Introduction to Smooth Manifolds & Lie Groups

Todd Kemp

Contents

Part 0

Review of Calculus.

1. Total Derivatives

DEFINITION 0.1. *Let $n, m \geq 1$ be integers, let $U \subseteq \mathbb{R}^n$ be open, and let $x_0 \in U$. A function $f\colon U \to \mathbb{R}^m$ is called* **differentiable** *at x_0 if there is a linear map $L\colon \mathbb{R}^n \to \mathbb{R}^m$ so that, for sufficiently small $v \in \mathbb{R}^n$,*

$$f(x_0 + v) - f(x_0) = L(v) + o(|v|).$$

More precisely, the statement is that

$$\lim_{v \to 0} \frac{f(x_0 + v) - f(x_0) - L(v)}{|v|} = 0.$$

When this is true, the linear map L is called the **derivative** *or* **total derivative** *of f at x_0, denoted $L = Df(x_0)$.*

PROPOSITION 0.2. *Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, and let $x_0 \in U$. The following facts are easy to verify.*

(1) *If f is differentiable at a point x_0, then it is continuous at x_0.*
(2) *If f is constant then it is differentiable at all points x, and $Df(x) = 0$.*
(3) *If f is a linear map, then f is differentiable at all points x, and $Df(x) = f$; that is, for any base point x, the derivative of f based at x is the linear map $[Df(x)]v = f(v)$.*

The usual "rules" of differentiation hold for this derivative, as follows.

PROPOSITION 0.3. *Let $f, g\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $x_0 \in U$. Let $y_0 = f(x_0) \in V \subseteq \mathbb{R}^m$, and let $h\colon V \to \mathbb{R}^p$ be differentiable at y_0.*

(1) *$f + g$ is differentiable at x_0, and $D(f + g)(x_0) = Df(x_0) + Dg(x_0)$.*
(2) *(Product Rule) Suppose $m = 1$. Then fg is differentiable at x_0, and $D(fg)(x_0) = f(x_0)Dg(x_0) + g(x_0)Df(x_0)$.*
(3) *(Chain Rule) $h \circ f$ is differentiable at x_0, and $D(h \circ f)(x_0) = Dh(y_0) \circ Df(x_0)$.*

The proofs of (1) and (2) are left as Exercises; for (3), see [**3**, Prop C.3, p.643].

2. Partial and Directional Derivatives

Given $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with $x_0 \in U$, and a vector $v \in \mathbb{R}^n$, the **directional derivative** of f at x_0 in the direction v, should it exist, is defined to be

$$D_v f(x_0) = \lim_{h \to 0} \frac{f(x_0 + hv) - f(x_0)}{h} = \frac{d}{dt} f(x_0 + tv)\bigg|_{t=0}.$$

As a special case, when we let v be a standard basis vector, we get the **partial derivatives**. Note: our standard notation for the coordinates in $\mathbb{R}^n$ is $(x^1, x^2, \ldots, x^n)$. The standard basis for $\mathbb{R}^n$ is denoted $\{e_1, e_2, \ldots, e_n\}$.

DEFINITION 0.4. *Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$, and let $x_0 \in U$. For $1 \leq j \leq n$, the* **partial derivative** *$\partial f / \partial x^j = \partial_j f$ at x_0 is the ordinary derivative of f at x_0 with respect to x^j, holding all other variables x^i with $i \neq j$ constant. That is*

$$\frac{\partial f}{\partial x^j}(x_0) = \lim_{h \to 0} \frac{f(x_0 + he_j) - f(x_0)}{h} = \frac{d}{dt} f(x_0 + te_j)\bigg|_{t=0} = D_{e_j} f(x_0),$$

should this limit exist.

If the partial derivative $\partial_j f$ exists at each point $x_0 \in U$, then we can interpret $\partial_j f$ as a function $U \to \mathbb{R}$ as well, and ask about its partial derivatives.

DEFINITION 0.5. *Let $U \subseteq \mathbb{R}^n$ be open, and let $k \in \mathbb{N}$. A function $f\colon U \to \mathbb{R}^m$ is called $C^k(U)$ if all mixed partial derivatives of length $\leq k$:*

$$\frac{\partial^\ell f}{\partial x^{j_1} \cdots \partial x^{j_\ell}}, \qquad 1 \leq \ell \leq k,\ j_1, \ldots, j_\ell \leq n$$

are continuous functions. If $f \in C^k(U)$ for all $k \in \mathbb{N}$, we say $f \in C^\infty(U)$, and call such a function **smooth**.

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be C^1. In terms of components $f = (f^1, \ldots, f^m)$, this means that the m real-valued functions $\{f^j\colon 1 \leq j \leq m\}$ are in $C^1(U)$. Then we can form the $m \times n$ matrix

$$[Jf(x_0)]^i_j = \frac{\partial f^i}{\partial x^j}(x_0).$$

This is the **Jacobian matrix** of f. (To be clear: A^i_j is the entry in the ith row and jth column of the matrix A, sometimes denoted A_{ij}.)

PROPOSITION 0.6. *Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. If f is differentiable at $x_0 \in U$, then the partial derivatives $\partial_j f(x_0)$, $1 \leq j \leq n$ exists, and in terms of the standard basis, the matrix of $Df(x_0)$ is $Jf(x_0)$. Conversely, if f is $C^1(U)$, then f is differentiable at each $x_0 \in U$, and so again $Df(x_0)$ has matrix $Jf(x_0)$.*

PROOF. By definition,

$$\partial_j f(x_0) = \frac{d}{dt} f(x_0 + te_j)\bigg|_{t=0} = D(f \circ \alpha_j)(0),$$

where $\alpha_j(t) = x_0 + te_j$ is an affine map $\mathbb{R} \to \mathbb{R}^n$ with $\alpha_j(0) = x_0$. Using Proposition 0.2, we have $[D\alpha_j(t)](s) = se_j$, and so by the chain rule, this derivative exists and

$$\partial_j f(x_0) = Df(x_0) \circ D\alpha_j(0).$$

In particular, we have $[D\alpha_j(0)](1) = e_j$, and so this tells us that $\partial_j f(x_0) = [Df(x_0)](e_j)$, which is precisely to say that $\partial_j f(x_0)$ is the jth column of the matrix of $Df(x_0)$ in the standard basis. This shows that $Jf(x_0)$ is the matrix.

For the converse, see [**6**]. □

We therefore usually make no distinction between Df and Jf when it is clear what basis we are working in. Note, then, mimicking the above proof using the chain rule, we have in general

$$D_v f(x_0) = \frac{d}{dt} f(x_0 + tv)\bigg|_{t=0} = [Df(x_0)](v).$$

In other words: if f is differentiable, then $v \mapsto D_v f(x_0)$ is a linear map (it is the total derivative). The linearity of this map is not at all clear from the definition of directional derivative; and, indeed, it is not true for non-differentiable functions (even ones with all partial derivatives existing).

COROLLARY 0.7 (Chain rule for partial derivatives). *Let $f = (f^1, \ldots, f^m)$ be a C^1 function $U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, let $x_0 \in U$, and let $g = (g^1, \ldots, g^p)$ be a C^1 function $V \subseteq \mathbb{R}^m \to \mathbb{R}^p$ with $f(x_0) \in V$. Then for $1 \leq i \leq p$*

$$\partial_j (g^i \circ f)(x_0) = \sum_{k=1}^m \partial_k g^i(f(x_0)) \partial_j f^k(x_0).$$

PROOF. We simply calculate

$$[\partial_j(g\circ f)(x_0)]^i = \partial_j(g^i\circ f)(x_0) = [J(g\circ f)(x_0)]^i_j = [D(g\circ f)(x_0)]^i_j = \sum_{k=1}^m [Dg(f(x_0))]^i_k [Df(x_0)]^k_j$$

and this is precisely what is stated above. □

COROLLARY 0.8. *A composition of smooth functions is smooth.*

The proof of Corollary 0.8 is left as an Exercise.

3. Taylor's Theorem

We will have frequent use for Taylor approximations of smooth functions. In multivariate form, it is useful to use multi-index notation. Denote $[n] = \{1, \dots, n\}$. For a positive integer k and $J = (j_1, \dots, j_k) \in [n]^k$, we denote for $x \in \mathbb{R}^n$

$$\partial_J = \partial_{j_1} \cdots \partial_{j_k}, \quad x^J = x^{j_1} x^{j_2} \cdots x^{j_k}.$$

Let k be a positive integer, and $f \in C^k(U)$ for some open set $U \subseteq \mathbb{R}^n$. For $p \in U$, the k**th order Taylor polynomial of** f **at** p is the polynomial function

$$P_k f(x;p) = \sum_{m=0}^k \frac{1}{m!} \sum_{J\in[n]^m} \partial_J f(p)(x-p)^J.$$

Notably we have the first and second order expansions

$$P_1 f(x;p) = f(p) + \sum_{j=1}^n \partial_j f(p)(x^j - p^j),$$

$$P_2 f(x;p) = P_1 f(x;p) + \frac{1}{2}\sum_{i,j=1}^n \partial_i\partial_j f(p)(x^i - p^i)(x^j - p^j).$$

Taylor's theorem says, in a precise sense, that $f(x) = P_k f(x;p) + o(|x-p|^k)$. Here we will state and prove this theorem in its most useful version: with integral remainder term.

THEOREM 0.9 (Taylor's Theorem). *Let $U \subseteq \mathbb{R}^n$ be open and convex, let k be a positive integer, and let and $f \in C^{k+1}(U)$. Then for any points $x, p \in U$,*

$$f(x) = P_k f(x;p) + \frac{1}{k!}\sum_{J\in[n]^{k+1}} (x-p)^J \int_0^1 (1-t)^k \partial_J f(p + t(x-p))\, dt. \tag{0.1}$$

PROOF. We prove this by induction on k. For $k = 0$, the desired statement is

$$f(x) = f(p) + \sum_{j=1}^n (x^j - p^j)\int_0^1 \partial_j f(p + t(x-p))\, dt.$$

To see why this is true, let $u(t) = f(p + t(x-p))$. Our assumption here is that f is C^1 in a convex neighborhood of x, p, and so the function u (defined along the line joining x to p) is $C^1[0,1]$. Hence, by the Fundamental Theorem of Calculus,

$$u(t) - u(0) = \int_0^1 u'(t)\, dt = \int_0^1 [Df(p + t(x-p))](x-p)\, dt$$

and this, together with the fact that $u(0) = f(p)$, yields the result.

Now, suppose (0.1) holds true for a given k. Fix a $J \in [n]^{k+1}$. Then, integrating by parts,

$$\begin{aligned}&\int_0^1 (1-t)^k \partial_J f(p + t(x-p))\, dt \\ &= \left[-\frac{(1-t)^{k+1}}{k+1}\partial_J f(p+t(x-p))\right]_{t=0}^{t=1} + \int_0^1 \frac{(1-t)^{k+1}}{k+1}\frac{\partial}{\partial t}\partial_J f(p+t(x-p))\, dt.\end{aligned}$$

The first term is simply $\frac{1}{k+1}\partial_J f(p)$. For the second term, we compute by the chain rule that that

$$\frac{\partial}{\partial t}(\partial_J f)(p+t(x-p)) = \sum_{j=1}^n \partial_j(\partial_J f)(p+t(x-p))(x^j - p^j).$$

Thus, summing over J, we have

$$\begin{aligned}&\sum_{J\in[n]^{k+1}} (x-p)^J \int_0^1 (1-t)^k \partial_J f(p+t(x-p))\, dt \\ =&\frac{1}{k+1}\sum_{J\in[n]^{k+1}} \partial_J f(p)(x-p)^J + \frac{1}{k+1}\sum_{J\in[n]^{k+1}, j\in[n]} (x-p)^J (x-p)^j \partial_j \partial_J f(p+t(x-p)) \\ =&\frac{1}{k+1}\sum_{J\in[n]^{k+1}} \partial_J f(p)(x-p)^J + \frac{1}{k+1}\sum_{J'\in[n]^{k+2}} (x-p)^{J'} \partial_{J'} f(p+t(x-p)).\end{aligned}$$

But, by the inductive hypothesis, this sum is $k!$ times $f(x) - P_k f(x)$. Dividing out and combining yields (0.1) at stage $k+1$, concluding the proof. □

COROLLARY 0.10. *Let U be an open convex subset of $\mathbb{R}^n$, let $p \in U$, and let $f \in C^{k+1}(U)$ for some positive integer k. If all partial derivatives of order $k+1$ of f are bounded on U – say $|\partial_J f(y)| \le M$ for $J \in [n]^{k+1}$ – then for $x \in U$,*

$$|f(x) - P_k f(x)| \le \frac{n^{k+1} M}{(k+1)!}|x-p|^{k+1}.$$

PROOF. There are n^{k+1} terms in the sum on the right-hand-side of (0.1). The Jth term is bounded by

$$\frac{1}{k!}|(x-p)^J| \int_0^1 (1-t)^k M\, dt = \frac{M}{(k+1)!}|(x^{j_1} - p^{j_1}) \cdots (x^{j_{k+1}} - p^{j_{k+1}})|.$$

Now, set $y = x - p$. The product term is then $|y^{j_1}| \cdots |y^{j_{k+1}}|$, which can be written uniquely in the form $|y^1|^{\epsilon_1} \cdots |y^n|^{\epsilon_n}$ for some non-negative integer exponents $\epsilon_1, \ldots, \epsilon_n$ satisfying $\epsilon_1 + \cdots + \epsilon_n = k+1$. Note that

$$\begin{aligned}|y|^{2(k+1)} = (|y^1|^2 + \cdots + |y^n|^2)^{k+1} &= \sum_{\substack{r_1,\ldots,r_n \ge 0 \\ r_1+\cdots+r_n = k+1}} \binom{k+1}{r_1 \cdots r_n} |y^1|^{2r_1} \cdots |y^n|^{2r_n} \\ &\ge |y^1|^{2\epsilon_1} \cdots |y^n|^{2\epsilon_n}.\end{aligned}$$

Taking square roots shows that $|y^{j_1}| \cdots |y^{j_{k+1}}| \le |y|^{k+1}$. Combining this with the above completes the proof. □

Taylor's Theorem is often stated only in the form of Corollary 0.10. This is useful for many applications, but it fails to make clear an important regularity result: the terms in the $o(|x-p|^k)$ remainder are all of the form $(x-p)^J$ (for some $J \in [n]^{k+1}$) times a nice function – if the original function is $C^{k+1+\ell}$, then these term are themselves C^ℓ. We record this as a proposition.

PROPOSITION 0.11. *In Taylor's Theorem 0.9, suppose $f \in C^{k+1+\ell}$ for some $\ell > 0$. Then the remainder terms are $(x-p)^J$ times functions*

$$x \mapsto \int_0^1 (1-t)^k \partial_J f(p + t(x-p))\, dt$$

that are C^ℓ.

PROOF. These are functions of the form $\int_0^1 g(x,t)\, dt$ where $g(x,t) = (1-t)^k \partial_J f(p+t(x-p))$. Since $f \in C^{k+1+\ell}$, $\partial_J f \in C^\ell$ for any $J \in [n]^{k+1}$, and so g is C^ℓ in both x and t. The result then follows by differentiating with respect to x (repeatedly) under the integral, which is easily justified in this situation (see, for example, [**6**]). □

4. Lipschitz Continuity

DEFINITION 0.12. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f\colon X \to X'$ is called* **Lipschitz** *if there is a constant $C > 0$ so that*

$$d_Y(f(x), f(x')) \le C \cdot d_X(x, x'), \qquad \textit{for all } x, x' \in X.$$

The function is said to be **locally Lipschitz** *if, given $x_0 \in X$, there is a neighborhood U containing x_0 such that $f|_U$ is Lipschitz.*

LEMMA 0.13. *Let $f\colon X \to Y$ be a locally Lipschitz map between metric spaces. If $K \subset X$ is compact, then $f|_K$ is Lipschitz.*

PROOF. By assumption, for each $x \in K$ there is neighborhood U_x of x on which f is Lipschitz; choose a radius $\delta(x)$ so that $B(x, 2\delta(x)) \subseteq U_x$; then there is a Lipschitz constant $C(x)$ so that

$$d_Y(f(x), f(x')) \le C(x) d_X(x, x'), \qquad \text{for all } x, x' \in B(x, 2\delta(x)).$$

Note that K is contained in the union of all the balls $B(x, \delta(x))$ for $x \in K$. Since K is compact, there are finitely many points $x_1, \dots, x_n \in K$ so that $K \subseteq B(x_1, \delta(x_1)) \cup \dots \cup B(x_n, \delta(x_n))$. Define $C = \max\{C(x_1), \dots, C(x_n)\}$ and $\delta = \min\{\delta(x_1), \dots, \delta(x_n)\}$.

- Suppose $d(x, x') < \delta$. We know there is a points $x_j \in \{x_1, \dots, x_n\}$ so that $d_X(x, x_j) < \delta(x_j)$. Then by the triangle inequality $d_X(x', x_j) \le d_X(x', x) + d_X(x, x_j) < \delta + \delta(x_j) \le 2\delta(x_j)$. Thus, both x, x' are in U_{x_j} where f is Lipschitz with constant $C(x_j) \le C$, and so $d_Y(f(x), f(x')) \le C d_X(x, x')$.
- Suppose, on the other hand, that $d_X(x, x') \ge \delta$. Since f is continuous, $f(K)$ is compact (*Exercise*), and hence it is bounded. Then we have

$$d_Y(f(x), f(x')) \le \operatorname{diam} f(K) \le \frac{\operatorname{diam} f(K)}{\delta} d_X(x, x').$$

This shows that f is Lipschitz on K, with constant $\le \max\{C, \operatorname{diam} f(K)/\delta\}$. □

The most prevalent examples of Lipschitz functions $\mathbb{R}^n \to \mathbb{R}^m$ (equipped with their usual Euclidean metrics) are C^1 functions.

PROPOSITION 0.14. *Let $U \subseteq \mathbb{R}^n$ be an open subset, and let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be C^1. Then f is locally Lipschitz.*

PROOF. Let $x_0 \in U$, and fix any closed ball $\overline{B}$ centered at x_0 with $\overline{B} \subset U$. Note that $\overline{B}$ is convex, so for any $a, b \in \overline{B}$ the whole line segment between a and b is contained in $\overline{B}$. By the chain rule, $t \mapsto f(a + t(b-a))$ is differentiable, and so by the fundamental theorem of calculus applied to its components, we have

$$f(b) - f(a) = \int_0^1 \frac{d}{dt} f(a + t(b-a))\, dt.$$

Applying the chain rule, this gives

$$f(b) - f(a) = \int_0^1 [Df(a + t(b-a))](b-a)\, dt.$$

We can therefore estimate

$$|f(b) - f(a)| \leq \int_0^1 |[Df(a + t(b-a))](b-a)|\, dt.$$

Given any $m \times n$ matrix A and vector $v \in \mathbb{R}^n$, by the Cauchy-Schwarz inequality we have $|Av| \leq |A|_2 |v|$ where $|A|_2$ is the Hilbert-Schmidt / Fröbenius norm of A:

$$|A|_2^2 \equiv \sum_{i=1}^m \sum_{j=1}^n |A_j^i|^2.$$

Thus, we have $|[Df(a + t(b-a))](b-a)| \leq |Df(a + t(b-a))|_2 |b-a|$. By assumption $f \in C^1(U)$, which means that $x \mapsto Df(x)$ is a continuous matrix-valued function, and therefore so is $x \mapsto |Df(x)|_2$. Hence, $M = \max_{\overline{B}} |Df|_2$ exists. Since $|Df(a + t(b-a))|_2 \leq M$, we therefore have

$$|f(b) - f(a)| \leq \int_0^1 |b-a|\, dt = M|b-a|.$$

So f is Lipschitz on the neighborhood B (the interior of $\overline{B}$) of x_0, with Lipschitz constant $\leq M$. $\square$

The converse of Proposition 0.14 is true in a sense: if f is locally Lipschitz on an open set in $\mathbb{R}^n$, then the f is differentiable *almost everywhere*.

