_{\phi_X^t(m)}$. So we define:

*Definition* 3.6. Let $\omega \in \Omega^p(M)$ be a differential form and let $X$ be a vector field on $M$. The Lie derivative of $\omega$ along $X$ is the differential $p$-form defined by

$$\mathscr{L}_X \omega = \lim_{t \to 0} \frac{(\phi_X^t)^*\omega - \omega}{t} = \left.\frac{d}{dt}\right|_{t=0} (\phi_X^t)^*\omega$$

*Remark* 3.7. It can be shown (either directly from the definition or from the formulas that we are about to prove) that our definitions of the Lie derivative of a vector field and of a differential form are compatible in the following sense. Suppose that $X, Y_1, \dots, Y_p$ are vector fields and $\omega$ is a $p$-form. Then $\omega(Y_1, \dots, Y_p)$ is a smooth function, *i.e.* a 0-form, so we can take its Lie derivative along $X$. On the other hand we can take the Lie derivatives along $X$ of $\omega$ and of the $Y_i$. These obey the Leibniz rule:

$$\mathscr{L}_X \left(\omega(Y_1, \dots, Y_p)\right) = (\mathscr{L}_X \omega)(Y_1, \dots, Y_p) + \sum_{j=1}^{p} \omega\left(Y_1, \dots, Y_{j-1}, \mathscr{L}_X Y_j, Y_{j+1}, \dots, Y_p\right).$$

The definition of the Lie derivative along $X$ makes it look somewhat impossible to compute; however we will presently give formulas which allow it to be quite easily computed from local coordinate expressions of $X$ and of the object being differentiated. We start with 0-forms, and remind the reader that a vector field can be viewed as a derivation on the space of $C^\infty$ functions; in particular if $X$ is a vector field and $f \in C^\infty(M)$ we have a well-defined function $Xf$.

**Proposition 3.8.** *If $f \in \Omega^0(M) = C^\infty(M)$ then $\mathscr{L}_X f = Xf$.*

*Proof.* For any point $m \in M$ we have, using the chain rule

$$\begin{aligned}(\mathscr{L}_X f)(m) &= \left.\frac{d}{dt}\right|_{t=0} (\phi_X^{t*} f)(m) = \left.\frac{d}{dt}\right|_{t=0} (f \circ \phi_X^t)(m) \\ &= \left.\frac{d}{dt}\right|_{t=0} f\left(\phi_X^t(m)\right) = df\left(\left.\frac{d}{dt}\right|_{t=0} \phi_X^t(m)\right) = df(X_m) = (Xf)_m\end{aligned}$$

□

Recall that, since vector fields are derivations on $C^\infty(M)$, they have well-defined commutators ($[X, Y] = X \circ Y - Y \circ X$), which are also vector fields. Interestingly, commutators fit into the story of Lie derivatives:

**Theorem 3.9.** *If $X$ and $Y$ are vector fields on $M$ then $\mathscr{L}_X Y = [X, Y]$*

*Proof.* Let $f \in C^\infty(M)$ and $m \in M$; we are to show that $\left((\mathscr{L}_X Y)(f)\right)(m) = \left(X(Yf)\right)(m) - \left(Y(Xf)\right)(m)$.

We see (recalling that an element of, *e.g.* $T_m M$ is a derivation from the algebra of germs of $C^\infty$ functions around $m$ to $\mathbb{R}$, so if $v \in T_m M$ and $f \in C^\infty(M)$ we have a number $v(f)$):

$$\begin{aligned}\left((\mathscr{L}_X Y)(f)\right)(m) &= \lim_{t \to 0} \frac{((\phi_X^{-t})_* Y_{\phi_X^t(m)})f - Y_m f}{t} = \lim_{t \to 0} \frac{Y_{\phi_X^t(m)}(f \circ \phi_X^{-t}) - Y_m f}{t} \\ &= \left.\frac{d}{dt}\right|_{t=0} Y_{\phi_X^t(m)}(f \circ \phi_X^{-t})\end{aligned}$$

where in the first inequality we have used the definition of the pushforward in terms of derivations: $(\phi_* v)(f) = v(f \circ \phi)$. Now define a function of two variable $H$ by

$$H(s, t) = (f \circ \phi_X^{-t})(\phi_Y^s(\phi_X^t(m))).$$

We observe

$$\frac{\partial H}{\partial s}(0, t) = d(f \circ \phi_X^{-t})(Y_{\phi_X^t(m)}) = Y_{\phi_X^t(m)}(f \circ \phi_X^{-t}).$$

Combining this we the previous displayed equation we see that

$$\left((\mathscr{L}_X Y)(f)\right)(m) = \frac{\partial^2 H}{\partial t \partial s}(0, 0).$$

Now (setting $u=-t$ and using the chain rule)[2]

$$\begin{aligned}\frac{\partial^2 H}{\partial t\partial s}(0,0) &= \left.\frac{\partial^2 H}{\partial s\partial t}\right|_{(0,0)} f(\phi_Y^s(\phi_X^t(m))) - \left.\frac{\partial^2 H}{\partial s\partial u}\right|_{(0,0)} f(\phi_X^u(\phi_Y^s(m)))\\ &= \left.\frac{\partial}{\partial t}\right|_{t=0}\left(\left.\frac{\partial}{\partial s}\right|_{s=0} f\circ\phi_Y^s\right)(\phi_X^t(m)) - \left.\frac{\partial}{\partial s}\right|_{s=0}\left(\left.\frac{\partial}{\partial u}\right|_{u=0} f\circ\phi_X^u\right)(\phi_Y^s(m))\\ &= \mathscr{L}_X(Yf)_m - \mathscr{L}_Y(Xf)_m = \big((XY-YX)f\big)(m),\end{aligned}$$

where in the last inequality we use Proposition 3.8. Since $f$ and $m$ are arbitrary this proves that $\mathscr{L}_X Y = XY - YX$. □

We now turn to the Lie derivative on differential forms; just as with vector fields there turns out to be a rather simple formula, which is quite useful for geometric applications. First we observe:

**Lemma 3.10.** *The Lie derivative $\mathscr{L}_X$ (defined by $\mathscr{L}_X\omega = \left.\frac{d}{dt}\right|_{t=0}\phi_X^{t*}\omega$) is the unique linear map $\mathscr{L}\colon \Omega^*(M)\to\Omega^*(M)$ obeying the following properties:*

1. *For all $f\in\Omega^0(M)=C^\infty(M)$, $\mathscr{L}f = Xf$.*
2. *$d\mathscr{L}\omega = \mathscr{L}(d\omega)$ for all $\omega\in\Omega^*(M)$.*
3. *For all $\omega,\theta\in\Omega^*(M)$, $\mathscr{L}(\omega\wedge\theta) = (\mathscr{L}\omega)\wedge\theta + \omega\wedge(\mathscr{L}\theta)$, and*
4. *If $U$ is an open set and $\omega,\omega'\in\Omega^*(M)$ are such that $\omega|_U = \omega'|_U$, then $(\mathscr{L}\omega)|_U = (\mathscr{L}\omega')|_U$.*

*Proof.* First we should check that $\mathscr{L}_X$ obeys properties (1)-(4).

Property (1) is Proposition 3.8.

For property (2), simply note that, since $d$ commutes with pullback,

$$d\left(\phi_X^{t*}\omega - \omega\right) = \phi_X^{t*}d\omega - d\omega,$$

and then (2) follows by dividing by $t$ and taking the limit as $t\to 0$

For property (3) we have

$$\begin{aligned}\left.\frac{d}{dt}\right|_{t=0}\phi_X^{t*}(\omega\wedge\theta) &= \left.\frac{d}{dt}\right|_{t=0}(\phi_X^{t*}\omega)\wedge(\phi_X^{t*}\theta)\\ &= \left.\frac{d}{dt}\right|_{t=0}(\phi_X^{t*}\omega)\wedge\theta + \omega\wedge(\phi_X^{t*}\theta) = (\mathscr{L}_X\omega)\wedge\theta + \omega\wedge(\mathscr{L}_X\theta).\end{aligned}$$

Property (4) is easily verified: for any point $m\in U$ we will have $\phi_X^t(m)\in U$ for sufficiently small $t$, and so $(\phi_X^{t*}\omega)_m = (\phi_X^{t*}\omega')_m$ for all sufficiently small $t$, from which the conclusion immediately follows.