DEFINITION 0.15. *A subset $N \subseteq \mathbb{R}^n$ is said to have* **measure zero** *if, for any $\epsilon > 0$, there is a countable collection of balls $\{B_j\}_{j=1}^\infty$ with $N \subseteq \bigcup_{j=1}^\infty B_j$, and $\sum_{j=1}^\infty \mathrm{Vol}(B_j) < \epsilon$.*

THEOREM 0.16 (Rademacher). *Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function. Then f is differentiable almost everywhere: the set of points in U at which f is* not *differentiable has measure zero.*

This theorem is beyond our present purview; the interested reader can find several proofs quickly by Googling.

The best kind of Lipschitz function is one whose Lipschitz constant is ≤ 1.

DEFINITION 0.17. *A function $f\colon X \to Y$ between metric spaces is called a* **contraction** *if*

$$d_Y(f(x), f(x')) \leq d_X(x, x'), \qquad \textit{for all } x, x' \in X.$$

If there is a constant $\lambda \in [0, 1)$ such that $d_Y(f(x), f(x')) \leq \lambda d_X(x, x')$ for all $x, x' \in X$, then f is called a **strict contraction**.

The most basic result for strict contraction is the Banach fixed-point theorem.

THEOREM 0.18 (Banach). *Let X be a non-empty* complete *metric space. If $f\colon X \to X$ is a strict contraction, then f has a unique fixed-point in X: a unique point $x \in X$ such that $f(x) = x$.*

PROOF. First, uniqueness: if $f(x) = x$ and $f(y) = y$, then we have

$$d(x, y) = d(f(x), f(y)) \leq \lambda d(x, y)$$

and since $\lambda < 1$ this is only possible if $d(x, y) \leq 0$, meaning $x = y$.

Now for existence. Fix any initial point $x_0 \in X$, and consider the sequence of iterates of f: $x_1 = f(x_0)$, $x_2 = f(x_1)$, and in general $x_n = f(x_{n-1}) = f^n(x_0)$. We will show that x_n converges to a point x, and that this limit is a fixed point of f (hence the unique fixed point). First, note that

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq \lambda d(x_n, x_{n-1}).$$

Iterating this n times yields

$$d(x_{n+1}, x_n) \leq \lambda^n d(x_1, x_0).$$

Thus, by the triangle inequality, for any $m \geq n$, we have

$$d(x_n, x_m) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{m-1}, x_m) \leq (\lambda^n + \lambda^{n+1} + \cdots + \lambda^{m-1}) d(x_0, x_1).$$

The sum is

$$\sum_{j=n}^{m-1} \lambda^j = \lambda^n \sum_{k=0}^{m-n-1} \lambda^k < \frac{\lambda^n}{1-\lambda} \to 0 \text{ as } n \to \infty.$$

This shows that $(x_n)_{n=0}^{\infty}$ is a Cauchy sequence. Since X is a complete metric space, it follows that there is a limit $x = \lim_{n\to\infty} x_n$. Now, using the continuity of f,

$$f(x) = f\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

Thus, x is a fixed-point of f. □

5. Inverse Function Theorem

Let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function differentiable at x_0. Suppose that f has a differentiable inverse in a neighborhood of x_0: so $f^{-1}(f(x)) = x$ for x in a neighborhood of x_0. Denote by I_n the $n \times n$ identity matrix. Then by the chain rule

$$I_n = D(x \mapsto x)(x_0) = D(f^{-1} \circ f)(x_0) = [D(f^{-1})(f(x_0))][Df(x_0)]. \tag{0.2}$$

That is: the derivative $Df(x_0)$ has an inverse matrix (which is $D(f^{-1})(f(x_0))$). A matrix can only be invertible if it is square, and so immediately we see this is only possible if $n = m$.

A smooth function from an open set $U \subseteq \mathbb{R}^n$ onto V which possesses a smooth inverse is called a **diffeomorphism**. The above calculation shows that diffeomorphisms only exists between like-dimensional open sets, and any diffeomorphism has invertible derivative at each point.

The inverse function theorem is a kind of converse to this: if the derivative $Df(x_0)$ is an invertible matrix, then f itself has a differentiable inverse in a small neighborhood of x_0: i.e. f is a **local diffeomorphism**. We will state this in the smooth (C^∞) category, but it holds true for C^k or just differentiable functions just as well.

THEOREM 0.19 (Inverse Function Theorem). *Let $U \subseteq \mathbb{R}^n$ be open, and let $f\colon U \to \mathbb{R}^n$ be smooth. Suppose that, for some point $x_0 \in U$, $Df(x_0)$ is invertible. Then there exists a connected neighborhood $U_0 \subseteq U$ of x_0, and a connected neighborhood $V_0 \subseteq f(U_0)$ of $f(x_0)$, such that $f|_{U_0} : U_0 \to V_0$ is a diffeomorphism.*

PROOF. To begin, we rescale and recenter: let $g(x) = [Df(x_0)]^{-1}(f(x + x_0) - f(x_0))$; then $f(x) = [Df(x_0)]g(x - x_0) + f(x_0)$, and so the statement of the theorem is equivalent to showing that g has an inverse in a connected neighborhood of 0. Note that $g(0) = 0$ and $Dg(0) = I_n$. Also note: since $x \mapsto Dg(x)$ is continuous, so is $x \mapsto \det Dg(x)$; since $\det Dg(0) = 1$, there is a neighborhood of 0 where Dg is invertible, and so by shrinking the domain of g (which is $U + x_0$) if necessary, we will assume Dg is invertible everywhere on the domain of g, which we denote U'.

Set $h(x) = x - g(x)$; then $Dh(0) = I_n - I_n = 0$. Since $x \mapsto Dh(x)$ is continuous, it follows that for all sufficiently small x, $|Dh(x)|_2 \le \frac{1}{2}$; so choose $\delta > 0$ small enough that $B(0, \delta) \subset U'$ and $|Dh(x)|_2 \le \frac{1}{2}$ for $|x| \le \delta$. Thus, by Proposition 0.14, if $x, y \in \overline{B}(0, \delta)$,

$$|h(x) - h(x')| \le \tfrac{1}{2}|x - x'|.$$

In particular, taking $x' = 0$ and noting that $h(0) = 0$, this gives $|h(x)| \le \frac{1}{2}|x|$ for $|x| \le \delta$.

Now, since $g(x) - g(x') + h(x) - h(x') = x - x'$, we have

$$|x - x'| \le |g(x) - g(x')| + |h(x) - h(x')| \le |g(x) - g(x')| + \tfrac{1}{2}|x - x'|, \quad \text{for } x, y \in \overline{B}(0, \delta).$$

Rearranging this yields

$$|x - x'| \le 2|g(x) - g(x')|. \tag{0.3}$$

This shows that g is one-to-one on $\overline{B}(0, \delta)$.

Claim: given $y \in B(0, \delta/2)$, there is a unique $x \in B(0, \delta)$ with $g(x) = y$. To see this, define $\tilde{h}(x) = y + h(x) = y + x - g(x)$; thus $g(x) = y$ iff $\tilde{h}(x) = x$. So it behooves us to show that $\tilde{h}$ has a unique fixed point. First note that

$$|\tilde{h}(x)| \le |y| + |h(x)| < \frac{\delta}{2} + \frac{1}{2}|x| \le \delta, \quad \text{for } |x| \le \delta$$

and so $\tilde{h}$ maps the complete metric space $\overline{B}(0, \delta)$ into itself. We note also that $|\tilde{h}(x) - \tilde{h}(x')| = |h(x) - h(x')| \le \frac{1}{2}|x - x'|$, and so $\tilde{h}$ is a strict contraction. Thus, the claim follows from Theorem 0.18.

Now, set $V_0' = B(0, \delta/2)$ and $U_0' = B(0, \delta) \cap g^{-1}(V_0')$. These are both open sets. The above argument shows that $g\colon U_0' \to V_0'$ is a bijection. From (0.3) we see that g^{-1} is Lipschitz continuous, and it follows that the preimage $g^{-1}(V_0')$ is also connected. All we have left to do is show that g^{-1} is smooth. If we knew this, (0.2) would show immediately that $D(g^{-1})(g(x)) = [Dg(x)]^{-1}$. In fact, we begin by showing directly that g^{-1} is differentiable, with total derivative given by this formula. That is: fix $y \in V_0'$, and let $x = g^{-1}(y)$. Then for $y' \ne y$ in V_0', setting $x' = g^{-1}(y') \ne x$, we have

$$\begin{aligned}\frac{g^{-1}(y') - g^{-1}(y) - [Dg(x)]^{-1}(y' - y)}{|y' - y|} &= [Dg(x)]^{-1}\left(\frac{Dg(x)(x' - x) - (y' - y)}{|y' - y|}\right) \\ &= -\frac{|x' - x|}{|y' - y|}[Dg(x)]^{-1}\left(\frac{g(x') - g(x) - Dg(x)(x' - x)}{|x' - x|}\right).\end{aligned}$$

Note that $|x' - x|/|y' - y| \le 2$ by (0.3). Since g^{-1} is continuous, as $y' \to y$, $x' \to x$. Now, as $x' \to x$, $Dg(x') \to Dg(x)$ since g is C^1. Since the matrix function $A \mapsto A^{-1}$ is continuous, it follows that $[Dg(x')]^{-1} \to [Dg(x)]^{-1}$. Since g is differentiable at x, the quantity inside the

braces tends to 0 as $x' \to x$. All together, this shows that g^{-1} is differentiable at x, with derivative $[Dg(x)]^{-1}$.

Thus far, we have shown that g^{-1} is differentiable, and its derivative at y is given by $[Dg(x)]^{-1} = [Dg(g^{-1}(y))]^{-1}$. As a function of y, this is a composition:

$$y \overset{g^{-1}}{\mapsto} g^{-1}(y) \overset{Dg}{\mapsto} Dg(g^{-1}(y)) \overset{\iota}{\mapsto} [Dg(g^{-1}(y))]^{-1}, \tag{0.4}$$

where here Dg denotes the map $x \mapsto Dg(x)$, and $\iota(A) = A^{-1}$ is the inverse map on matrices. Each of these functions is continuous (indeed ι is smooth by, say, Cramer's rule, which expresses the entries of A^{-1} as rational functions of the entries of A), and this shows that the map $y \mapsto D(g^{-1})(y)$ is continuous; hence g^{-1} is C^1.

Finally, we show g^{-1} is smooth inductively: suppose we have shown g^{-1} is C^k. Then the maps g^{-1} and Dg in (0.4) are C^k, and ι is C^∞, which shows that $D(g^{-1})$ is C^k. Hence its components, the partial derivatives of g^{-1}, are C^k, which is precisely to say that g^{-1} is C^{k+1}. Induction now shows that $g \in C^\infty$, concluding the proof. $\square$

6. Implicit Function Theorem

An alternative way to view the Inverse Function Theorem is as follows: given $f\colon \mathbb{R}^n \to \mathbb{R}^n$, consider the function $\Phi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ given by $\Phi(x, y) = y - f(x)$. Then the level set $\Phi(x, y) = 0$ can be (globally) described as the graph of y as a function (f) of x. The Inverse Function Theorem tells us that, if $Df(x_0)$ is invertible, then this level set may *also* be described (locally) as the graph of x as a function of y.

Put in those terms, there is a natural generalization to the case when $n \neq m$.

THEOREM 0.20 (Implicit Function Theorem). *Let $U \subseteq \mathbb{R}^n \times \mathbb{R}^m$ be open. Denote the coordinates on $\mathbb{R}^n \times \mathbb{R}^m$ as $(x^1, \dots, x^n, y^1, \dots, y^m)$. Let $\Phi\colon U \to \mathbb{R}^m$ be a smooth function, fix $(x_0, y_0) \in U$, and let $z_0 = \Phi(x_0, y_0)$. Suppose that the $m \times m$ matrix*

$$\left[\frac{\partial \Phi^i}{\partial y^j}(x_0, y_0)\right]_{1 \le i, j \le m}$$

is invertible. Then there are neighborhoods $V_0 \subseteq \mathbb{R}^n$ of x_0 and $W_0 \subseteq \mathbb{R}^m$ of y_0, and a smooth function $f\colon V_0 \to W_0$ such that the level set $\Phi^{-1}(z_0) \cap (V_0 \times W_0)$ is the graph of $y = f(x)$; that is

$$\textit{for all } (x, y) \in V_0 \times W_0, \quad \Phi(x, y) = z_0 \textit{ if and only if } y = f(x).$$

Before giving the proof, we give an example to illustrate the theorem.

EXAMPLE 0.21. Do the equations

$$\begin{aligned} x^2 - y &= a \\ y^2 - z &= b \\ z^2 - x &= 0 \end{aligned}$$

determine (x, y, z) implicitly as functions of (a, b) in a neighborhood of $(x, y, z, a, b) = (0, 0, 0, 0, 0)$? To answer, we phrase the question in terms of the level set of a smooth function, and calculate its derivative:

$$\Phi(x, y, z, a, b) = (x^2 - y - a, y^2 - z - b, z^2 - x), \quad D\Phi(x, y, z, a, b) = \begin{bmatrix} 2x & -1 & 0 & -1 & 0 \\ 0 & 2y & -1 & 0 & -1 \\ -1 & 0 & 2z & 0 & 0 \end{bmatrix}.$$

We want to know if (x, y, z) are locally functions of (a, b); to apply the implicit function theorem, we must consider the 3×3 sub-matrix corresponding to ∂_x, ∂_y, and ∂_z, which is

$$\begin{bmatrix} 2x & -1 & 0 \\ 0 & 2y & -1 \\ -1 & 0 & 2z \end{bmatrix}.$$

We can readily compute that the determinant of this sub-matrix is $8xyz - 1$. So, at the point $x = y = z = a = b = 0$ (which does indeed satisfy the equations), this matrix has determinant -1 which is $\neq 0$. It therefore follows from the Implicit Function Theorem that, indeed, (x, y, z) can be expressed locally as a function of (a, b) near this point. In fact, this holds true in a neighborhood of any point (x, y, z, a, b) satisfying the equations *except* possibly when $xyz = \frac{1}{8}$.

We might ask instead if (x, z, a) can be expressed locally as a function of (y, b). To answer this we need to look at the sub-matrix corresponding to ∂_x, ∂_z, and ∂_a, which is

$$\begin{bmatrix} 2x & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 2z & 0 \end{bmatrix}, \quad \text{with determinant constantly equal to } 1.$$

So this matrix is non-singular at all points, and so indeed (x, z, a) can be expressed locally as a function of (y, b) near *any* point. Indeed, a little calculation will show that (x, z, a) can be expressed *globally* as

$$\begin{aligned} x &= (y^2 - b)^2 \\ z &= y^2 - b \\ a &= (y^2 - b)^4 - y. \end{aligned}$$

Caution, though: in general, even if the appropriate sub-matrix is always non-singular, the only guarantees that there is a *local* impact function near each point. It does not generally imply that there is a global implicit function, as there is in this example.

PROOF. We consider the auxiliary function $\Psi\colon U \to \mathbb{R}^n \times \mathbb{R}^m$ defined by $\Psi(x, y) = (x, \Phi(x, y))$. Then we compute the Jacobian:

$$D\Psi(x_0, y_0) = \begin{bmatrix} I_n & 0_{n\times m} \\ \left[\frac{\partial \Phi^i}{\partial x^j}(x_0, y_0)\right]_{\substack{1\le i\le m \\ 1\le j\le n}} & \left[\frac{\partial \Phi^i}{\partial y^j}(x_0, y_0)\right]_{1\le i,j\le m} \end{bmatrix}.$$

This is a block lower-triangular matrix, with both (square) diagonal blocks invertible (by assumption of the theorem for the lower right block) at $(x, y) = (x_0, y_0)$. It follows that $D\Psi(x_0, y_0)$ is invertible. It therefore follows from the Inverse Function Theorem that there exist connected neighborhoods U_0 of (x_0, y_0) and Y_0 of (x_0, z_0) such that $\Psi\colon U_0 \to Y_0$ is a diffeomorphism. U_0 contains a product neighborhood $V \times W$ of (x_0, y_0), so shrinking U_0 if necessary and setting $Y_0 = \Psi(V \times W)$, wlog we assume $U_0 = V \times W$.

On this neighborhood, we have a local inverse function $\Psi^{-1}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$. We may then write it in the form $\Psi^{-1}(x, y) = (A(x, y), B(x, y))$, and thus

$$(x, y) = \Psi(\Psi^{-1}(x, y)) = \Psi(A(x, y), B(x, y)) = (A(x, y), \Phi(A(x, y), B(x, y))). \tag{0.5}$$

Thus $A(x, y) = x$, and so $\Psi^{-1}(x, y) = (x, B(x, y))$ for some smooth function $B\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$.

Now, let $V_0 = \{x \in V : (x, z_0) \in Y_0\}$ and $W_0 = W$, and define $f : V_0 \to W_0$ as $f(x) = B(x, z_0)$. Again from (0.5), it follows that

$$z_0 = \Phi(x, B(x, z_0)) = \Phi(x, f(x)), \quad \text{for } x \in V_0.$$

This shows that the graph of f is indeed contained in $\Phi^{-1}(z_0)$. Conversely, let $(x, y) \in V_0 \times W_0$ be any point where $\Phi(x, y) = z_0$. Then $\Psi(x, y) = (x, \Phi(x, y)) = (x, z_0)$, and so

$$(x, y) = \Psi^{-1}(x, z_0) = (x, B(x, z_0)) = (x, f(x)),$$

showing that $y = f(x)$. So any such point is in the graph of f. This completes the proof. □

7. Solutions of ODEs

An **initial value problem** is a system of ODEs (ordinary differential equations) of the following form: given an open interval $J \subseteq \mathbb{R}$, an open subset $U \subseteq \mathbb{R}^n$, a continuous function $F : J \times U \to \mathbb{R}^n$, an initial time $t_0 \in J$, and an initial condition $y_0 \in U$, we seek an $\mathbb{R}^n$-valued differentiable function y defined on (a subset of) J satisfying

$$\dot{y}(t) = F(y_0, y(t)), \qquad y(t_0) = y_0.$$

Note: $\dot{y}(t) = y'(t) = \frac{dy}{dt}$; the notation $\dot{y}$ is convenient when there are superscript around, and is common (especially in physics) to denote a time derivative, which will be our perspective in this course.

The fundamental theorem, stated below, is that if the driving vector field F is sufficiently smooth (at least locally Lipschitz), then there always exists a unique solution to the initial value problem for any given initial data x_0. (The solution may only exist for a short time.) Moreover, the dependence of the solution on the initial data is as smooth as the driving field. This last point is not often covered in ODE courses (because it is much harder to prove), but it will be very important to us in this course.

The special case in which the driving field F does not depend explicitly on time t is called the **autonomous** case. It turns out the general case can be reduced to the autonomous case, as we will see later; for now, we restrict to autonomous systems.