It remains to show that properties (1)-(4) uniquely specify a linear map. If $\mathscr{L}$ is any map obeying (1)-(3), and if $f, g_1,\dots,g_p\in C^\infty(M)$, then we will have $\mathscr{L}f = \mathscr{L}_X f$ and $\mathscr{L}g_i = \mathscr{L}_X g_i$ by (1), and then $\mathscr{L}(dg_i) = \mathscr{L}_X(dg_i)$ by (2), and then

$$\mathscr{L}\left(f\,dg_1\wedge\cdots\wedge dg_p\right) = \mathscr{L}_X\left(f\,dg_1\wedge\cdots\wedge dg_p\right)$$

by (3). So by linearity $\mathscr{L}$ and $\mathscr{L}_X$ coincide on any forms which are finite linear combinations of forms of the shape $f\,dg_1\wedge\cdots\wedge dg_p$. Now we proved earlier (Proposition 4.19 of Part 1) that any differential form $\omega$ can be written as a locally finite sum of forms of the shape $f\,dg_1\wedge\cdots\wedge dg_p$,

[2]Here and below I will make use of the following point (and similar ones) without comment: if $f(x,y)$ is some function and if we set $g(z) = f(-z,z)$, then the chain rule gives that $g'(0) = (\nabla f)\cdot\langle -1,1\rangle = \frac{\partial f}{\partial y}(0,0) - \frac{\partial f}{\partial x}(0,0)$

*i.e.*, $M$ is covered by open sets on each of which $\omega$ is a finite linear combinations of forms of the shape $f\,dg_1 \wedge \cdots \wedge dg_p$. Now if $\mathscr{L}$ (like $\mathscr{L}_X$) obeys condition (4) then the restriction of $\mathscr{L}\omega$ to any open set is determined by the restriction of $\omega$ to that set, so by considering the restriction of $\omega$ to the open sets in the cover in the previous paragraph we see that $\mathscr{L}\omega = \mathscr{L}_X\omega$. □

Accordingly if we find a simple formula for an operation obeying (1)-(4) above then we can deduce that $\mathscr{L}_X$ is given by that formula. To prepare for this, recall the operation of "interior multiplication" of a form by a vector field: For any vector field $X$ we get a map $\iota_X \colon \Omega^p(M) \to \Omega^{p-1}(M)$ defined by

$$(\iota_X\omega)(v_1, \ldots, v_{p-1}) = \omega(X, v_1, \ldots, v_{p-1}).$$

Here is the promised formula:

**Theorem 3.11** (Cartan's Magic Formula). *For any $\omega \in \Omega^p(M)$ we have*

$$\mathscr{L}_X\omega = d\iota_X\omega + \iota_X d\omega.$$

*Proof.* We just have to show that $\mathscr{L}'_X := d\iota_X + \iota_X d$ obeys condition (1)-(4) above.

If $f \in \Omega^0(M)$ we see that

$$\mathscr{L}'_X f = 0 + \iota_X df = df(X) = Xf,$$

confirming (1).

(2) holds, since both $d\mathscr{L}'_X$ and $\mathscr{L}'_X d$ are equal to $d\iota_X d$ (as $d^2 = 0$).

(4) is immediate from the definition of $\mathscr{L}'_X$.

So the only nontrivial part is (3). And this isn't too hard: the key point is (as I will leave you to verify, using formula (7) on p. 23 of part 1) the identity

$$\iota_X(\omega \wedge \theta) = (\iota_X\omega) \wedge \theta + (-1)^p \omega \wedge (\iota_X\theta) \quad \text{if } \omega \in \Omega^p(M).$$

Of course this is the same "anti-derivation" property as is satisfied by $d$. Combining these we get, if $\omega \in \Omega^p(M)$,

$$\begin{aligned}\big(d\iota_X + \iota_X d\big)(\omega \wedge \theta) &= d\big((\iota_X\omega) \wedge \theta + (-1)^p\omega \wedge (\iota_X\theta)\big) + \iota_X\left((d\omega) \wedge \theta + (-1)^p \omega \wedge (d\theta)\right)\\ &= (d\iota_X\omega) \wedge \theta + (-1)^{p-1}\iota_X\omega \wedge d\theta + (-1)^p d\omega \wedge (\iota_X\theta) + (-1)^{2p}\omega \wedge d\iota_X\theta\\ &\quad + (\iota_X d\omega) \wedge \theta + (-1)^{p+1} d\omega \wedge \iota_X\theta + (-1)^p \iota_X\omega \wedge d\theta + (-1)^{2p}\omega \wedge \iota_X d\theta,\end{aligned}$$

and after cancellation one ends up with precisely $\big(d\iota_X\omega + \iota_X d\omega\big) \wedge \theta + \omega \wedge \big(d\iota_X\theta + \iota_X d\theta\big)$, as desired. □

We have, by definition,

$$\mathscr{L}_X\omega = \left.\frac{d}{dt}\right|_{t=0} \phi_X^{t*}\omega;$$

to find the derivative at a time other than zero, we compute, using (2),

$$\left.\frac{d}{dt}\right|_{t=s} \phi_X^{t*}\omega = \lim_{h\to 0}\frac{\phi_X^{(s+h)*}\omega - \phi_s^*\omega}{h} = \lim_{h\to 0}\frac{\phi_X^{s*}(\phi_X^{h*}\omega - \omega)}{h} = \phi_X^{s*}\mathscr{L}_X\omega \tag{3}$$

**Corollary 3.12.** *Suppose that $\omega \in \Omega^*(M)$ is closed: $d\omega = 0$. Then a vector field $X$ has the property that $\phi_X^{t*}\omega = \omega$ for all $t$ if and only if $d(\iota_X\omega) = 0$.*

*Proof.* We have $\phi_X^{t*}\omega = \omega$ for all $t$ if and only if $\frac{d}{dt}\Big|_{t=s}\phi_X^{t*}\omega = 0$ for all $s$, which by (3) is equivalent to $\phi_X^{s*}\mathscr{L}_X\omega = 0$ for all $s$, which of course is equivalent to $\mathscr{L}_X\omega = 0$. Cartan's Magic Formula reveals that this in turn is equivalent to

$$d\iota_X\omega + \iota_X d\omega = 0,$$

and of course the second term on the left is zero since we assume $\omega$ is closed. □

3.2. **Volume forms and the Moser argument.** From now on let $M$ be a compact oriented $n$-manifold (without boundary). A *volume form* on $M$ is by definition a differential form $\omega \in \Omega^n(M)$ which is nowhere zero. In particular since volume forms have top degree they are obviously automatically closed. If $\omega$ is a volume form, then for any open subset $U \subset M$, by restricting $\omega$ to $U$ and then integrating we can define the volume of $U$:

$$vol_\omega(U) = \int_U \omega$$

(this is a finite number by virtue of the ambient manifold $M$ being compact). A diffeomorphism $\phi$ is called *volume-preserving* (with respect to the volume form $\omega$) if $\phi^*\omega = \omega$; this terminology is justified by recalling the behavior of integrals under pullbacks by diffeomorphisms: we have

$$vol_\omega(\phi(U)) = \int_{\phi(U)} \omega = \int_U \phi^*\omega = \int_U \omega = vol_\omega(U)$$

if $\phi$ is volume-preserving.

Corollary 3.12 shows how to construct many volume-preserving diffeomorphisms. Namely, the time-$t$ flow of a vector field $X$ will be volume-preserving provided that $d(\iota_X\omega) = 0$. To get a feel for this condition, note that we can write $\omega$ in local coordinates $(x_1, \ldots, x_n)$ as

$$\omega = g\,dx_1 \wedge \cdots \wedge dx_n$$

for some smooth nowhere-zero function $g$. Then if a vector field $X$ is given locally by $X = \sum_i f_i \frac{\partial}{\partial x_i}$, we will have

$$\iota_X\omega = \sum_i g f_i \iota_{\frac{\partial}{\partial x_i}} dx_1 \wedge \cdots \wedge dx_n = \sum_i (-1)^{i-1} g f_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n$$

and so

$$d(\iota_X\omega) = \left(\sum_i \frac{\partial}{\partial x_i}(g f_i)\right) dx_1 \wedge \cdots \wedge dx_n.$$

Thus the condition for the flow of $X$ to preserve $\omega$ is just that

$$\sum_{i=1}^n \frac{\partial}{\partial x_i}(g f_i) = 0.$$

In case the function $g$ is 1 (*i.e.*, $\omega$ restricts to the coordinate chart as the standard volume form $dx_1 \cdots \wedge \cdots dx_n$; it is actually always possible to find coordinates around any given point such that this holds—see Remark 3.14), this condition just reads that the divergence (in the standard multivariable calculus sense) of $X = \sum f_i \frac{\partial}{\partial x_i}$ should be zero. Thus *the flow of a divergence-free vector field preserves volume*.

It should be clear from the local coordinate formulas above that, given a volume form $\omega$ and *any* $\alpha \in \Omega^{n-1}(M)$, a unique vector field $X$ can be chosen so that $\iota_X\omega = \alpha$ (this can be done locally in coordinate charts, and then the local solutions can be pieced together with a partition of unity). Of course, there are many closed $(n-1)$-forms (for instance, the derivative of any $(n-2)$-form will do), and so there are many vector fields $X$ with $d\iota_X\omega = 0$. As such, from Cartan's Magic Formula we have seen that for any volume form on a compact oriented manifold there are many diffeomorphisms which preserve the volume form.