THEOREM 0.22 (Picard–Lindelöf Theorem for Autonomous ODEs). *Suppose $U \subseteq \mathbb{R}^n$ and $J \subseteq \mathbb{R}$ are open, and $F : U \to \mathbb{R}^n$ is locally Lipschitz. Fix a point $(t_0, y_0) \in J \times U$, and consider the initial value problem*

$$\dot{y}(t) = F(y(t)), \qquad y(t_0) = y_0. \tag{ODE}$$

(1) LOCAL EXISTENCE: *There exists an open interval J_0 containing t_0, and a C^1 function $y : J_0 \to U$ that solves* (ODE).
(2) UNIQUENESS: *Any two differentiable solutions to* (ODE) *agree on the intersection of their domains.*
(3) SMOOTH DEPENDENCE ON INITIAL CONDITIONS: *Let $J_0 \subseteq \mathbb{R}$ be an open interval containing t_0 and let $U_0 \subseteq U$ be open. Suppose that $\theta : J_0 \times U_0 \to U$ is a function such that $y(t) = \theta(t, y_0)$ is a solution to* (ODE) *with initial condition y_0. Then θ is as smooth as F (i.e. if $F \in C^k(U)$ then $\theta \in C^k(J_0 \times U_0)$).*

REMARK 0.23. As we will see in the proof, a priori the interval J_0 on which the solution y exists depends on the initial condition x_0; however, this interval depends on x_0 in a smooth way *if the boundary of U is smooth*. In this common kind of situation, it then makes sense for the function

θ in (3) to be defined on a uniform interval J_0 for varying initial conditions, as we explicitly assume there.

An important technical tool in the proof of Theorem 0.22 is the following ODE comparison theorem (which is a generalized version of Gronwall's inequality).

LEMMA 0.24 (Gronwall's Lemma). *Let $J \subseteq \mathbb{R}$ be an open interval, and let $f\colon [0,\infty) \to [0,\infty)$ be Lipschitz. Suppose $u\colon J \to \mathbb{R}^n$ is a differentiable function which satisfies the differential inequality*

$$|\dot{u}(t)| \leq f(|u(t)|), \qquad \text{for all } t \in J.$$

Fix $t_0 \in J$, and suppose $v\colon [0,\infty) \to [0,\infty)$ is continuous, and differentiable on $(0,\infty)$, and satisfies the initial value problem

$$\dot{v}(t) = f(v(t)), \quad v(0) = |u(t_0)|.$$

Then, for all $t \in J$, it follows that $|u(t)| \leq v(|t - t_0|)$.

PROOF. First, by shifting $\tilde{u}(t) = u(t + t_0)$ and $\tilde{J} = J - t_0$, we may wlog take $t_0 = 0$ (and so we rename $\tilde{J}$ to J and $\tilde{u}$ to u). Let $J^+ = J \cap [0,\infty)$. If t is such that $u(t) = 0$, then the desired inequality $|u(t)| \leq v(|t|)$ holds trivially as $v \geq 0$; so we restrict our attention to the set $\{t \in J^+\colon u(t) > 0\}$. Since u is differentiable, hence continuous, this is an open set; on this set, $|u(t)|$ is also differentiable. We can then calculate

$$\frac{d}{dt}|u(t)| = \frac{1}{|u(t)|}u(t) \cdot \dot{u}(t) \leq \frac{1}{|u(t)|}|u(t)||\dot{u}(t)| = |\dot{u}(t)| \leq f(|u(t)|), \tag{0.6}$$

where the first inequality is the Cauchy-Schwarz inequality, and the second is an assumption of the lemma.

Now, let A be a Lipschitz constant for f, and define $w(t) = e^{-At}(|u(t)| - v(t))$. Then $w(0) = 0$, and for $t \in J^+$, the desired conclusion $|u(t)| \leq v(t)$ is equivalent to $w(t) \leq 0$. Well, if $t \in J^+$ and $w(t) > 0$ (meaning that $|u(t)| > v(t) \geq 0$), it follows that w is differentiable, and we can calculate

$$\begin{aligned}\dot{w}(t) &= -Ae^{-At}(|u(t)| - v(t)) + e^{-At}\frac{d}{dt}(|u(t)| - v(t)) \\ &\leq -Ae^{-At}(|u(t)| - v(t)) + e^{-At}(f(|u(t)|) - f(v(t)))\end{aligned}$$

which follows from (0.6) along with the assumption $\dot{v}(t) = f(v(t))$. But A is a Lipschitz constant for f, and so

$$f(|u(t)|) - f(v(t)) \leq |f(|u(t)|) - f(v(t))| \leq A||u(t)| - v(t)| = A(|u(t)| - v(t)).$$

This shows that $\dot{w}(t) \leq 0$.

Since $w(0) = 0$, if w were differentiable for all $t \in J^+$ we could now conclude that $w(t) \leq 0$ which would conclude the proof. As it stands, we need to be a little more careful. Suppose, for a contradiction, that there is some time $t_1 \in J^+$ where $w(t_1) > 0$. Define

$$\tau = \sup\{t \in [0, t_1]\colon w(t) \leq 0\}.$$

Then, since w is continuous, $w(\tau) = 0$, and by definition $w(t) > 0$ for $\tau < t \leq t_1$. But then w is differentiable on (τ, t_1), and continuous on $[\tau, t_1]$, and so by the Mean Value Theorem there is some $t \in (\tau, t_1)$ where $\dot{w}(t) > 0$, even though $w(t) > 0$, contradicting the above calculations showing that if $w(t) > 0$ for $t \in J^+$ then $\dot{w}(t) \leq 0$. Thus, we have shown that $w(t) \leq 0$ for $t \in J^+$.

To conclude, we simply replace t by $-t$ in the above argument to show the result also holds for $t \in J^- = J \cap (-\infty, 0]$. $\square$

Let us now proceed to the proof of Theorem 0.22. We begin with the easy case: uniqueness.

PROOF OF THEOREM 0.22(2). Let y_1, y_2 be two solutions of (ODE) defined on the same interval J_0, but with potentially different initial conditions. Shrink the interval slightly to an open interval J_1 containing t_0 such that $\overline{J_1} \subset J_0$. The union of the two continuous paths $y_2(\overline{J_1})$ and $y_2(\overline{J_1})$ is a compact subset of U. So, by Lemma 0.13, the locally Lipschitz vector field F is globally Lipschitz on this union of paths; let C be a Lipschitz constant. Thus

$$\left|\frac{d}{dt}(y_1(t) - y_2(t))\right| = |F(y_1(t)) - F(y_2(t))| \leq C|y_1(t) - y_2(t)|.$$

Applying Lemma 0.24 with $u(t) = y_1(t) - y_2(t)$, $f(v) = Cv$, and $v(t) = e^{Ct}|y_1(t_0) - y_2(t_0)|$ yields

$$|y_1(t) - y_2(t)| \leq e^{C|t-t_0|}|y_1(t_0) - y_2(t_0)|, \quad t \in \overline{J_1}.$$

Thus, if the initial conditions $y_1(t_0) = y_2(t_0)$ are the same, then $y_1(t) = y_2(t)$ for all $t \in \overline{J_1}$. Since every point of J_0 is contained in some closed subinterval $\overline{J_1}$ of J, the result now follows. $\square$

No we continue with the existence proof.

PROOF OF THEOREM 0.22(1). By assumption F is locally Lipschitz on U. We begin by restricting to some closed ball $\overline{B} = \overline{B}_r(y_0) \subset U$ that containst y_0 (we choose r small enough that this closed ball is contained in U); by Lemma 0.13, F is Lipschitz on $\overline{B}$.

They key idea is to rewrite (ODE) as an integral equation:

$$y(t) = x_0 + \int_{t_0}^{t} F(y(s))\, ds. \tag{0.7}$$

Eq. (0.7) is equivalent to (ODE) by the fundamental theorem of calculus. Indeed, if y is any solution to (ODE), since F is (assumed) continuous, it follows from (ODE) that $y \in C^1(J_0)$, and so the fundamental theorem of calculus applies. Conversely, if $y\colon J_0 \to U$ is any continuous function satisfying (0.7), again the continuity of F shows that, by the fundamental theorem of calculus, y satisfies (ODE) (and therefore is C^1 as above). Henceforth, we show that (0.7) has a solution.

We now introduce a linear operator on the space $C^0(J_0; \overline{B})$ of continuous maps $J_0 \to U$:

$$(Iy)(t) = x_0 + \int_{t_0}^{t} F(y(s))\, ds.$$

(Note that we have not yet defined the interval J_0; for the time being, the definition depends on an arbitrarily chosen open interval $J_0 \subseteq J$.) For any continuous y, as F is continuous, the integrand of Iy is continuous, and hence (by the fundamental theorem of calculus) $Iy \in C^1(J_0; \mathbb{R}^n) \subset C^0(J_0; \mathbb{R}^n)$. That is,

$$I\colon C^0(J_0; \overline{B}) \to C^0(J_0; \mathbb{R}^n)$$

is a linear operator. Note that any solution of (0.7) is a fixed point for I, and vice versa. To prove such a fixed point exists (and conclude the proof), we want to put a metric space structure on $C^0(J_0; U)$ and show that I is a strict contraction, to use the Banach fixed point theorem. The delicate point is that I does not a priori map $C^0(J_0; U)$ back into itself. This is where the choice of J_0 comes in: we can make it small enough that I *does* map it into itself, simply because F is

bounded. Let $M = \sup_{\overline{B}} |F|$, and choose $0 < \epsilon < r/M$. We will take $J_0 = (t_0 - \epsilon, t_0 + \epsilon)$. Then we have the following.

Claim. *The linear operator I maps $C^0(J_0; \overline{B})$ into itself.* To see this, since we've already discussed above that I maps continuous functions to continuous functions, it suffices to show that $Iy(t) \in \overline{B}$ for any $t \in J_0$. To that end, we estimate as follows:

$$|Iy(t) - y_0| = \left| y_0 + \int_{t_0}^{t} F(y(s))\, dx - y_0 \right| \leq \int_{t_0}^{t} |F(y(s))|\, ds \leq \int_{t_0}^{t} M\, ds < M\epsilon \leq r.$$

This proves the claim.

Now we must introduce a metric on $C^0(J_0; \overline{B})$ and show that I is a strict contraction, which will allow us to use the Banach Fixed Point Theorem 0.18 to prove that I has a unique fixed point, concluding the proof. There is a standard metric for continuous functions on a compact set, given by the sup-norm: $\|y\|_\infty \equiv \sup_{J_0} |y|$. Thus, we define

$$d_\infty(y_1, y_2) \equiv \sup_{j \in J_0} |y_1(t) - y_2(t)|.$$

It is a standard result that this is a *complete* metric on $C^0(J_0; \overline{B})$ if J_0 is compact; in our case, we have functions y that are continuous on a bigger interval J. We are taking $J_0 = (t_0 - \epsilon, t_0 + \epsilon)$ where (so far) we know $\epsilon < r/M$. We must also choose ϵ small enough that the closed interval $[t_0 - \epsilon, t_0 + \epsilon]$ is contained in J. Then we know that our solution is in $C^0(\overline{J_0}; \overline{B})$, which is a complete metric space in d_∞. (Note: completeness is just the statement that any uniformly Cauchy sequence of continuous functions has a continuous limit.)

So, we are left to show that I is a strict contraction on the metric space $C^0(J_0; \overline{B})$. To guarantee this, we need one more constraint on ϵ. : Let C be a Lipschitz constant for F on $\overline{B}$; then we make sure that $\epsilon < 1/C$. Then we have for any two $y_1, y_2 \in C^0(J_0; \overline{B})$,

$$\begin{aligned} d_\infty(Iy_1, Iy_2) = \sup_{t \in J_0} \left| \int_{t_0}^{t} F(y_1(s))\, ds - \int_{t_0}^{t} F(y_2(s))\, ds \right| &\leq \sup_{t \in J_0} \int_{t_0}^{t} |F(y_1(s)) - F(y_2(s))|\, ds \\ &\leq \sup_{t \in J_0} \int_{t_0}^{t} C|y_1(s) - y_2(s)|\, ds \\ &\leq C \cdot \sup_{t \in J_0} \int_{t_0}^{t} d_\infty(y_1, y_2)\, ds \\ &= C|t - t_0| d_\infty(y_1, y_2) < C\epsilon \cdot d_\infty(y_1, y_2). \end{aligned}$$

Since we choose $\epsilon < 1/C$, the constant $C\epsilon$ is strictly < 1, and hence, by the Banach fixed point theorem, I possesses a unique fixed point in $C^0(J_0; \overline{B})$. As discussed above, this fixed point is the desired solution to (ODE).

To summarize: we have shown that for any $\epsilon > 0$ small enough that $J_0 = (t_0 - \epsilon, t_0 + \epsilon)$ is contained in J, and satisfying $C\epsilon < 1$ and $M\epsilon < r$ (where C is a Lipschitz constant for F, M is the maximum of F, and r is the distance from y_0 to the boundary of U), it follows that (ODE) has a solution which exists on the interval J_0. □

REMARK 0.25. This proof is often called "Picard iteration": we showed there is a solution by appealing to the Banach fixed point Theorem 0.18. The proof of that theorem is by iteration. So, in total, the above proof shows that the solution to (ODE) can be achieved by iteration of the operator I starting at any continuous function: i.e. the solution is given by $y(t) = \lim_{n \to \infty} I^n(y_0)(t)$ (where, for convenience, we choose our starting point to be the constant function $y = y_0$).

For extra nerdiness, we will refer to this proof as "the Picard maneuver". :)

As for Theorem 0.22(3), which is an important result for us: the proof is quite involved and technical, so we will leave it for the time being. The interested reader may find it on [**3**, pp. 667-672].

We conclude this section by showing how the general (non-autonomous) case follows as a special case of the autonomous case (in one dimension higher).

THEOREM 0.26 (Full Picard–Lindelöf Theorem). *Suppose $U \subseteq \mathbb{R}^n$ and $J \subseteq \mathbb{R}$ are open, and $F\colon J \times U \to \mathbb{R}^n$ is locally Lipschitz. Fix a point $(t_0, y_0) \in J \times U$, and consider the initial value problem*

$$\dot{y}(t) = F(t, y(t)), \qquad y(t_0) = y_0. \tag{ODE'}$$

(1) LOCAL EXISTENCE: *There exists an open interval J_0 containing t_0, and a C^1 function $y\colon J_0 \to U$ that solves* (ODE).
(2) UNIQUENESS: *Any two differentiable solutions to* (ODE) *agree on the intersection of their domains.*
(3) SMOOTH DEPENDENCE ON INITIAL CONDITIONS: *Let $J_0 \subseteq \mathbb{R}$ be an open interval containing t_0 and let $U_0 \subseteq U$ be open. Suppose that $\theta\colon J_0 \times U_0 \to U$ is a function such that $y(t) = \theta(t, y_0)$ is a solution to* (ODE) *with initial condition y_0. Then θ is as smooth as F (i.e. if $F \in C^k(U)$ then $\theta \in C^k(J_0 \times U_0)$).*

PROOF. Let $\mathbf{F} = (1, F)$ be the $\mathbb{R}^{n+1}$-valued vector field whose first component (which we label F^0 for convenience) is constantly 1; this vector field is locally Lipschitz since F is. Consider the following autonomous ODE in $\mathbb{R}^{n+1}$:

$$\dot{y}(t) = \mathbf{F}(\mathbf{y}(t)) = (1, F(y^0(t), y(t))), \qquad \mathbf{y}(t_0) = (t_0, y_0). \tag{0.8}$$

The first component of this equation is simply $y^0(t) = t$ with initial condition $y(t_0) = t_0$, whose unique solution is $y(t) = t$. Hence, given the unique solution $\mathbf{y} = (y^0, y)$ of (0.8) (guaranteed to exist by Theorem 0.22(1)), the last n components y give a solution to (ODE'), which is unique by Theorem 0.22(2) (since uniqueness of $\mathbf{y}$ clearly implies uniqueness of y), and depends as smoothly as F on initial conditions by Theorem 0.22(3) since $\mathbf{F} \in C^k$ iff $F \in C^k$. $\square$

Part 1

Manifolds and Differential Geometry.

CHAPTER 1

Smooth Manifolds

1. Smooth Surfaces in $\mathbb{R}^d$

A smooth manifold is a generalization of a classical smooth surface. The most basic kind of surface is the graph of a smooth function. That is: if $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is smooth, then the graph $\Gamma(f, U) = \{(x, y) \in U \times \mathbb{R}^m \colon y = f(x)\}$ is a smooth surface. What does this mean? Consider the associated function $\psi\colon U \to \mathbb{R}^n \times \mathbb{R}^m$ given by

$$\psi(x) = (x, f(x)).$$

By definition, ψ is a bijection between U and $\Gamma(f, U)$. In fact, it is more than a bijection: it is a *homeomorphism*. The set $\Gamma(f, U)$ is a (topological) subspace of $\mathbb{R}^n \times \mathbb{R}^m$, and it inherits a topology from this imbedding; the function ψ is a continuous bijection from U onto this space $\Gamma(f, U)$, and its inverse is also continuous. (This is left as an exercise.) So, the surface $\Gamma(f, U)$ is, topologically, the same as the open set U in $\mathbb{R}^n$.

Not every surface is the graph of a function, globally.

EXAMPLE 1.1. The *n-sphere* is the subset of $\mathbb{R}^{n+1}$ of points unit distance from the origin: $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} \colon |x| = 1\}$. It is not the graph of any function of any variable x^j in terms of the other n variables, as any coordinate line $x^j = c$ with $|c| < 1$ passes through $\mathbb{S}^n$ in two points. In terms of coordinates, $\mathbb{S}^n$ is the set of points $(x^1, \dots, x^{n+1})$ such that $(x^1)^2 + \cdots + (x^{n+1})^2 = 1$; that is, $\mathbb{S}^n$ is the level set of the function $\Phi(x^1, \dots, x^{n+1}) = (x^1)^2 + \cdots + (x^{n+1})^2$ at height 1. Now, $D\Phi(x) = 2[x^1, \dots, x^{n+1}]$. This row matrix is never 0 for $x \in \mathbb{S}^n$, which means that it is always full rank. Thus, at every point $x_0 \in \mathbb{S}^n$, we can find some coordinate x^j such that $\frac{\partial \Phi}{\partial x^j} \neq 0$, and by the Implicit Function Theorem, there exists a smooth local function $f\colon V \subseteq \mathbb{R}^n \to \mathbb{R}$ so that, near x_0, $\mathbb{S}^n$ is the graph of $x^j = f(x^1, \dots, x^{j-1}, x^{j+1}, \dots, x^n)$. Thus, even though $\mathbb{S}^n$ is not the graph of a smooth function globally, it *is* locally.

This motivates our more general definition.

DEFINITION 1.2. *Let $n < d$ be positive integers. A* **smooth n-dimensional surface in $\mathbb{R}^d$** *is a set $S \subset \mathbb{R}^d$ with the property that, for each point $x_0 \in S$, there is an open set $V_{x_0} \subseteq \mathbb{R}^n$, and a smooth function $f\colon V_{x_0} \to \mathbb{R}^{d-n}$, and some partition of coordinates $v = (x^{j_1}, \dots, x^{j_n})$ and $w = (x^{k_1}, \dots, x^{k_{d-n}})$ (where $\{j_1, \dots, j_n, k_1, \dots, k_{d-n}\} = \{1, \dots, d\}$), such that $x \in S$ with $x \sim (v, w)$ and $v \in V_{x_0}$ if and only if $w = f(v)$. In other words, a smooth n-dimensional surface in $\mathbb{R}^d$ is a subset which is locally the graph of $d - n$ variables as functions of the other n.*

This definition is set up precisely to conform to the Implicit Function Theorem. Indeed, let us make one more definition.

DEFINITION 1.3. *Let $\Phi\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be a smooth function. A point $c \in \mathbb{R}^m$ is called a* **regular value** *for Φ if, for any x in the level set $\Phi^{-1}(c)$, the derivative $D\Phi(x)$ has full rank m.*

COROLLARY 1.4. *If $\Phi\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is a smooth function and $c \in \mathbb{R}^m$ is a regular value of Φ, then the level set $\Phi^{-1}(c)$ is a smooth n-dimensional surface in $\mathbb{R}^{n+m}$.*

PROOF. Choose m pivotal columns in the matrix $D\Phi(x)$; then the $m \times m$ submatrix composed of these columns is invertible. The Implicit Function Theorem therefore implies that the level set $\Phi^{-1}(c)$ is locally the graph of a function of these pivotal m variables in terms of the remaining n variables. This is the definition of a smooth n-dimensional surface. $\square$

EXAMPLE 1.5. The *Orthogonal Group* $O(n)$ consists of those $n \times n$ real matrices Q with the property that $Q^\top Q = I_n$ (i.e. invertible matrices with $Q^{-1} = Q^\top$). Note that the condition is that, if Q_i is the ith column of Q, then $[Q^\top Q]^i_j = \langle Q_i, Q_j \rangle = \delta^i_j$; so the columns of Q form an orthonormal basis for $\mathbb{R}^n$. Of course, since $Q^{-1} = Q^\top$ we also have $QQ^\top = I_n$, and so the rows of Q also form an orthonormal basis for $\mathbb{R}^n$. It is a group under matrix multiplication: if $Q_1, Q_2 \in O(n)$, then $(Q_1Q_2)^{-1} = Q_2^{-1}Q_1^{-1} = Q_2^\top Q_1^\top = (Q_1Q_2)^\top$; and $(Q^{-1})^\top = (Q^\top)^{-1} = Q$, so $O(n)$ is closed under product and inverse (and clearly contains the identity I_n). Note that, if $x, y \in \mathbb{R}^n$ are vectors, and $Q \in O(n)$, then $\langle Qx, Qy \rangle = \langle Q^\top Qx, y \rangle = \langle x, y \rangle$; that is, $O(n)$ consists of linear isometries of $\mathbb{R}^n$. In fact, it is the group of all linear isometries of $\mathbb{R}^n$.

By definition, $O(n)$ is the level set $\Phi^{-1}(I_n)$ where $\Phi(A) = A^\top A$. The function Φ is a smooth map from $\mathbb{M}_n \to \mathbb{M}_n$ (indeed, all its entries are polynomial functions); here $\mathbb{M}_n$ is the n^2-dimensional vector space of $n \times n$ real matrices. It is better to view this map having a smaller codomain: the set $\mathbb{M}_n^{sa}$ of self-adjoint (i.e. symmetric) matrices, since $\Phi(A)^\top = (A^\top A)^\top = A^\top A = \Phi(A) \in \mathbb{M}_n^{sa}$. It is easy to check that the dimension of $\mathbb{M}_n^{sa}$ is $\frac{n(n+1)}{2}$. So we have a smooth function

$$\Phi \colon \mathbb{M}_n \to \mathbb{M}_n^{sa}$$

and our set of interest is the level set $O(n) = \Phi^{-1}(I_n)$. Now, let us compute the derivative of Φ. For any given $A \in \mathbb{M}_n$, $D\Phi(A) \colon \mathbb{M}_n \to \mathbb{M}_n^{sa}$ is the linear map given by

$$[D\Phi(A)](H) = A^\top H + H^\top A.$$

(Showing this is the case, from the definition of the derivative, is left as an exercise.) Now, consider this derivative at any point $Q \in O(n)$. In fact, if X is any symmetric matrix $X \in \mathbb{M}_n^{sa}$, then by taking $H = \frac{1}{2}QX$, note that

$$[DF(Q)](\tfrac{1}{2}XQ) = \tfrac{1}{2}Q^\top QX + \tfrac{1}{2}X^\top Q^\top Q = \tfrac{1}{2}(X + X^\top) = X.$$

This shows that the linear map $D\Phi(Q) \colon \mathbb{M}_n \to \mathbb{M}_n^{sa}$ is surjective, i.e. full-rank, for any $Q \in O(n)$. Thus, I_n is a regular value of Φ, and so by Corollary 1.4, $O(n)$ is a smooth surface in $\mathbb{M}_n$, whose dimension is $n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$.

2. Topological Manifolds

A smooth manifold is a generalization of a smooth surface. In fact, we will eventually see that it is no generalization at all: every n-dimensional smooth manifold *is* a smooth surface in $\mathbb{R}^d$ for some $d \leq 2n$. However, we do not want to be biased by the imbedding of our manifold, and instead view it as an abstract space. Before we get to *smooth* manifolds, we must first define (topological) manifolds.

DEFINITION 1.6. *An n-**dimensional manifold** $M = M^n$ is a second-countable Hausdorff space that is locally Euclidean. To be precise: M is*

- SECOND-COUNTABLE*: there is a countable collection $\mathscr{U} = \{U_n\}_{n \in \mathbb{N}}$ of open sets in M with the property that* any *open set in M is a union of elements of $\mathscr{U}$.*

- HAUSDORFF*: given any two distinct points $p, q \in M$, there are two open sets $U, V \subset M$ so that $p \in U$, $q \in V$, and $U \cap V = \varnothing$.*
- LOCALLY EUCLIDEAN*: given any $p \in M$, there is an open neighborhood $U \subseteq M$ of p that is homeomorphic to an open set in $\mathbb{R}^n$. Let $\widehat{U}$ be such an open set in $\mathbb{R}^n$, and let $\varphi\colon U \to \widehat{U}$ be a homeomorphism. The pair (U, φ) is called a* **coordinate chart** *near p.*

The second-countability condition is a requirement to prevent exotic topological spaces that are "too big" from being manifolds. The Hausdorff condition is there to prevent us from including exotic, pathological spaces (like the real line with two origins) from being manifolds. These are technical conditions; the heart of the definition is (3), that M is locally Euclidean.

PROPOSITION 1.7 (Smooth surfaces are manifolds). *If S is an n-dimensional smooth surface in $\mathbb{R}^d$, then S in an n-dimensional manifold.*

PROOF. Since $\mathbb{R}^d$ is second-countable and Hausdorff, so is any subspace, including S. Now, from the definition of smooth surface, we know that, for any $x_0 \in S$, there is a partition of coordinates $x \sim (v, w)$ with $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^{d-n}$, and an open set $V \subseteq \mathbb{R}^n$, and a smooth function $f\colon V \to \mathbb{R}^{d-n}$, so that the graph $\Gamma(f, V)$ coincides with the surface S near x_0. The function $\psi(x) = (x, f(x))$ is a homeomorphism onto its image, and so the set $U = \psi(V)$ is an open neighborhood of x_0 in S, with a homeomorphism $\varphi = \psi^{-1}\colon U \to V$. $\square$

EXAMPLE 1.8 (The n-sphere). As shown in the previous section, the n-sphere $\mathbb{S}^n$ is a smooth surface, and so by the previous proposition, $\mathbb{S}^n$ is an n-dimensional manifold. Let's looks specifically at some standard coordinate charts on it. For each $j \in \{1, \dots, n+1\}$, consider the (open) half spaces:

$$H_j^{\pm} = \{(x^1, \dots, x^{n+1}) \in \mathbb{R}^n\colon \pm x^j > 0\}.$$

Define $U_j^{\pm} = H_j^{\pm} \cap S^n$, the northern and southern hemispheres. It is easy to check that $U_j^{\pm}$ is the graph of the function $x_j = \pm f(x^1, \dots, \widehat{x^j}, \dots, x^{n+1})$, where $f\colon \mathbb{B}^n \to \mathbb{R}$ is the smooth function $f(u) = \sqrt{1 - |u|^2}$. Here $\mathbb{B}^n = \{u \in \mathbb{R}^n\colon |u| < 1\}$ is the unit ball, and the notation $(x^1, \dots, \widehat{x^j}, \dots, x^{n+1})$ means x^j is omitted from the list: i.e. this is the point

$$(x^1, \dots, \widehat{x^j}, \dots, x^{n+1}) = (x^1, \dots, x^{j-1}, x^{j+1}, \dots, x^{n+1}) \in \mathbb{R}^n.$$

Now, any point $x \in S^n$ is contained in at least one of the hemispheres $U_1^{\pm}, \dots, U_{n+1}^{\pm}$. Thus, for a point in $U_j^{\pm}$, we have the coordinate chart $\varphi_j^{\pm}\colon U_j^{\pm} \to \mathbb{B}^n$ given by

$$\varphi_j^{\pm}(x) = (x^1, \dots, \widehat{x^j}, \dots, x^{n+1})$$

which is a homeomorphism whose inverse is $y \mapsto (y, \pm f(y))$ (reordered so that the $f(y)$ coordinate is in the jth entry).

Now, let's consider a manifold that is *not* evidently a subset of a Euclidean space.

EXAMPLE 1.9 (Real Projective Space). The real projective space $\mathbb{RP}^n$ is defined to be the set of all 1-dimensional subspaces of $\mathbb{R}^{n+1}$; that is, points in $\mathbb{RP}^n$ are lines through the origin in $\mathbb{R}^{n+1}$. It is made into a topological space via the quotient map $\pi\colon \mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{RP}^n$, where $\pi(x) = \operatorname{span}\{x\}$. (This means that a set U in $\mathbb{RP}^n$ is open iff $\pi^{-1}(U)$ is open in $\mathbb{R}^{n+1} \setminus \{0\}$.) Note that $\mathbb{RP}^n$ can also be thought of as the quotient of the n-sphere by the equivalence relation $x \sim -x$ (i.e. antipodal points are identified). It is left as an exercise to the topologically-minded reader to prove (or look up the proof) that $\mathbb{RP}^n$ is second-countable and Hausdorff.

For $1 \leq j \leq n+1$, let $V_j \subset \mathbb{R}^{n+1} \setminus \{0\}$ be the set of points where $x^j \neq 0$, and let $U_j = \pi(V_j) \subset \mathbb{RP}^n$. Thus U_j is the set of 1-dimensional subspaces of $\mathbb{R}^{n+1}$ that are not parallel to the $x_j = 0$ plane. The preimage $\pi^{-1}(U_j)$ is easily seen to precisely equal V_j, which is open, and so by definition U_j is open in $\mathbb{RP}^n$. Since no 1-dimensional subspaces of $\mathbb{R}^{n+1}$ are parallel to all coordinate planes, it follows that every point in $\mathbb{RP}^n$ is contained in at least one of the U_j.

Now, define $\varphi_j \colon U_j \to \mathbb{R}^n$ by

$$\varphi(\pi(x^1, \ldots, x^{n+1})) = \frac{1}{x^j}(x^1, \ldots, \widehat{x^j}, \ldots, x^{n+1}).$$

This map is well-defined since $\pi(x) = \pi(\lambda x)$ for any non-zero scalar λ. Note that $\varphi \colon \pi \colon V_j \to \mathbb{R}^n$ is continuous, and so by the definition of the (quotient) topology on $\mathbb{RP}^n$, it follows that φ_j is continuous. What's more, it is a homeomorphism: it is straightforward to verify that its inverse is given by

$$\varphi_j^{-1}(u^1, \ldots, u^n) = \pi(u^1, \ldots, u^{j-1}, 1, u^{j+1}, \ldots, u^n),$$

which is also evidently continuous. Hence, $\mathbb{RP}^n$ is locally Euclidean (with patches in $\mathbb{R}^n$), and hence is an n-dimensional manifold.

Finally, an example for building manifolds out of others.

EXAMPLE 1.10 (Product Manifolds). Let $M_1, \ldots, M_k$ be manifolds of dimensions $n_1, \ldots, n_k$. Then $M_1 \times \cdots \times M_k$ is a manifold of dimension $n_1 + \cdots + n_k$. Indeed, the product of any finite (or countable) collection of second-countable Hausdorff spaces is second-countable Hausdorff. For the local Euclidean property, if $(p_1, \ldots, p_k) \in M_1 \times \cdots \times M_k$ is a point in the product, then choose a coordinate chart (U_j, φ_j) near each p_j, and note that $(U_1 \times \cdots \times U_k, \varphi_1 \times \cdots \times \varphi_k)$ is a coordinate chart near $(p_1, \ldots, p_k)$.

In particular, taking the circle $\mathbb{S}^1$ (which is the n-sphere in the case $n = 1$), the n-**torus** is the product $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1 = (\mathbb{S}^1)^n$. It is a (fun) n-dimensional manifold.

3. Smooth Charts and Atlases

Näively, we might expect that a smooth manifold should be one where the charts (U, φ) come with smooth functions φ. (This is certainly the case for the hemisphere charts on the n-sphere given in Example 1.8.) However, this doesn't really make sense in general: a manifold M is an abstract topological space, and so we do not know how to define smoothness for a function $\varphi \colon U \subseteq M \to \mathbb{R}^n$. In Example 1.8, the set M was constructed already as a subset of $\mathbb{R}^{n+1}$, and so the notion of smoothness is present. In general, we want to avoid this case, and work with manifolds abstractly.

In fact, we are going to turn this question around: the point will be to define what it means for a function $f \colon M \to \mathbb{R}^n$ to be smooth. Again, the näive choice is to define this locally: taking $p \in M$ and a chart (U, φ) at p, we should insist that the composite function $f \circ \varphi^{-1} \colon \varphi(U) \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be smooth. This is, in fact, the definition we will use, but we must be careful. In general, there will be many charts at p, and the definition of smoothness should not depend on which one we use. Thus, for this notion of smoothness to be well-defined, we require a consistency condition.

DEFINITION 1.11. *Let M be a manifold, and let (U, φ) and (V, ψ) be two charts such that $U \cap V \neq \varnothing$. The* **transition map** *from (U, φ) to (V, ψ) is the composite map*

$$\psi \circ \varphi^{-1} \colon \varphi(U \cap V) \to \psi(U \cap V).$$

Note that $\varphi(U \cap V)$ and $\psi(U \cap V)$ are open subsets of $\mathbb{R}^n$, and $\psi \circ \varphi^{-1}$ is a homeomorphism between them. The two charts are said to be **smoothly compatible** *if the transition map $\psi \circ \varphi^{-1}$ is a diffeomorphim.*

A collection of charts whose domains cover all of M is called an **atlas** *$\mathscr{A}$. If all the charts in $\mathscr{A}$ are smoothly compatible, we call it a* **smooth atlas**.

EXAMPLE 1.12. Following Example 1.8, consider the two charts (U_1^+, φ_1^+) and (U_2^+, φ_2^+). Note that $U_1^+ \cap U_2^+$ is the portion of the circle in the first quadrant, and we have $\varphi_1^+(U_1^+ \cap U_2^+) = \varphi_2^+(U_1^+ \cap U_2^+) = (0,1)$, and $\varphi_1^+(x,y) = y$ while $\varphi_2^+(x,y) = x$. Then $(\varphi_2^+)^{-1}(x) = (x, \sqrt{1-x^2})$, and so the transition map from φ_2^+ to φ_1^+ is

$$\varphi_1^+ \circ (\varphi_2^+)^{-1}(x) = \sqrt{1-x^2}.$$

This is a smooth map on the unit interval; in fact it is a diffeomorphism there. Similar calculations show that all the transition maps in the atlas $\mathscr{A} = \{(U_j^\pm, \varphi_j^\pm) \colon 1 \leq j \leq n+1\}$ are diffeomorphisms, making this a smooth atlas on $\mathbb{S}^{n+1}$.

EXAMPLE 1.13. Following Example 1.9, consider the atlas of charts (U_j, φ_j) on $\mathbb{RP}^n$. Assuming, for convenience, that $i > j$, we can compute that

$$\varphi_j \circ \varphi_i^{-1}(u^1, \ldots, u^n) = \frac{1}{u^j}\left(u^1, \ldots, \widehat{u^j}, \ldots, u^{i-1}, 1, u^i, \ldots, u^n\right),$$

which is a smooth map from $\varphi_i(U_i \cap U_j)$ onto $\varphi_j(U_i \cap U_j)$. The inverse is given by a similar formula (with the 1 first and omitted variable second), and so the transition maps are all diffeomorphisms. Hence, this is a smooth atlas for $\mathbb{RP}^n$.

EXAMPLE 1.14. Given smooth atlases on $M_1, \ldots, M_k$, their products give a smooth atlas on $M_1 \times \cdots \times M_k$.

We would now like to define a **smooth manifold** to be a manifold in possession of a smooth atlas. But this is not careful enough, for two reasons.

- It is possible for a manifold to possess two or more smooth atlases that are not smoothly compatible with each other. This means that the definition of smooth function on the manifold will again be ill-defined: a given function's smoothness depends on *which* atlas one chooses to use. Hence, the smooth atlas itself must be a part of the definition – different smooth atlases on the same manifold define *different* smooth manifolds.
- On the other hand, it may well be that two distinct smooth atlases give rise to precisely the same class of smooth functions. In this case, we do not wish to consider the two manifold-atlas pairs to be distinct smooth manifolds. We could address this point by defining a smooth manifold to be an equivalence class of manifold-atlas pairs, but there is an easier way, as follows.

4. Smooth Structures

DEFINITION 1.15. *Let M be a manifold. A smooth atlas $\mathscr{A}$ on M is called* **complete** *or* **maximal** *if it is not properly contained in any larger smooth atlas. That is, $\mathscr{A}$ is complete if, given any chart (U, φ) that is smoothly compatible with all the charts in $\mathscr{A}$, it follows that $(U, \varphi) \in \mathscr{A}$.*

A complete atlas on M is called a **smooth structure** *on M. A* **smooth manifold** *is a pair $(M, \mathscr{A})$, where M is a manifold, and $\mathscr{A}$ is a smooth structure on M.*

The apparent disadvantage of the complete atlas definition of a smooth structure is that such objects are very large, and difficult to describe. For example, the set of all charts smoothly compatible with the $2(n+1)$ charts in the "hemispheres" atlas of $\mathbb{S}^n$ in Example 1.8 includes charts with *all possible open sets* U as domains, and then on each one a huge uncountable mess of coordinate chart functions φ. Not to worry, though: we can still talk about *the* smooth structure on the sphere by specifying only those $2(n+1)$ charts, due to the following.

LEMMA 1.16. *Let M be a manifold, and let $\mathscr{A}$ be a smooth atlas on M. Then there is a unique smooth structure $\overline{\mathscr{A}}$ on M that contains $\mathscr{A}$. (It is called the* **smooth structure determined by** *$\mathscr{A}$.)*

PROOF. We define $\overline{\mathscr{A}}$ be be the set of all charts that are smoothly compatible with every chart in $\mathscr{A}$. Since $\mathscr{A}$ is a smooth atlas, $\mathscr{A} \subseteq \overline{\mathscr{A}}$. In particular, $\overline{\mathscr{A}}$ is an atlas. We need to show three things.

(1) $\overline{\mathscr{A}}$ is a smooth atlas (i.e. all its charts are smoothly compatible).
(2) $\overline{\mathscr{A}}$ is a smooth structure (i.e. it is complete).
(3) $\overline{\mathscr{A}}$ is the *unique* smooth structure containing $\mathscr{A}$.

(1) We must show that if (U,φ) and (V,ψ) are two charts in $\overline{\mathscr{A}}$ with $U \cap V \neq \varnothing$, they are smoothly compatible. That is, we must show that $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ is smooth. (We actually need to show it is a diffeomorphism, but in the proof we do this for all pairs of charts, including these two in reverse, showing that the inverse is also smooth.)

Fix a point $p \in U \cap V$, and let $x = \varphi(p) \in \varphi(U \cap V)$. Since $\mathscr{A}$ is an atlas, there is some chart $(W,\vartheta) \in \mathscr{A}$ with $p \in W$. By the definition of $\overline{\mathscr{A}}$, (U,φ) and (V,ψ) are both smoothly compatible with (W,ϑ). Hence, both of the maps $\vartheta \circ \varphi^{-1}$ and $\psi \circ \vartheta$ are smooth on their domains. Thus $\psi \circ \varphi^{-1} = (\psi \circ \vartheta^{-1}) \circ (\vartheta \circ \varphi^{-1})$ is smooth on a neighborhood of x. This was for an arbitrary point $x \in \varphi(U \cap V)$, and smoothness is a local property, so $\psi \circ \varphi^{-1}$ is indeed smooth on $\varphi(U \cap V)$. This shows that the two charts are smoothly compatible, and so $\overline{\mathscr{A}}$ is an atlas.