Now let us call two volume forms $\omega_0$ and $\omega_1$ on $M$ *equivalent* if there is a diffeomorphism $f: M \to M$ so that $f^*\omega_1 = \omega_0$. In view of the behavior of the integral under pullbacks, if $\omega_1$ is equivalent to $\omega_0$ it is obviously necessary to have $\int_M \omega_1 = \int_M \omega_0$. An argument of Moser shows that this condition is also sufficient:

**Theorem 3.13** ([Mos]). *Let $M$ be a compact connected oriented manifold without boundary and $\omega_0, \omega_1 \in \Omega^n(M)$ two volume forms such that $\int_M \omega_0 = \int_M \omega_1$. Then there is a diffeomorphism $f: M \to M$ so that $f^*\omega_1 = \omega_0$.*

*Sketch of proof.* First of all note that since the only closed 0-forms on a connected manifold are the constants, the assumption says that the $\omega_i$ have the same integral when wedged with any closed 0-form. By Poincaré duality, this then implies that $\omega_0$ and $\omega_1$ represent the same cohomology class in $H^n(M)$. So there is some $\alpha \in \Omega^{n-1}(M)$ such that $\omega_1 = \omega_0 + d\alpha$. Now for $0 \le t \le 1$ let

$$\omega_t = \omega_0 + t\,d\alpha = (1-t)\omega_0 + t\omega_1.$$

The fact that $\omega_0$ and $\omega_1$ have equal integrals (or even just integrals of the same sign) means that they induce the same orientation on $M$. So if $m \in M$ and $\{e_1, \ldots, e_n\}$ is a basis for $T_mM$ such that $\omega_0(e_1, \ldots, e_n) > 0$, then it will also hold that $\omega_1(e_1, \ldots, e_n) > 0$. But then for all $t \in [0,1]$

$$\omega_t(e_1, \ldots, e_n) = ((1-t)\omega_0 + t\omega_1)(e_1, \ldots, e_n) > 0.$$

Since $m \in M$ was an arbitrary point this shows that *the $\omega_t = \omega_0 + t\,d\alpha$ are all volume forms*.

The plan now is to find a *time-dependent* vector field $X_t$ on $M$ so that where $\{\phi_t\}$ is the flow of $X_t$ (*i.e.* $\phi_0 = id_M$ and $\frac{d}{dt}\phi_t(m) = X_t(\phi_t(m))$) we have $\phi_t^*\omega_t = \omega_0$ for all $t$. If we can do this then $f = \phi_1$ would be our desired diffeomorphism.

In this direction, Cartan's Magic Formula together with the chain rule can be seen to imply that, if $X_t$ has flow $\phi_t$:

$$\begin{aligned}\frac{d}{dt}(\phi_t^*\omega_t) &= \phi_t^*\frac{d\omega_t}{dt} + \phi_t^*\mathscr{L}_{X_t}\omega_t \\ &= \phi_t^*\left(d\alpha + d\iota_{X_t} + \iota_{X_t}d\omega_t\right) = \phi_t^* d\left(\alpha + \iota_{X_t}\omega_t\right)\end{aligned}$$

So one need only solve the equation $\iota_{X_t}\omega_t = -\alpha$, which by local coordinate considerations as above can be done in a unique way, producing a vector field $X_t$ which depends smoothly on $t$. So indeed we can just set $f$ equal to the time-one map of the flow of $X_t$. □

*Remark* 3.14. With sufficient care, one can localize this argument to show that for any volume form $\omega = g(x_1, \ldots, x_n)dx_1 \wedge \cdots \wedge dx_n$ on a neighborhood $U$ of the origin in $\mathbb{R}^n$, there are coordinates $(y_1, \ldots, y_n)$ on a smaller neighborhood $U'$ of the origin so that $\omega|_{U'} = dy_1 \wedge \cdots \wedge dy_n$. This justifies a statement made earlier that for any volume form, the manifold is covered by coordinate charts in which the volume form is given by the standard formula $dx_1 \wedge \cdots \wedge dx_n$.

## References

[Lee] John M. Lee. *Introduction to Smooth Manifolds* GTM 218, Springer, 2003.

[Mos] Jürgen Moser. *On the volume elements of a manifold*. Trans. Amer. Math. Soc. **120** (1965), 286–294.