(2) We now show that $\overline{\mathscr{A}}$ is a complete smooth atlas. But this is pretty easy: since $\mathscr{A} \subseteq \overline{\mathscr{A}}$, if (U,φ) is any chart that is smoothly compatible with every chart in $\overline{\mathscr{A}}$, then it is automatically smoothly compatible with every chart in $\mathscr{A}$ – which means, by definition, that $(U,\varphi) \in \overline{\mathscr{A}}$.

(3) Let $\mathscr{B}$ be any smooth structure containing $\mathscr{A}$. Since $\mathscr{B}$ is an atlas, all its charts are smoothly compatible, and so in particular every chart in $\mathscr{B}$ is smoothly compatible with every chart in $\mathscr{A}$. But that means, by definition, that $\mathscr{B} \subseteq \overline{\mathscr{A}}$. But $\mathscr{B}$ is assumed to be complete, and since every chart in $\overline{\mathscr{A}}$ is smoothly compatible, it follows that $\mathscr{B} = \overline{\mathscr{A}}$, concluding the proof. □

EXAMPLE 1.17. The Euclidean space $\mathbb{R}^n$ comes with a standard smooth structure, which is the smooth structure determined by the single coordinate chart $(\mathbb{R}^n, \mathrm{Id}_{\mathbb{R}^n})$. But there are other distinct smooth structures on $\mathbb{R}^n$ as well. For example when $n = 1$, consider the chart $(\mathbb{R},\psi)$ where $\psi(x) = x^3$, which is a homeomorphism $\mathbb{R} \to \mathbb{R}$. By Lemma 1.16, there is a unique smooth structure on $\mathbb{R}$ containing this chart. It is *not* the standard smooth structure; for, if it were, the two charts $(\mathbb{R}, \mathrm{Id}_{\mathbb{R}})$ and $(\mathbb{R},\psi)$ would be smoothly compatible. But note that $\mathrm{Id}_{\mathbb{R}} \circ \psi^{-1}$ is the map $y \mapsto y^{1/3}$, which is not smooth.

Let's consider now a few more examples of smooth manifolds.

EXAMPLE 1.18. Let M be a smooth manifold, with complete smooth atlas $\mathscr{A}$. If $W \subseteq M$ is any open subset, then W is also a smooth manifold, with smooth structure determined by the atlas $\{(U,\varphi)\colon (U,\varphi) \in \mathscr{A}, U \subseteq W\}$. Note: since $\mathscr{A}$ is complete, its coordinate charts that happened

to be contained in W already included a covering of W, so the resulting collection of restricted charts is indeed an atlas.

So, in particular, any open subset of $\mathbb{R}^n$ is a smooth manifold. An example of this form gives us another member of the classical family of Lie groups: $GL(n)$, the group of all invertible $n \times n$ matrices. Since it can be described as $\det^{-1}(\mathbb{R} \setminus \{0\})$, and $\det$ is continuous while $\mathbb{R} \setminus \{0\}$ is open, it follows that $GL(n)$ is an open subset of $\mathbb{M}_n \cong \mathbb{R}^{n^2}$. Thus, $GL(n)$ is a smooth manifold (with the subspace smooth structure).

In fact, $GL(n)$ is an open *dense* subset of $\mathbb{M}_n$. To see this, let $A \in \mathbb{M}_n$ be any matrix. Let $p_A(\lambda) = \det(A - \lambda I_n)$ be its characteristic polynomial. This polynomial has n complex roots (the eigenvalues of A). Let $r = \min\{|\lambda| \colon \lambda \neq 0, p_A(\lambda) = 0\}$. Then $A - \epsilon I_n$ is invertible for $0 < \epsilon < r$. This shows $GL(n)$ is dense in $\mathbb{M}_n$.

The next example, the *Grassmannian manifold* of k-planes in $\mathbb{R}^n$, is a generalization of real projective space.

EXAMPLE 1.19 (Grassmannians). For $0 \leq k \leq n$, let $\mathrm{Gr}(n, k)$ denote the set of all k-dimensional subspaces of $\mathbb{R}^n$. (So, for example, $\mathrm{Gr}(n+1, 1) = \mathbb{RP}^n$ as in Example 1.9.) Problem 7 on Homework 1 shows how to define the standard smooth structure on the Grassmannian that makes it into a smooth manifold.

In fact, Homework 1 Problem 7 relies on Homework 1 Problem 6, which is essentially the same as Lemma 1.35 (the "Smooth Manifold Charts Lemma") in [**3**]. We state this again as a proposition here, since it allows us to dispense with a lot of the separate topological preliminaries when working with a smooth manifold. Also, to be slightly more general, instead of using a fixed Euclidean space $\mathbb{R}^n$, we allow the actual model Euclidean space to vary from point to point (as long as the dimension is fixed), specifying a Hilbert space for each chart. Note that all the usual notions (the norm topology, open sets, continuous and smooth maps) make sense in a finite-dimensional Hilbert space, without imposing the coordinate structure of $\mathbb{R}^n$.

PROPOSITION 1.20. *Let M be a nonempty set. Let J be an index set. Let n be a positive integer. For each $j \in J$, suppose*

- *U_j is a nonempty subset of M,*
- *H_j is a real Hilbert space of dimension n, and*
- *$\varphi_j \colon U_j \to H_j$ is an injective map whose image is open in H_j.*

Furthermore, suppose that

- *there is a countable subset $I \subset J$ so that $\bigcup_{i \in I} U_i = M$, and*
- *for any two points $p \neq q$ in M, either there is a single U_j with $p, q \in U_j$, or there are disjoint $U_j \cap U_k = \varnothing$ with $p \in U_j$ and $q \in U_k$.*

Finally, suppose that each of the transition maps

$$\varphi_k \circ \varphi_j^{-1} \colon \varphi_j(U_j \cap U_k) \to \varphi_k(U_j \cap U_k), \quad j, k \in J$$

is smooth (as a map from an open subset of H_j to H_k). Then there is a unique topology on M for which each φ_j is a homeomorphism onto its image, making M a topological manifold. Moreover, if $\lambda_j \colon H_j \to \mathbb{R}^n$ is a given isometric isomorphism, then the set $\mathscr{A} \equiv \{(U_j, \lambda_j \circ \varphi_j) \colon j \in J\}$ is a smooth atlas on M.

The proof of Proposition 1.20 is left as Problem 6 on Homework 1. We will use this proposition frequently. Moreover, from here on, we will use the terms *chart* and *smooth atlas* to refer to the nominally more abstract objects above, where the coordinate patches are in finite-dimensional

Hilbert spaces (of fixed dimension) whose representation may vary from chart to chart, rather than necessarily a fixed $\mathbb{R}^n$.

REMARK 1.21. We could even loosen the a priori requirement that all the model spaces H_j have the same dimension n, only requiring that they be finite dimensional. In this case, it follows from the last requirement (that the transitions maps be smooth) that the dimension of H_j is constant along each connected component of M; the proof of this is also part of Problem 6 on Homework 1. But this definition would allow, for example, the disjoint union of $\mathbb{S}^1$ and $\mathbb{S}^2$ to be considered a smooth manifold, meaning dimension would not be well-defined for disconnected manifolds. We will exclude this possibility.

CHAPTER 2

Smooth Maps

Now having defined a smooth manifold (with the idea of giving meaning to smooth functions on said manifolds), we can define smooth maps. Note: the words *map* and *function* are usually used interchangeably; here, we will try to be consistent about reserving the word function for a map whose codomain is a Euclidean space $\mathbb{R}^n$, while map could mean a function between any two manifolds.

NOTATION 2.1. A smooth manifold is a pair $(M, \mathscr{A})$ where $\mathscr{A}$ is a smooth structure on M. It will be more convenient to just say "let M be a smooth manifold," with the given smooth structure understood as part of the given data. In this case, if we must refer to the smooth structure explicitly, we will use the notation $\mathscr{A}_M$.

1. Definitions, Basic Properties

DEFINITION 2.2. *Let M and N be smooth manifolds. A map $F\colon M \to N$ is called* smooth *if, for each $p \in M$, there is a chart $(U, \varphi) \in \mathscr{A}_M$ with $p \in U$, and a chart $(V, \psi) \in \mathscr{A}_N$ with $F(p) \in V$, such that $F(U) \subseteq V$, and the composite map*

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U) \to \psi(V)$$

is smooth.

In other words: a smooth map is a map that is smooth in local coordinates. This is well-defined: suppose we have two charts (U_j, φ_j) at p and (V_j, ψ_j) at $F(p)$ for $j = 1, 2$. Since they are each chosen from a smooth atlas, the transition maps between them are smooth. Thus, if F is smooth (in the above sense) with respect to (U_1, φ_1) and (V_1, ψ_1), then on $U_1 \cap U_2$

$$\psi_2 \circ F \circ \varphi_2^{-1} = (\psi_2 \circ \psi_1^{-1}) \circ (\psi_1 \circ F \circ \varphi_1^{-1}) \circ (\varphi_1 \circ \varphi_2^{-1})$$

is a composition of smooth maps, and is also smooth. This is precisely the reason for the smooth compatibility condition in the definition of smooth atlas to begin with.

There is one slight subtlety in the definition: we require that the chart in the codomain (V, ψ) satisfy $F(U) \subseteq V$. This is necessary to even make sense of the composition $\psi \circ F$ on all of U.

PROPOSITION 2.3. *Let M, N be smooth manifolds. A smooth map $F\colon M \to N$ is continuous.*

PROOF. Fix a point $p \in M$. Choose charts (U, φ) at p and (V, ψ) at $F(p)$ so that $F(U) \subseteq V$ and $\psi \circ F \circ \varphi^{-1}$ is smooth on $\varphi(U)$. This composite map is therefore continuous. Thus, on U, we have

$$F|_U = \psi^{-1} \circ (\psi \circ F \circ \varphi^{-1}) \circ \varphi$$

is a composition of continuous maps, and hence is continuous. This shows that F is continuous on a neighborhood of the arbitrary point $p \in M$; thus F is continuous. $\square$

Using the continuity of F, the following alternate characterization of smoothness is often useful.

PROPOSITION 2.4. *Let M, N be smooth manifolds. A map $F\colon M \to N$ is smooth if and only if the following holds true. For every $p \in M$, there is a smooth chart (U, φ) at p and a smooth chart (V, ψ) containing $F(p)$ such that $U \cap F^{-1}(V)$ is open in M and*

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \to \psi(V)$$

is a smooth map.

The proof is left as an exercise. Note: the condition that $U \cap F^{-1}(V)$ is open here is a replacement for the choice of U, V with $F(U) \subseteq V$ in the definition. Without this explicit requirement, the alternate characterization of smoothness would allow for discontinuous maps.

EXAMPLE 2.5. Let $F\colon \mathbb{R} \to \mathbb{R}$ be $F = \mathbb{1}_{[0,\infty)}$: $F(x) = 0$ if $x < 0$ and $F(x) = 1$ if $x \geq 1$. Here we take the usual smooth structure determined by $(\mathbb{R}, \mathrm{Id}_{\mathbb{R}})$ for $\mathbb{R}$ in the domain and codomain. For any $x \neq 0$, one can take either $(-\infty, 0)$ or $(0, \infty)$ (equipped with the identity map) as a chart in the domain, and $\mathbb{R}$ in the codomain, to see that F is smooth (of course) near x. For $x = 0$, take $U = (-\frac{1}{2}, \frac{1}{2})$ and $V = (\frac{1}{2}, \frac{3}{2})$. Then $F^{-1}(V) = [0, \infty)$, and so $U \cap F^{-1}(V) = [0, \frac{1}{2})$, which is not open in $\mathbb{R}$. Ignoring this condition, however, and noting that all the coordinate functions are $\mathrm{Id}_{\mathbb{R}}$, the composition $\psi \circ F \circ \varphi^{-1}$ on $\varphi(U \cap F^{-1}(V)) = [0, \frac{1}{2})$ is just the map $\psi \circ F \circ \varphi^{-1}(x) = x$, which is smooth (in the sense that it is the restriction of a smooth function on $\mathbb{R}$ to the set $[0, \frac{1}{2})$). Hence, the condition $U \cap F^{-1}(V)$ be open is crucial for smoothness to carry its usual meaning (and imply continuity).

The next lemma follows immediately from the definition of smoothness, and its proof is left as an exercise.

LEMMA 2.6 (Smoothness is Local). *Let M, N be smooth manifolds. A map $F\colon M \to N$ is smooth iff for every $p \in M$, there is an open neighborhood $U \subseteq M$ containing p so that $F|_U$ is smooth.*

COROLLARY 2.7 (Smooth Gluing Lemma). *Let M, N be smooth manifolds. Let $\{U_j : j \in J\}$ be an open cover of M. For each j, suppose $F_j\colon U_j \to N$ is a smooth map. Moreover, suppose that $F_j|_{U_j \cap U_k} = F_k|_{U_j \cap U_k}$ for all $j, k \in J$. Then there is a unique smooth map $F\colon M \to N$ so that $F|_{U_j} = F_j$ for each $j \in J$.*

PROOF. For any $p \in M$, choose some $U_j \ni p$, and define $F(p) = F_j(p)$. This is well-defined by the overlapping consistency condition: if $p \in U_k$ as well, then $F_k(p) = F_j(p) = F(p)$. Clearly this is the unique function F with the desired property, so it remains only to show that F is smooth. This follows from Lemma 2.6: for any $p \in M$, choose some $U_j \ni p$; on this open neighborhood, $F|_{U_j} = F_j$ is smooth, and hence F is smooth. □

2. Smooth Functions, and Examples

In the special case that our target manifold is a Euclidean manifold, we refer to a smooth map $f\colon M \to \mathbb{R}^k$ as a *smooth function*. By definition, this means that, for each $p \in M$, there is a chart (U, φ) at p, and a chart (V, ψ) at $f(p)$ in $\mathbb{R}^k$, such that $f(U) \subseteq V$ and $\psi \circ f \circ \varphi^{-1}$ is smooth on $\varphi(U)$. In fact, it suffices to always take $V = \mathbb{R}^k$ and $\psi = \mathrm{Id}$, giving the following apparently stronger definition.

DEFINITION 2.8. *Let M be a smooth manifold, and let k be a positive integer. A function $f\colon M \to \mathbb{R}^k$ is called* **smooth** *if, for each $p \in M$, there is a chart (U, φ) so that the function*

$$f \circ \varphi^{-1}\colon \varphi(U) \to \mathbb{R}^k$$

is smooth. This composite function is called the **coordinate representation** *of f in the chart (U, φ).*

If this holds true, then evidently f is smooth in the sense of Definition 2.2 (taking $(V, \psi) = (\mathbb{R}^k, \mathrm{Id})$). Conversely, if $f\colon M \to \mathbb{R}^k$ is smooth in the sense of Definition 2.2 with some codomain chart (V, ψ) satisfying $f(U) \subseteq V$, then note that

$$f \circ \varphi^{-1} = \psi^{-1} \circ (\psi \circ f \circ \varphi^{-1})$$

is well-defined since $f(U)$ is contained in V, and is a composition of smooth functions, thus is is smooth. Thus, the two definitions are equivalent. A similar argument shows that if M is also an open subset of a Euclidean space $\mathbb{R}^n$, then this notion of smoothness coincides with the usual calculus definition.

In the special case $k = 1$, the set of smooth functions $f\colon M \to \mathbb{R}$ is denoted $C^\infty(M)$. In this case, since we can add and multiply the valued $f(p)$ for any $p \in M$, $C^\infty(M)$ has the structure of a commutative algebra. This algebra can be used to recover the topological properties of the manifold; this approach is the beginning of algebraic geometry.

Let us now consider some examples of smooth maps.

EXAMPLE 2.9. Let $M = \{(x, y) \in \mathbb{R}^2\colon x > 0\}$. The function $f(x, y) = x^2 + y^2$ is smooth in the classical sense. Let's consider its coordinate representation in a different chart (other than the identity chart on M, which is an open subset of $\mathbb{R}^2$. Let $\varphi\colon (-\frac{\pi}{2}, \frac{\pi}{2}) \times (0, \infty) \to M$ be the global polar coordinate chart $\varphi(r, \theta) = (r\cos\theta, r\sin\theta)$. Then the coordinate representation of f, $\hat{f} = f \circ \varphi^{-1}\colon (-\frac{\pi}{2}, \frac{\pi}{2}) \times (0, \infty)$, is given by $\hat{f}(r, \theta) = r^2$ (which is, just as clearly, smooth). It is not uncommon in differential geometry to confuse the notations f and $\hat{f}$, and simply refer to local coordinate representation of f in the (r, θ) coordinate as $f(r, \theta) = r^2$. (This is common even in most multivariate calculus classes.) We will try to avoid this, as it is confusing; but you will come across it at some point, so try not to be afraid of what will look like obvious nonsense.

EXAMPLE 2.10. It is easy to verify that all of the following maps are smooth on any manifold M:

- Any constant map.
- The identity map $M \to M$.
- If $U \subseteq M$ is open, then the inclusion map $U \hookrightarrow M$ is smooth.

EXAMPLE 2.11. Let $\iota\colon \mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1}$ denote the inclusion. Then, using the atlas of hemispheres (cf. Example 1.8) with $\varphi_j^\pm(x^1, \dots, x^{n+1}) = (x^1, \dots, \widehat{x^j}, \dots, x^{n+1})$, we have the coordinate representations of ι, $\hat{\iota}\colon \varphi_j^\pm(U_j^\pm) \to \mathbb{R}^{n+1}$ given by

$$\hat{\iota}(u^1, \dots, u^n) = (u^1, \dots, u^{j-1}, \pm\sqrt{1 - (u^1)^2 - \cdots - (u^n)^2}, u^j, \dots, u^n)$$

which is smooth on the domain $\varphi_j^\pm(U_j^\pm) = \mathbb{B}^n = \{(u^1, \dots, u^n)\colon (u^1)^2 + \cdots + (u^n)^2 < 1\}$. Thus, ι is a smooth map.

EXAMPLE 2.12. Let $\pi\colon \mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{RP}^n$ be the defining projection $\pi(x) = \operatorname{span}_{\mathbb{R}}\{x\}$. Then, in the coordinate charts (U_j, φ_j) for $\mathbb{RP}^n$ given in Example 1.9, and using the identity coordinates on $\mathbb{R}^{n+1} \setminus \{0\}$ (restricted to the set $\{x_j \neq 0\}$ so that π maps them into U_j), we have

$$\varphi_j \circ \pi(x^1, \dots, x^{n+1}) = \frac{1}{x^j}(x^1, \dots, \widehat{x^j}, \dots, x^{n+1}).$$

This is a smooth map on its domain, and so $\pi\colon \mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{RP}^n$ is smooth.

We can also build new smooth functions from old ones in the usual ways.

PROPOSITION 2.13. *Let M, N, P be smooth manifolds. Let k be a positive integer.*

(a) *If $f, g\colon M \to \mathbb{R}^k$ are smooth functions, then $f + g\colon M \to \mathbb{R}^k$ is a smooth function, and $f \cdot g\colon M \to \mathbb{R}$ (the dot product in the range) is a smooth function.*
(b) *If $F\colon M \to N$ and $G\colon N \to P$ are smooth, then so is $G \circ F\colon M \to P$.*

PROOF. Part (a) is left as an exercise. For part (b): fix a point $p \in M$. By smoothness of G, there is a chart (V, ψ) at $F(p)$ in N, and a chart (W, ϑ) a $G(F(p))$ in P with $G(V) \subseteq W$, such that $\vartheta \circ G \circ \psi^{-1}\colon \psi(V) \to \vartheta(W)$ is smooth. Since F is smooth, it is continuous, and so $F^{-1}(V)$ is an open neighborhood of p in M. By the completeness of the atlas on M, we may choose a chart (U, φ) at p so that $U \subseteq F^{-1}(V)$, which implies that $F(U) \subseteq V$. By the smoothness of F (and the smooth compatibility of the atlas $\mathscr{A}_M$), we have $\psi \circ F \circ \varphi^{-1}\colon \varphi(U) \to \psi(V)$ is smooth. Now, $(G \circ F)(U) = G(F(U)) \subseteq G(V) \subseteq W$, and on $\varphi(U)$,

$$\vartheta \circ (G \circ F) \circ \varphi^{-1} = (\vartheta \circ G \circ \psi^{-1}) \circ (\psi \circ F \circ \varphi^{-1})$$

is a composition of smooth maps between Euclidean spaces, and so is smooth. Thus, by definition, $G \circ F$ is smooth. □

EXAMPLE 2.14. The composition $\pi \circ \iota\colon \mathbb{S}^n \to \mathbb{RP}^n$, with ι and π from Examples 2.11 and 2.12, is therefore a smooth map. This is the map which sense $x \in \mathbb{S}^n$ to $\operatorname{span}_{\mathbb{R}}\{x\} \in \mathbb{RP}^n$. It is a 2-to-1 covering map, since the preimage of any element $\pi(x) \in \mathbb{RP}^n$ consists of the two antipodal points $\{x, -x\}$ in $\mathbb{S}^n$. This is the usual map by which we visualize $\mathbb{RP}^n$, as the quotient of $\mathbb{S}^n$ by identifying antipodal points. We now see that this quotient map is smooth.

As usual, "vector-valued" maps are smooth iff all their components are smooth.

PROPOSITION 2.15. *Let $M_1, \dots, M_k$ and N be smooth manifolds. Let $\pi_j\colon M_1 \times \cdots M_k \to M_j$ be the projection map. Then π_j is smooth for each j. Moreover, if $F\colon N \to M_1 \times \cdots M_k$ is any map, then F is smooth if and only if $F_j = \pi_j \circ F\colon N \to M_j$ is smooth for $1 \leq j \leq k$.*

The proof of Proposition 2.15 is left as an exercise.

3. Diffeomorphisms

Now that we know what "smooth" means for maps between manifolds, we can define the basic morphism in the category of smooth manifolds.

DEFINITION 2.16. *Let M, N be smooth manifolds. A* **diffeomorphism** *$F\colon M \to N$ is a smooth map that possesses a smooth inverse. If there exists a diffeomorphism $M \to N$, we say that M and N are* **diffeomorphic***; we write $M \cong N$.*

EXAMPLE 2.17. (1) The open unit ball $\mathbb{B}^n$ is diffeomorphic to the whole Euclidean space $\mathbb{R}^n$, via the map

$$F(x) = \frac{x}{\sqrt{1-|x|^2}}, \quad \text{whose inverse is } \; F^{-1}(y) = \frac{y}{\sqrt{1+|y|^2}}.$$

So $\mathbb{B}^n \cong \mathbb{R}^n$.

(2) Let M be a smooth manifold and let (U, φ) be a chart in a smooth atlas. Then $\varphi\colon U \to \varphi(U)$ is a diffeomorphism (its local coordinate representation in the φ-coordinates is the identity map, in either direction). So: a smooth manifold is *locally diffeomorphic* to $\mathbb{R}^n$.

One key immediate property of diffeomorphisms is invariance of dimension (which is extremely hard to prove for homeomorphisms).

THEOREM 2.18 (Invariance of Dimension). *Let M, N be smooth manifolds. If $M \cong N$, then* $\dim M = \dim N$.

(Here $\dim M = n$ if M is an n-dimensional manifold.)

PROOF. Let $F\colon M \to N$ be a diffeomorphism. For any point $p \in M$, choose smooth charts (U, φ) at p and (V, ψ) and $F(p)$ with $F(U) \subseteq V$. Because F is a homeomorphism, it is an open map, and so $F(U)$ is open; by restriction if necessary, wlog we may therefore take $V = F(U)$. Then $\psi \circ F \circ \varphi^{-1}\colon \varphi(U) \to \psi(V)$ is a diffeomorphism between Euclidean spaces, with inverse $\varphi \circ F^{-1} \circ \psi^{-1}$. But we know (cf. Section 5) that diffeomorphisms can only exist between same-dimensional Euclidean spaces. This concludes the proof. □

Here is a list of basic properties of diffeomorphisms that are straightforward to verify.

PROPOSITION 2.19. (1) *A composition of diffeomorphisms is a diffeomorphism.*
(2) *A finite Cartesian product of diffoeomorphisms is a diffeomorphism.*
(3) *A diffeomorphism is a homeomorphism.*
(4) *If $F\colon M \to N$ is a diffeomorphism between smooth manifolds, and $U \subseteq M$ is open, then $F|_U : U \to F(U)$ is a diffeomorphism.*
(5) *The relation $M \cong N$ (diffeomorphic) is an equivalence relation on smooth manifolds.*

The last point in the proposition shows that we really want to think of a smooth manifold as an equivalence class under the diffeomorphism relation: we don't want to consider two smooth manifolds to be different if they are, in fact, diffeomorphic.

EXAMPLE 2.20. Consider the alternate smooth structure on $\mathbb{R}$ from Example 1.17, where the smooth atlas is determined by the global chart $(\mathbb{R}, \varphi(x) = x^3)$. Denote this smooth manifold by $\widetilde{R}$. This is not the standard smooth structure on $\mathbb{R}$. However, consider the map

$$F\colon \mathbb{R} \to \widetilde{R}, \quad F(x) = x^{1/3}.$$

Then in local (i.e. global) coordinates, with $(\mathbb{R}, \mathrm{Id})$ be the chart on $\mathbb{R}$, we have

$$\varphi \circ F \circ \mathrm{Id}^{-1}(x) = \varphi(x^3) = x, \qquad \mathrm{Id} \circ F^{-1} \circ \varphi^{-1}(y) = F^{-1}(y^{1/3}) = y.$$

Thus, in local coordinates, F is the identity map, which is clearly a diffeomorphism. We have therefore shown that $\mathbb{R} \cong \widetilde{R}$.

So $\mathbb{R}$ and $\widetilde{\mathbb{R}}$ are "the same" after all (as smooth manifolds, they are diffeomorphic). This leads naturally to the question: are there any smooth structures on the topological manifold $\mathbb{R}$ that are *not* diffeomorphic to the standard smooth structure? It turns out the answer is *no* (and we may get

to develop all the tools needed to prove that theorem in this course). A wider question is: given two topological manifolds M and N that are homeomorphic, is it possible to find smooth structures on them that are *not diffeomorphic*? This is a very difficult question in general, and is still a topic of much current research. Here are are few answers in specific cases that are pretty mind-blowing.

- If $\dim M \leq 3$, then there exists a unique (up to diffeomorphism) smooth structure on M. [Munkres 1960; Moise 1977]
- If $n \neq 4$, then the standard smooth structure on $\mathbb{R}^n$ is the only one up to diffeomorphism. *However*, there are *uncountably many* non-diffeomorphic smooth structures on $\mathbb{R}^4$. These are called "fake $\mathbb{R}^4$s". [Donaldson and Freedman, 1984].
- There are exactly 15 non-diffeomorphic smooth structures on $\mathbb{S}^7$. [Milnor, 1956; Kervaire and Smale, 1963].
- If $n > 3$, there exists a compact topological n-manifold which possesses no smooth structures at all.

4. Partitions of Unity

Since we must always work locally on a manifold, we want to be able to put local objects together into a global object. The "smooth gluing lemma" Corollary 2.7 is an example. But it is too weak for most purposes: there we must know how to extend our smooth function from one neighborhood to another where they are *identical* on the (big) intersection; this turns out to be generally impossible to arrange if we didn't already know of the existence of a global smooth function in the first place. If we only wanted continuity, we only need to arrange agreement on the intersections of *closed* sets, and this is much easier. But there's no way to get smoothness this way.

The main tool that we will use throughout this course to overcome this technical problem is called a *partition of unity*. The first step is to refine our cover of coordinate charts so that they do not overlap *too much*, and the resulting diffeomorphic patches of $\mathbb{R}^n$ are open balls.

Let M be a manifold, and let $\mathscr{O}$ be an open cover. Another open cover $\mathscr{O}'$ of M is called a **refinement** of $\mathscr{O}$ if, for any $U \in \mathscr{O}'$, there is some $V \in \mathscr{O}$ with $U \subseteq V$. A cover $\mathscr{O}$ is called **locally-finite** if, for any $p \in M$, there is a niehgborhood of p that only intersects finitely many $U \in \mathscr{O}$. Our first lemma here shows the real reason we need to assume our manifolds are second-countable: this guarantees the existence of *locally-finite* open covers.

LEMMA 2.21. *Let M be a smooth manifold, and let $\mathscr{O}$ be an open cover of M. Then there is an open cover $\mathscr{O}'$ of coordinate charts that refines $\mathscr{O}$, is locally-finite, and each coordinate chart in $\mathscr{O}'$ has compact closure, and is diffeomorphic to $\mathbb{R}^n$ (or $\mathbb{B}^n$).*

PROOF. By second-countability, M is σ-compact (Exercise): there is a countable collection of compact sets $K_1, K_2, K_3, \ldots$ so that $M = \bigcup_n K_n$. Now, let $V_1 \supset K_1$ be an open set with compact closure. Then $\overline{V_1} \cup K_2$ is compact, and similarly has an open neighborhood V_2 with compact closure. Continuing this way, we produce a nested sequence $V_1, V_2, V_3, \ldots$ of open sets, with $\overline{V_n} \subset V_{n+1}$, each with compact closure, such that $K_1 \cup \cdots \cup K_n \subset V_n$. It follows that $M = \bigcup_n V_n$. Then we also have $M = \bigcup_n A_n$ where A_n are the closed annular regions $A_n = \overline{V_n} \setminus V_{n-1}$ (where we take $V_0 = \varnothing$). These are all compact.

Fix $n \geq 1$. For each $p \in A_n$, there is some $W \in \mathscr{O}$ with $p \in U$. There is also some chart (U, φ) at p, and so $U \cap W$ is an open set containing p. Let $\mathbb{B}$ be a ball in $\varphi(U \cap W)$ small enough that $\varphi^{-1}(\mathbb{B}) \subseteq V_{n+1} \setminus V_{n-2}$ (this is possible by the continuity of φ); then $B_p = \varphi^{-1}(\mathbb{B})$ is a neighborhood of p that is diffeomorphic to a ball (via φ) and is contained in U and in $V_{n+1} \setminus V_{n-2}$.

(We call such neighborhoods *coordinate balls*.) Now, since A_n is compact, we may choose finitely many B_p of these that cover A_n. Take $\mathscr{O}'$ to be this countable collection of open sets for $n \geq 1$, which are all by construction diffeomorphic to balls (and since each element is contained in some V_{n+1}, it has compact closure). This gives an open cover of M subordinate to $\mathscr{O}$; we need only show it is locally-finite.

Fix any $p \in M$, and let $B \in \mathscr{O}'$ be a coordinate ball containing p, that was introduced at stage n above. Then $B \subseteq V_{n+1}$. Now, if $B' \in \mathscr{O}'$ was introduced at stage m, then it is not in V_{m-2}, so as long as $m - 2 > n + 1$, B does not intersect B'. Thus, the set of elements B' that *do* intersect the given neighborhood B of p are among those that were introduced at stages $1, 2, \ldots, n + 2$. There are only finitely many of these, and so $\mathscr{O}'$ is locally-finite, as desired. □

EXERCISE 2.21.1. *The above construction produced a locally-finite open cover $\mathscr{O}'$ that is countable. In fact, in any second-countable space, any locally-finite open cover is countable. Prove this.*

Another technical result that will be useful in what follows is the *shrinking lemma*: any locally-finite open cover can be shrunk a little bit (in the sense of shrinking each open set in the cover) and still remain an open cover.

LEMMA 2.22 (Shrinking Lemma). *Let $\mathscr{O}$ be a locally-finite open cover of M. Then each $U \in \mathscr{O}$ contains a subset U' with $\overline{U'} \subset U$ such that the collection of U' is also a locally-finite open cover.*

PROOF. As discussed above, the cover $\mathscr{O}$ is countable, so list it $\mathscr{O} = \{U_1, U_2, U_3, \ldots\}$. Consider the set

$$C_1 = U_1 \setminus \bigcup_{j \geq 2} U_j.$$

This set is closed: indeed, it is equal to $\overline{U_1} \setminus \bigcup_{j \geq 2} U_j$, for if $p \in \partial U_1$ then, since U_1 is open, $p \notin U_1$, and so $p \in \bigcup_{j \geq 2} U_j$. This shows $\overline{U_1} \setminus \bigcup_{j \geq 2} U_j \subseteq U_1 \setminus \bigcup_{j \geq 2} U_j$, and the reverse containment is immediate.

Thus C_1 is a closed set which is closed and contained in U_1, and satisfies $M = C_1 \cup U_2 \cup U_3 \cup \cdots$. Choose any open set U_1' with the property that $C_1 \subset U_1' \subset \overline{U_1'} \subset U_1$. Then $\{U_1', U_2, U_3, \ldots\}$ is an open cover of M. Now proceed by induction: having produced $U_1', \ldots, U_k'$, let

$$C_{k+1} = U_{k+1} \setminus (U_1' \cup \cdots \cup U_k' \cup U_{k+2} \cup U_{k+3} \cup \cdots).$$

By the same argument as above, C_{k+1} is closed, contains U_2, and $M = U_1' \cup \cdots \cup U_k' \cup C_{k+1} \cup U_{k+2} \cup U_{k+3} \cup \cdots$. Choose any open set U_{k+1}' which satisfies $C_{k+1} \subset U_{k+1}' \subset \overline{U_{k+1}'} \subset U_{k+1}$.

We need to show that $\{U_1', U_2', U_3', \ldots\}$ is an open cover; this is where local-finiteness comes in. Fix $p \in M$; then there is a finite collection of U_j containing p, and so let $n = \max\{j \colon p \in U_j\}$. By the above induction argument, $\{U_1', \ldots, U_n', U_{n+1}, U_{n+2}, \ldots\}$ is an open cover of M, and so $p \in U_1' \cup U_2' \cup \cdots U_n' \cup U_{n+1} \cup U_{n+2} \cup \cdots$. But since p is not in U_j for $j > n$, it follows that $p \in U_1' \cup \cdots \cup U_n'$. This holds for any p, and so $\{U_1', U_2', U_3', \ldots\}$ indeed form an open cover. (That it is locally-finite follows from the fact that it is a refinement of a locally-finite open cover.) □

This brings us to our local tool for building smooth functions: a (smooth) *partition of unity*.

DEFINITION 2.23. *Let $\mathscr{O}$ be an open cover of M. A* **partition of unity** *subordinate to $\mathscr{O}$ is a collection of functions $\{\psi_U \colon M \to \mathbb{R} \colon U \in \mathscr{O}\}$ with the following properties:*

- *supp* $\psi_U \subset U$

- $0 \leq \psi_U(p) \leq 1$ *for all* $p \in M$
- *The refinement* $\mathscr{O}'$ *of* $\mathscr{O}$*, given by* $U \in \mathscr{O}'$ *iff* ψ_U *is not identically* 0*, is locally-finite.*
- $\sum_{U \in \mathscr{O}} \psi_U(p) = 1$ *for each* $p \in M$*.*

Note that the last condition (that the functions sum to 1) makes sense in light of the other conditions: for any fixed $p \in M$, the set of $U \in \mathscr{O}'$ containing p is finite, and so the set of ψ_U for which $\psi_U(p) \neq 0$ is finite; hence the sum is always a finite sum.

The main theorem of this section is the assertion that every smooth manifold admits a *smooth* partition of unity subordinate to the open cover of its smooth structure. This is an enormously powerful theorem for building global smooth functions out of local ones. Indeed, let $\{(U_j, \varphi_j)\}_{j \in J}$ be the smooth structure of M. For each j, choose some smooth function $\hat{f}_j \colon \mathbb{R}^n \to \mathbb{R}$. Then $f_j = \hat{f}_j \circ \varphi_j$ is a smooth function on the smooth manifold U_j. We can extend it to a function on all of M by setting $f_j = 0$ on $M \setminus U_j$. This is not generally a smooth function, but that doesn't matter: we can define the global function $f = \sum_j \psi_j f_j$ for a smooth partition of unity subordinate to the open cover $\{U_j\}_{j \in J}$, and then f will be smooth (because the non-smooth parts of f_j are smashed to 0 by the compact support of ψ_j inside U_j).

We will see how to use this technique in many important examples in what follows. First, we need to see why a smooth partition of unity exists. The key is the existence of "smooth bump functions" on $\mathbb{R}^n$.

PROPOSITION 2.24. *Let* $0 < r < R < \infty$*. There exists a smooth function* $h \colon \mathbb{R}^n \to \mathbb{R}$ *with the following properties:*

(1) $h(x) = 1$ *for* $|x| \leq r$*.*
(2) $h(x) = 0$ *for* $|x| \geq R$*.*
(3) $0 < h(x) < 1$ *for* $r < |x| < R$*.*

PROOF. First, it suffices to prove the proposition in the case $n = 1$: if h_1 is such a function in the $n = 1$ case, then we can set $h(x) = h_1(|x|)$, which clearly possesses properties (1)–(3). The function $x \mapsto |x|$ is smooth on $\mathbb{R}^n \setminus \{0\}$, and so h (being a composition of a smooth function with $x \mapsto |x|$) is as well; it is also smooth at 0 since it is, by construction, $= 1$ on a neighborhood of 0.

To prove the existence of such an function $h_1 \colon \mathbb{R} \to \mathbb{R}$, we first build a smooth version of a cutoff function. Set $f(t) = e^{-1/t} \mathbb{1}_{t \geq 0}$. Then f is smooth on $\mathbb{R} \setminus \{0\}$. In fact, it is also smooth at 0. First, continuity at 0 follows by l'Hôpital's rule since $\lim_{t \to 0} e^{-1/t} = 0 = f(0)$. Continuing by induction, it is straightforward to verify that, for $t \neq 0$ and any positive integer k, $f^{(k)}(t) = p_k(t) e^{-1/t} / t^{2k}$ for some polynomial p_k. This means, by l'Hôpital's rule, that $\lim_{t \to 0} f^{(k)}(t) = 0$. On the other hand, we also have

$$f^{(k)}(0) = \lim_{t \to 0} \frac{f^{(k-1)}(t) - f^{(k-1)}(0)}{t} = \lim_{t \to 0} \frac{p_{k-1}(t) e^{-1/t} / t^{2(k-1)}}{t} = 0$$

again by induction. Thus, $f^{(k)}$ exists and is continuous, for each k, and so f is smooth, as desired.

Note that $f(t) = 0$ for $t \leq 0$, while $f(t) > 0$ for $t > 0$. It is easy to then verify that

$$h_1(t) = \frac{f(R - |t|)}{f(R - |t|) + f(|t| - r)}$$

has the desired properties, concluding the proof. □

COROLLARY 2.25. *Let* M *be a smooth manifold. Let* $K \subset U \subset M$ *with* K *compact and* U *open. Then there is a smooth function* $f \colon M \to [0, 1]$ *with* $f \equiv 1$ *on* K *and* $f \equiv 0$ *on* $M \setminus U$*.*

PROOF. For each $p \in K$, choose a chart (U_p, φ_p) at p such that $\overline{U_p} \subseteq U$, and $\varphi_p(p) = 0$. Let $R_p > 0$ be such that the ball $\mathbb{B}(0, 2R_p)$ is contained in $\varphi_p(U_p)$. Choose some $r_p \in (0, R_p)$, and let h_p be a smooth bump function as in Proposition 2.24 with radii r_p and R_p. Then $g_p = h_p \circ \varphi_p$ is a smooth function on U_p. We can extend it to all of M by setting $g_p(q) = 0$ for $q \in M \backslash \varphi_p^{-1}(\overline{\mathbb{B}}(0, R_p))$ (this agrees with the bump function on the preimage of the annulus $\mathbb{B}(0, 2R_p) \setminus \overline{\mathbb{B}}(0, R_p)$, and so the function is smooth by the smooth gluing lemma, Corollary 2.7). Hence, for each p we have a smooth function g_p which is 0 outside U_p, and hence outside U, and g_p is strictly positive on $U'_p = \varphi_p^{-1}(\mathbb{B}(0, R_p))$.

Now, as K is compact, we can choose finitely many $p_1, \ldots, p_k$ so that $K \subset U'_{p_1} \cup \cdots \cup U'_{p_k}$. Thus, the function $g = g_{p_1} + \cdots + g_{p_k}$ is a smooth function on M which is 0 outside U, and which is strictly positive on K. By the compactness of K, this means that there is some $\delta > 0$ with $g(p) \geq \delta$ for all $p \in K$. Let $k \colon \mathbb{R} \to \mathbb{R}$ be a smooth transition function on the interval $[0, \delta]$: $k(x) = 0$ for $x \leq 0$ and $k(x) = 1$ for $x \geq \delta$, while $0 \leq k(x) \leq 1$ for all x. For example, if h_1 is a bump function as in Proposition 2.24, with inner radius δ and outer radius 2δ, then we could take

$$k(x) = \begin{cases} h_1(\delta - x), & x \leq \delta \\ 1, & x \geq \delta \end{cases}.$$

Then the function $f = k \circ g$ has the desired properties. □

THEOREM 2.26. *Let M be a smooth manifold, and let $\mathscr{O}$ be an open cover. There is a smooth partition of unity subordinate to $\mathscr{O}$.*

PROOF. To begin, refine the open cover of smooth charts to a locally finite one $\mathscr{O}$ all of whose elements are coordinate balls with compact closure, cf. Lemma 2.21, and define $\psi_U \equiv 0$ for $U \notin \mathscr{O}$. Now, apply the shrinking lemma to produce a new locally-finite cover $\mathscr{O}'$ refining $\mathscr{O}$ so that for each $U \in \mathscr{O}$ there is a $U' \in \mathscr{O}'$ with $\overline{U'} \subset U$. By our construction, the open sets $U \in \mathscr{O}$ have compact closure, and therefore $\overline{U'}$ is compact. By Corollary 2.25, there is a smooth function $f_U \colon M \to [0, 1]$ such that $f_U \equiv 1$ on $\overline{U'}$ and supp $f_U \subset U$. Define $f = \sum_{U \in \mathscr{O}} f_U$. As the cover is locally-finite, this is a finite sum at each point. For each p, there is some neighborhood V so that only finitely many $U \in \mathscr{O}$ intersect V; thus, on V, f is this fixed finite sum of smooth functions, and so is smooth. Moreover, Since the U' cover M, it follows that $f > 0$ on M. Thus, we may define $\psi_U = f_U / f$. These are smooth functions, whose supports are subordinate to the locally-finite cover $\mathscr{O}$; they clearly take values in $[0, 1]$, and have been designed so that $\sum_{U \in \mathscr{O}} \psi_U = 1$. □

5. Applications of Partitions of Unity

PROPOSITION 2.27. *Let M be a smooth manifold, and let $A, B \subset M$ be disjoint closed sets. There exists a smooth function $f \colon M \to [0, 1]$ such that $f^{-1}(0) = A$ and $f^{-1}(1) = B$.*

PROOF. This is an exercise on Homework 2. □

We may talk about smooth functions on *any* subset of a smooth manifold.

DEFINITION 2.28. *Let M, N be smooth manifolds, and let $A \subseteq M$ be any subset. A map $F \colon A \to N$ is called smooth if, for each $p \in A$, there is an open neighborhood $U_p \subseteq M$ and a smooth map $F_p \colon U_p \to N$ so that $F|_{A \cap U_p} = F_p|_{A \cap U_p}$.*

One might expect the definition of smoothness on A to be the nominally stronger statement that F is the restriction of a smooth function on a neighborhood of all of A. As it turns out, these are equivlalent for closed sets A and functions taking values in $\mathbb{R}^k$.

PROPOSITION 2.29. *Let $A \subseteq M$ be closed, and let $f\colon A \to \mathbb{R}^k$ be a smooth function in the sense of Definition 2.28. Let U be any open neighborhood of A. Then there is a smooth function $\tilde{f}\colon M \to \mathbb{R}^k$ such that $\tilde{f}\Big|_A = f$ and supp $\tilde{f} \subseteq U$.*

PROOF. For each $p \in M$, choose a neighborhood U_p and a local smooth extension $\tilde{f}_p$ of f to U_p, as in Definition 2.28; wlog assump $U_p \subseteq U$ (i.e. replace U_p by $U_p \cap U$ if necessary). Then the collection $\{U_p\colon p \in M\} \cup \{M \setminus A\}$ is an open cover of M. Fix a smooth partition of unity $\{\psi_p\colon p \in M\} \cup \{\psi_0\}$ subordinate to this cover (so, in particular, $\operatorname{supp}\psi_p \subset U_p$ and $\operatorname{supp}\psi_0 \subset M \setminus A$).

For each $p \in A$, the function $\psi_p \tilde{f}_p$ is smooth on U_p and 0 outside the closed subset $\operatorname{supp}\psi_p \subset U_p$; hence, we can extend it to a smooth function on all of M by setting it equal to 0 outside $\operatorname{supp}\psi_p$. Then define

$$\tilde{f} = \sum_{p \in A} \psi_p \tilde{f}_p.$$

By local-finiteness, at each point $q \in M$ there is a neighborhood where this is a fixed finite sum, and thus defined a smooth function on that neighborhood; thus $\tilde{f} \in C^\infty(M, \mathbb{R}^k)$. Note that $\psi_0 = 0$ on A. Thus, by the partition of unity property, we have for $q \in A$,

$$1 = \sum_{p \in M} \psi_p(q) + \psi_0(q) = \sum_{p \in M} \psi_p(q),$$

and so since $\tilde{f}_p(q) = f(q)$ for $q \in A$, we have

$$\tilde{f}(q) = \sum_{p \in M} \psi_p(q) \tilde{f}_p(q) = \sum_{p \in M} \psi_p(q) f(q) = f(q).$$

This shows $\tilde{f}$ is a global smooth extension of f as desired. We need only show that $\operatorname{supp}\widetilde{f} \subseteq U$. This is left as an exercise. $\square$

REMARK 2.30. We may use a partition of unity here since the codomain of f is $\mathbb{R}^k$ where we can employ vector space operations. If we don't have this, not only does the above proof not work, but the result is generally false. For example: the identity map $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ is smooth, both in the intrinsic sense of a smooth map between smooth manifolds, and in the above sense of a map being smooth on a closed subset of $\mathbb{R}^2$; but this map has no smooth (or even continuous) extension to a map $\mathbb{R}^2 \to \mathbb{S}^1$. Indeed, suppose $f\colon \mathbb{R}^2 \to \mathbb{S}^1$ is C^1 and satisfies $f(u) = u$ for $u \in \mathbb{S}^1$. Then the curve $\gamma_r(t) = (r\cos t, r\sin t)$ in $\mathbb{R}^2$ has a C^1 image $f \circ \gamma_r$ in $\mathbb{S}^1$. We can compute the winding number $n(\gamma_r, 0)$ of this curve about 0 in the usual way:

$$n(\gamma_r, 0) = \oint_{f \circ \gamma_r} \frac{x}{\sqrt{x^2+y^2}} dy - \frac{y}{\sqrt{x^2+y^2}} dx.$$

By assumption $f \circ \gamma_r$ has image contained in $\mathbb{S}^1$, so this is the same as

$$\begin{aligned} n(\gamma_r, 0) &= \oint_{f\circ\gamma_r} x dy - y dx \\ &= \int_0^{2\pi} \left[f_1(r\cos t, r\sin t)\frac{d}{dt} f_2(r\cos t, r\sin t) - f_2(r\cos t, r\sin t)\frac{d}{dt} f_1(r\cos t, r\sin t)\right] dt \\ &= \int_0^{2\pi} f_1(r\cos t, r\sin t)[-\partial_1 f_2(r\cos t, r\sin t) r\sin t + \partial_2 f_2(r\cos t, r\sin t) r\cos t]\, dt \\ &\quad - \int_0^{2\pi} f_2(r\cos t, r\sin t)[-\partial_1 f_1(r\cos t, r\sin t) r\sin t + \partial_2 f_1(r\cos t, r\sin t) r\cos t]\, dt. \end{aligned}$$

Because f_j and $\partial_i f_j$ are continuous functions for $i, j \in \{1, 2\}$, this is evidently a continuous function of r. It is also integer valued (as a winding number), so it must be constant. By assumption, when $r = 1$, $n(\gamma_1, 0) = 1$, and so $n(\gamma_r, 0) = 1$. But, by inspection, $n(\gamma_0, 0) = 0$, a contradiction.

Another application of partitions of unity that demonstrates their typical application is the existence of a *smooth exhaustion function*. This is a smooth function $f\colon M \to \mathbb{R}$ with the property that $f^{-1}(-\infty, c]$ is compact for each $c \in \mathbb{R}$. Since f is defined everywhere, each point in M is in one (and hence all larger) of these *sublevel sets*. On $\mathbb{R}^n$, $f(x) = |x|^2$ is a smooth exhaustion function; on $\mathbb{B}^n$, $f(x) = \frac{1}{1-|x|^2}$ is one.

For any such function, the collection $K_n = f^{-1}(-\infty, n]$ forms a nested sequence of compact sets that covers M. So a smooth exhaustion function is a kind of smooth version of such a sequence. (Of course, this only really makes sense for a non-compact M.)

PROPOSITION 2.31. *Let M be a smooth manifold. Then there exists a strictly positive smooth exhaustion function $f\colon M \to (0, \infty)$.*

PROOF. Let $\{V_j\}$ be a locally-finite (thence countable) open cover of M, such that each V_j has compact closure. Let $\{\psi_j\}$ be a smooth partition of unity subordinate to this cover. Define $f = \sum_j j\psi_j$. For any $p \in M$, there is a neighborhood on which only finitely many ψ_j are non-zero, and so f is smooth on a neighborhood of p, thus $f \in C^\infty(M)$. Moreover, $f(p) \geq \sum_j \psi_j(p) = 1$, so f is strictly positive.

Now, fix $c \in \mathbb{R}$, and let $N > c$ be a positive integer. If $p \notin \bigcup_{j=1}^n \overline{V_j}$, since $\operatorname{supp} \psi_j \subset V_j$, it follows that $\psi_j(p) = 0$ for $1 \leq j \leq N$. Thus

$$f(p) = \sum_{j=N+1}^{\infty} j\psi_j(p) \geq N \sum_{j=N+1}^{\infty} \psi_j(p) = N \sum_{j=1}^{\infty} \psi_j(p) = N > c.$$

That is: if $p \notin \bigcup_{j=1}^n \overline{V_j}$ then $f(p) > c$. The contrapositive statement is: if $f(p) \leq c$, then $p \in \bigcup_{j=1}^N \overline{V_j}$. So the sublevel set $f^{-1}(-\infty, c]$, which is closed since f is continuous, is contained in the compact set $\bigcup_{j=1}^N \overline{V_j}$, and hence is compact. This shows f is a smooth exhaustion function. $\square$

CHAPTER 3

Tangent Vectors

1. Tangent Spaces

We typically view element in $\mathbb{R}^n$ with dual meaning: either as points in the metric space, or as vectors in the vector space. In particular, if $f\colon \mathbb{R}^n \to \mathbb{R}^m$, then the derivative *in a direction* $v \in \mathbb{R}^n$ at a *point* $x \in \mathbb{R}^n$ is $D_v f(x) = [Df(x)]v$. Here we view v as a vector in $\mathbb{R}^n$, which we must therefore draw as an arrow anchored at the origin; but geometrically we think of it as anchored at the point x. We could be more formal about this.

DEFINITION 3.1. *The* **tangent space** *to* $\mathbb{R}^n$ *as a point* $x \in \mathbb{R}^n$ *is the vector space* $\{x\} \times \mathbb{R}^n$. *We denote it by* $T_x\mathbb{R}^n$.

This notation helps us record that the vectors are tangent *to the point* x. This is helpful when we want to talk about vectors tangent to a surface, for example.

EXAMPLE 3.2. The smooth surface $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ has a tangent space at each point x, consisting of those vectors that are orthogonal to the position vector x itself. That is:

$$T_x\mathbb{S}^n = \{(x, v) \in T_x\mathbb{R}^{n+1} \colon v \perp x\} \subset T_x\mathbb{R}^n,$$

a subspace of the tangent space to $\mathbb{R}^{n+1}$ at x.

This notation allows us to conveniently talk about tangent spaces to smooth surfaces imbedded in $\mathbb{R}^n$. We would like to talk about tangent vectors to manifolds that are not (a priori) imbedded in $\mathbb{R}^n$. The presentation above of $T_x\mathbb{S}^n$ requires the imbedding: the tangent vectors live in the larger space $\mathbb{R}^{n+1}$ (even if they form an n-dimensional subspace thereof). We therefore need another way to talk about such tangent vectors that is *intrinsic* to the manifold itself.

There are many different (but, in the end, isomorphic) ways to do this. The one we will follow is the most common: having essentially defined the smooth structure on M by selecting which functions will be called smooth, we also define tangent vectors in terms of smooth functions. The key observation is that the space $T_x\mathbb{R}^n$ can be identified with the space of directional derivative differential operators $D_v|_x$ at the point x, for $v \in \mathbb{R}^n$. We will therefore *define* tangent vectors to be such operators. The question is: how can we intrinsically define such operators without already having an explicit $v \in \mathbb{R}^n$ to deal with? The answer lies in the product rule: for any directional derivative $D_v|_x$ at the point $x \in \mathbb{R}^n$ (acting on $C^\infty(\mathbb{R}^n)$), we have the product rule

$$D_v(fg)|_x = f(x)\, D_v g|_x + g(x)\, D_v f|_x\,.$$

It will turn out that this property uniquely defines all directional derivative operators at x. We call operators satisfying this property **derivations**.

DEFINITION 3.3. *Let* M *be a smooth manifold, and* $p \in M$. *The space of* **derivations** $\mathrm{Der}_p(M)$ *at* p *is the space of linear operators* $X_p\colon C^\infty(M) \to \mathbb{R}$ *with the property*

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

Let us record two key properties of derivations that follow immediately from the definition.

LEMMA 3.4. *Let M be a smooth manifold and $p \in M$, and let $X_p \in \mathrm{Der}_p M$. Then*

(a) *for any constant function $f \in C^\infty(M)$, $X_p f = 0$, and*
(b) *if $f, g \in C^\infty(M)$ with $f(p) = g(p) = 0$, then $X_p(fg) = 0$.*

PROOF. For part (a), note that if $f \equiv c$ then $f = c \cdot \mathbb{1}$ where $\mathbb{1}$ is the constant function taking value 1. Thus, by linearity $X_p f = X_p(c\mathbb{1}) = cX_p\mathbb{1}$, and so it suffices to prove that $X_p\mathbb{1} = 0$. This follows from the fact that $\mathbb{1} = \mathbb{1}^2$ and the derivation property:

$$X_p\mathbb{1} = X_p(\mathbb{1}^2) = \mathbb{1}(p)X_p\mathbb{1} + \mathbb{1}(p)X_p\mathbb{1} = 2X_p\mathbb{1}$$

and this implies that $X_p\mathbb{1} = 0$. Part (b) follows directly from the derivation property:

$$X_p(fg) = f(p)X_p g + g(p)X_p f = 0 + 0 = 0.$$

□

Elementary as these properties are, they actually allow us to prove that, on $\mathbb{R}^n$, $\mathrm{Der}_x\mathbb{R}^n$ consists exactly of the directional derivative operators $D_v|_x$ for $v \in \mathbb{R}^n$.

PROPOSITION 3.5. *Fix $p \in \mathbb{R}^n$. The map $\vartheta_p \colon (p, v) \mapsto D_v|_p$ is an vector space isomorphism $T_p\mathbb{R}^n \to \mathrm{Der}_p\mathbb{R}^n$.*

PROOF. We already noted that directional derivative operators $D_v|_p$ are derivations at p, and so the map is well-defined. It is straightforward to verify that it is a linear map. Now, let $v \in \ker \vartheta_p$, so that $D_v f(p) = 0$ for all $f \in C^\infty(\mathbb{R}^n)$. Taking $f(x) = x^j$ for $1 \le j \le n$, we have $0 = [Df(p)]v = e^j \cdot v = v^j$, the jth component of the vector in the standard basis. This shows that $v = 0$, and so ϑ_p is one-to-one.

It remains only to show that it is onto. Fix $X_p \in \mathrm{Der}_p\mathbb{R}^n$. Motivated by the previous argument, define the vector v by taking its jth component in the standard basis to be $X_p(x \mapsto x^j)$. We will show that $X_p = \vartheta_p(p, v) = D_v|_x$. By Taylor's Theorem 0.9, we have

$$f(x) = f(p) + \sum_{i=1}^n \partial_i f(p)(x^i - p^i) + \frac{1}{2}\sum_{i,j=1}^n (x^i - p^i)(x^j - p^j)\int_0^1 (1-t)\partial_i\partial_j f(p + t(x-p))\,dt.$$

From Proposition 0.11, the functions $x \mapsto \int_0^1 (1-t)\partial_i\partial_j f(p + t(x-p))\,dt$ are smooth. Thus, setting $g^i(x) = x^i - p^i$ and $h^j(p) = (x^j - p^j)\int_0^1 (1-t)\partial_i\partial_j f(p + t(x-p))\,dt$, the functions g^i, h^j are smooth and satisfy $g^i(p) = h^j(p) = 0$ for $1 \le i, j \le n$, and we have

$$f(x) = f(p) + \sum_{i=1}^n \partial_i f(p)(x^i - p^i) + \frac{1}{2}\sum_{i,j=1}^n g^i(x)h^j(x).$$

By Lemma 3.4, it follows that

$$X_p f = \sum_{i=1}^n X_p\left(\partial_i f(p)(x^i - p^i)\right) = \sum_{i=1}^n \partial_i f(p) v^i = D_v f|_p$$

as desired. □

We thus have an intrinsic realization of the geometric tangent space $T_p\mathbb{R}^n$: we can view it as the space of derivations $\mathrm{Der}_p\mathbb{R}^n$. So, for example, instead of talking about the vector $[1, 0, -2]^\top$ tangent to $p \in \mathbb{R}^3$, we could instead think of this vector as the first-order differential operator

$(\partial_1 - 2\partial_3)|_p$ acting on $C^\infty(\mathbb{R}^3)$. With this description in mind, we are prompted to define tangent spaces on manifolds accordingly.

DEFINITION 3.6. *Let M be a smooth manifold, and $p \in M$. The* **tangent space** *T_pM to M at p is defined to be the space Der_pM of derivations at p. That is, the tangent space is the set of linear operators $X_p\colon C^\infty(M) \to \mathbb{R}$ satisfying $X_p(fg) = f(p)X_pg + g(p)X_pf$ for all $f, g \in C^\infty(M)$.*

There is an important technical point to understand here: while the derivations formally act on smooth functions defined on all of M, they really only depend on the behavior of the functions in an arbitrarily small neighborhood of p.

LEMMA 3.7. *Let $X_p \in \mathrm{Der}_pM$, let $f, g \in C^\infty(M)$, and let U be any open neighborhood of p in M. If $f|_U = g|_U$, then $X_p(f) = X_p(g)$.*

PROOF. Let $h = f - g$, so $h \in C^\infty(M)$ satisfies $h = 0$ on U. The support $\operatorname{supp} h$ is a closed set disjoint from U. Now, let ψ be a bump function that is identically equal to 1 on $\operatorname{supp} h$, and satisfies $\psi(p) = 0$. Then ψh is identically equal to h: if $h(p) \neq 0$, then $(\psi h)(p) = 1 \cdot h(p) = h(p)$, and otherwise $(\psi h)(p) = 0 = h(p)$. But then, by Lemma 3.4, since $\psi(p) = h(p) = 0$, we have $X_p(h) = X_p(\psi h) = 0$. By linearity of X_p, this means $0 = X_p(h) = X_p(f - g) = X_p(f) - X_p(g)$, so $X_p(f) = X_p(g)$ as desired. □

By virtue of Proposition 3.5, when $M = \mathbb{R}^n$, Der_pM is an n-dimensional vector space. A priori, it may not be clear that this space is finite-dimensional (or non-zero) on a general manifold. We will shortly see that it has the same dimension at every point p, equal to $\dim M$. Indeed, this will follow from the fact that M looks like $\mathbb{R}^n$ in some neighborhood of p. To understand how this works, we first need to understand the generalized version of the total derivative.

2. The Differential / Tangent Map

For smooth functions $F\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, we have the total derivative $DF(p)\colon \mathbb{R}^n \to \mathbb{R}^m$, the best linear approximation of F near p. In fact, in our new language of anchoring vectors at base points, it is clear that we should think of $DF(p)$ as a linear map from $T_p\mathbb{R}^n$ to $T_{F(p)}\mathbb{R}^m$. That is, after all, how we draw the associated vectors. It also makes a great deal of sense in terms of the chain rule: if we have $F\colon \mathbb{R}^n \to \mathbb{R}^m$ and $G\colon \mathbb{R}^m \to \mathbb{R}^k$, then we have $D(G \circ F)(p)\colon T_p\mathbb{R}^n \to T_{G\circ F(p)}\mathbb{R}^k$ satisfies $D(G\circ F)(p) = DG(F(p)) \circ DF(p)$, where $DF(p)\colon T_p\mathbb{R}^n \to T_{F(p)}\mathbb{R}^m$, and so we see that the evaluation point $F(p)$ in DG is necessary for the composition to even make sense, so that $DG(F(p))\colon T_{F(p)}\mathbb{R}^n \to T_{G\circ F(p)}\mathbb{R}^k$.

Now, by Proposition 3.5, we may view $T_p\mathbb{R}^n$ as $\mathrm{Der}_p\mathbb{R}^n$ via the isomorphism ϑ_p. Thus, we can view $DF(p)\colon T_p\mathbb{R}^n \to T_{F(p)}\mathbb{R}^m$ as a linear map $\mathrm{Der}_p\mathbb{R}^n \to \mathrm{Der}_{F(p)}\mathbb{R}^m$; i.e. let us define

$$dF_p\colon \mathrm{Der}_p\mathbb{R}^n \to \mathrm{Der}_{F(p)}\mathbb{R}^m, \qquad dF_p = \vartheta_{F(p)} \circ DF(p) \circ \vartheta_p^{-1}. \tag{3.1}$$

This conjugated total derivative is called the **differential** of F at p. That is, for any $X_p \in \mathrm{Der}_p\mathbb{R}^n$, $dF_p(X_p)$ is a derivation in $\mathrm{Der}_p\mathbb{R}^m$. So

$$dF_p(X_p) = \vartheta_{F(p)}\left(DF(p)(\vartheta_p^{-1}(X_p))\right).$$

Let $\vartheta_p^{-1}(X_p) = (p, v)$. Then $DF(p)(\vartheta_p^{-1}(X_p)) = [DF(p)](p, v) = (F(p), [DF(p)]v) \in T_{F(p)}\mathbb{R}^m$. The action of $\vartheta_{F(p)}$ is to convert a vector into the directional derivative operator in the direction of that vector, and so

$$dF_p(X_p) = D_{[DF(p)]v}\big|_{F(p)}.$$

In other words, if $f \in C^\infty(\mathbb{R}^m)$, then $dF_p(X_p)$ is the derivation in $\mathrm{Der}_{F(p)}\mathbb{R}^m$ with action

$$dF_p(X_p)(f) = D_{[DF(p)]v}(f)\big|_{F(p)} = [Df(F(p))][DF(p)]v.$$

Now, employing the chain rule, this means that

$$dF_p(X_p)(f) = [D(f \circ F)(p)]v. \tag{3.2}$$

On the other hand, näively, given an element $X_p \in \mathrm{Der}_p M$, how could we make it act on $C^\infty(\mathbb{R}^m)$? Well, for any $f \in C^\infty(\mathbb{R}^m)$, the push-forward function $f \circ F$ is in $C^\infty(\mathbb{R}^n)$, which is the domain of the operator X_p. Since we have $X_p = \vartheta_p(p, v) = D_v\big|_p$, we have

$$X_p(f \circ F) = D_v(f \circ F)\big|_p = [D(f \circ F)(p)]v. \tag{3.3}$$

Comparing (3.2) and (3.3) leads to the following immediate, but deep, observation.

PROPOSITION 3.8. *Let $F\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be smooth. The differential map $dF_p\colon \mathrm{Der}_p\mathbb{R}^n \to \mathrm{Der}_{F(p)}\mathbb{R}^m$ of* (3.1) *is given by*

$$dF_p(X_p)(f) = X_p(f \circ F), \qquad X_p \in \mathrm{Der}_p\mathbb{R}^n,\ f \in C^\infty(\mathbb{R}^m).$$

In other words: by building the derivatives right into the definition of the vectors, the total derivative map becomes nothing more than the (pointwise) push-forward by F: the natural composition map needed to fit the pieces together (thanks to the chain rule).

We can thus discuss the differential of a smooth map between manifolds.

DEFINITION 3.9. *Let M, N be smooth manifolds, and let $F\colon M \to N$ be smooth. For $p \in M$, the* **differential** *dF_p is the map $T_pM \to T_{F(p)}N$ defined by*

$$dF_p(X_p)(f) = X_p(f \circ F), \qquad X_p \in T_pM,\ f \in C^\infty(N).$$

Note that, if $f, g \in C^\infty(N)$, then setting $Y_{F(p)} = dF_p(X_p)$ we have

$$\begin{aligned} Y_{F(p)}(f \cdot g) = X_p((f \cdot g) \circ F) &= X_p((f \circ F) \cdot (g \circ F)) \\ &= (f \circ F)(p)X_p(g \circ F) + (g \circ F)(p)X_p(f \circ F) \\ &= f(F(p))Y_{F(p)}(g) + g(F(p))Y_{F(p)}(f). \end{aligned}$$

That is: $Y_{F(p)}$ is indeed in $\mathrm{Der}_{F(p)}N$, and so dF_p is well-defined.

The collection of names/notations for this map is essentially uncountable. Some authors call it the *differential*, others the *total derivative*, still others the unimaginative *tangent map*. You might see it denoted as $dF_p = DF(p) = F'(p) = T_pF$, or even as $dF = DF = F' = TF$ with the p suppressed. Another popular notation is F_*, which highlights the fact that the map (in this language) is nothing more than the (pointwise) push-forward from Der_pM to $\mathrm{Der}_{F(p)}N$, via the map F. We will try to be consistent with the notation dF_p, and the name *differential*.

Here are some basic properties of differentials of smooth maps than can be readily verified from the definition.

PROPOSITION 3.10. *Let M, N, P be smooth manifolds, with $F\colon M \to N$ and $G\colon N \to P$ smooth maps. Let $p \in M$.*

(a) *$dF_p\colon T_pM \to T_{F(p)}N$ is a linear map.*
(b) *$d(G \circ F)_p = dG_{F(p)} \circ dF_p$.*
(c) *$d(\mathrm{Id}_M)_p = \mathrm{Id}_{T_pM}$.*
(d) *If F is a diffeomorphism, then dF_p is a linear isomorphism, and $(dF_p)^{-1} = d(F^{-1})_{F(p)}$.*

PROOF. We will prove (d), leaving the other parts to the reader. Let $q = F(p)$; then we have $dF_p \colon T_pM \to T_qN$ and $d(F^{-1})_q \colon T_qN \to T_pM$. We compose them: for $X_p \in T_pM$, $Y_p \in T_qN$, $f \in C^\infty(M)$, and $g \in C^\infty(N)$,

$$[d(F^{-1})_q \circ dF_p(X_p)](f) = [dF_p(X_p)](f \circ F^{-1}) = X_p((f \circ F^{-1}) \circ F) = X_p(f)$$
$$[dF_p \circ d(F^{-1})_q(Y_q)](g) = [d(F^{-1})_q(Y_q)](g \circ F) = Y_q((g \circ F) \circ F^{-1}) = Y_q(g).$$

Thus $d(F^{-1})_q \circ dF_p = \mathrm{Id}_{T_pM}$ and $dF_p \circ d(F^{-1})_q = \mathrm{Id}_{T_qN}$. □

From our (pre-derivation) definition of the tangent space $T_p\mathbb{R}^n$ to $\mathbb{R}^n$, it's clear that if we restrict to an open subset $U \ni p$, $T_pU = T_p\mathbb{R}^n$. In terms of derivations, this equality cannot be strictly true (since derivations in $\mathrm{Der}_p\mathbb{R}^n$ act on a different space than those in Der_pU). Nevertheless, the differential allows us to identify them in a natural way, on a general smooth manifold. After all, Lemma 3.7 shows that derivations act locally.

PROPOSITION 3.11. *Let M be a smooth manifold, and let $U \subseteq M$ be open. Denote by $\iota \colon U \hookrightarrow M$ the inclusion map (which is smooth). For any $p \in U$, the differential $d\iota_p \colon T_pU \to T_pM$ is a linear isomorphism.*

PROOF. Suppose $X_p \in T_pU$ is in $\ker(d\iota_p)$. Fix any $f \in C^\infty(U)$, and some open neighborhood B of p such that $\overline{B} \subset U$. By Proposition 2.29, there is a smooth extension $\tilde{f} \in C^\infty(M)$ so that $\tilde{f}|_{\overline{B}} = f|_{\overline{B}}$. Then f and $\tilde{f}|_U$ are smooth functions in $C^\infty(U)$ that agree on the open neighborhood B of p, and so by Lemma 3.7 $X_p(f) = X_p(\tilde{f}|_U)$. But $\tilde{f}|_U = \tilde{f} \circ \iota$, and so

$$X_p(f) = X_p(\tilde{f} \circ \iota) = d\iota_p(X_p)(\tilde{f}) = 0.$$

This shows that $X_p = 0$. Thus, $d\iota_p$ is injective.

For surjectivity, let $Y_p \in T_pM$. We define $X_p \in T_pU$ as follows: as above, fix some open neighborhood B of p with $\overline{B} \subset U$, and for $f \in C^\infty(U)$, define $X_p(f) = Y_p(\tilde{f})$ where $\tilde{f}$ is *any* smooth function on M which agrees with f on $\overline{B}$. This is well-defined by Lemma 3.7, and it is easy to verify that X_p is a derivation. We claim that $d\iota_p(X_p) = Y_p$. Indeed, for any $g \in C^\infty(M)$, we have

$$d\iota_p(X_p)(g) = X_p(g \circ \iota).$$

By definition, $X_p(g \circ \iota) = Y_p(h)$ for *any* smooth extension of $g \circ \iota|_{\overline{B}}$ to M. The function g is such an extension, and so $d\iota_p(X_p)(g) = Y_p(g)$, as desired. Thus, $d\iota_p$ is surjective, completing the proof. □

We therefore canonically identify $T_pU \cong T_pM$, keeping in mind that the derivations in T_pU act on functions in the larger space $C^\infty(M)$ by acting on any smooth extension from a function in a neighborhood of p in U.

REMARK 3.12. One way to avoid this slight complication is to define derivations a little differently: in light of Lemma 3.7, we may think of the domain of a derivation X_p not as the space $C^\infty(M)$, but as the space of equivalences classes of smooth functions that agree on any neighborhood of p. Such an equivalence class is called a *germ*, and the space of germs at p is usually denoted $C_p^\infty(M)$. Then we might define the tangent space to be the set of derivations of $C_p^\infty(M)$. This makes the local nature of the tangent space clearer, and it greatly simplifies the proof of Proposition 3.11 (since $C_p^\infty(M) = C_p^\infty(U)$). In particular, this means we don't need to use smooth extensions, whose existence depends on bump functions. If we were interested in *analytic manifolds*, we would have no choice but to use the germ definition, since analytic functions are too rigid

to allow the kinds of restriction/extension arguments above. But we will stick with the present formalism, since the topic is already highly abstract, and we don't want to complicate it more than necessary.

With this precise statement that the tangent space is local, combined with the local Euclideanness of the manifold, we can quickly prove that T_pM is an n-dimensional vector space at each $p \in M$ (where $n = \dim M$).

COROLLARY 3.13. *If M is an n-dimensional smooth manifold, then for each $p \in M$, T_pM is an n-dimensional vector space.*

PROOF. Fix $p \in M$, and let (U, φ) be a chart at p. Then $\varphi\colon U \to \hat{U}$ is a diffeomorphism onto an open set $\hat{U} \subseteq \mathbb{R}^n$. By Proposition 3.10(d), it follows that $d\varphi_p$ is an isomorphism $T_pU \to T_{\varphi(p)}\hat{U}$. By Proposition 3.11, we have $T_pU \cong T_pM$, and $T_{\varphi(p)}\hat{U} \cong T_{\varphi(p)}\mathbb{R}^n$. By Proposition 3.5, the latter is isomorphic to $\mathbb{R}^n$, concluding the proof. □

One more general useful identification is the tangent space to a product, which can be thought of as the direct sum of the tangent spaces.

PROPOSITION 3.14. *Let $M_1, \dots, M_k$ be smooth manifolds, and let $M = M_1 \times \cdots \times M_k$. Denote by $\pi_j\colon M \to M_j$ the projection map for $1 \le j \le k$. Fix a point $p = (p_1, \dots, p_k) \in M$. Then the map*

$$d(\pi_1)_p \oplus \cdots \oplus d(\pi_k)_p\colon T_pM \to T_{p_1}M_1 \oplus \cdots \oplus T_{p_k}M_k$$

is a linear isomorphism.

The proof is left as an exercise.

3. Local Coordinates

Tangent vectors are now viewed as abstract derivations. However, we can make these concrete in local coordinates. Fix $p \in M$, and let (U, φ) be a chart at p. Denote by $\hat{p} = \varphi(p)$ and $\hat{U} = \varphi(U)$. Let $\iota\colon U \hookrightarrow M$ and $\hat{\iota}\colon \hat{U} \hookrightarrow \mathbb{R}^n$ to be the inclusion maps. Then, following the proof of Corollary 3.13,

$$d\hat{\iota}_{\hat{p}} \circ d\varphi_p \circ (d\iota_p)^{-1}\colon T_pM \to T_{\hat{p}}\mathbb{R}^n$$

is a linear isomorphism. Note: we will usually ignore the maps $d\iota$ and $d\hat{\iota}$ and consider them to be the identity: that is, we use the isomorphism property of Proposition 3.11 to *identify* the tangent space of an open subset of a manifold with the tangent space of the manifold. Thus, we will (somewhat informally, but consistently) think of $d\varphi_p$ as an isomorphism from T_pM onto $T_{\hat{p}}\mathbb{R}^n$.

Now, $T_{\hat{p}}\mathbb{R}^n$ has a canonical basis: starting with the canonical basis $\{e^1, \dots, e^n\}$ of $\mathbb{R}^n$, the images under the isomorphism $\vartheta_{\hat{p}}$ are the derivations $D_{e^j}\big|_{\hat{p}} = \frac{\partial}{\partial x^j}\big|_{\hat{p}}$ for $1 \le j \le n$. Since $d\varphi_p$ is an isomorphism, the preimages of these derivations form a basis for T_pM. Here is another standard abuse of notation: we denote these preimages $\frac{\partial}{\partial x^j}\big|_p$:

$$\left.\frac{\partial}{\partial x^j}\right|_p \equiv (d\varphi_p)^{-1}\left(\left.\frac{\partial}{\partial x^j}\right|_{\hat{p}}\right) = d(\varphi^{-1})_{\hat{p}}\left(\left.\frac{\partial}{\partial x^j}\right|_{\hat{p}}\right).$$

So $\frac{\partial}{\partial x^j}\Big|_p$ is a derivation acting on $C^\infty(M)$ (or $C^\infty(U)$). How does it act? Untwisting the definition, given $f \in C^\infty(U)$, we write it in local coordinates as $\hat{f} = f \circ \varphi^{-1} \colon \hat{U} \to \mathbb{R}$. Then by the definition of the differential, we have

$$\frac{\partial}{\partial x^j}\bigg|_p (f) = d(\varphi^{-1})_{\hat{p}} \left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}} \right)(f) = \frac{\partial}{\partial x^j}\bigg|_{\hat{p}} (f \circ \varphi^{-1}) = \frac{\partial \hat{f}}{\partial x^j}(\hat{p}).$$

In other words: a canonical basis for $T_pU \cong T_pM$ is given by the n derivations whose actions are to take the partial derivatives of smooth functions in the chart coordinates. The vectors $\frac{\partial}{\partial x^1}\Big|_p, \dots, \frac{\partial}{\partial x^n}\Big|_p$ are called the **coordinate vectors at** p; it is important to note that they depend on the chart φ of choice.

EXAMPLE 3.15. Let $M = \mathbb{R}^3$, and consider the chart (U, φ) where $U = \{(x, y, z) \in \mathbb{R}^3 \colon x > 0\}$, and $\varphi \colon U \to (0, \infty) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \times (0, \pi)$ is the spherical polar map whose inverse is $\varphi^{-1}(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$. Fix a point $p = (x_0, y_0, z_0) \in U$, with spherical polar coordinates $\hat{p} = \varphi(p) = (\rho_0, \theta_0, \phi_0)$. The 3 coordinate vectors at this point are the derivations

$$\frac{\partial}{\partial \rho}\bigg|_p, \frac{\partial}{\partial \theta}\bigg|_p, \frac{\partial}{\partial \phi}\bigg|_p \in T_{\hat{p}}\hat{U}. \tag{3.4}$$

These are just the images, under the isomorphism $\vartheta_{\hat{p}}$, of the standard basis vectors e^1, e^2, e^3 in (ρ, θ, ϕ)-space. The more interesting question of expressing them as vectors in terms of the original (x, y, z)-coordinates is discussed below, in Example 3.18.

Thus, every vector $X_p \in T_pM$ can be expressed uniquely as a linear combination

$$X_p = \sum_{j=1}^n X_p^j \frac{\partial}{\partial x^j}\bigg|_p, \quad X_p^j \in \mathbb{R},\ 1 \le j \le n.$$

How do we compute the components X_p^j for a given X_p? Well, for any $f \in C^\infty(U)$,

$$X_p(f) = \sum_{j=1}^n X_p^j \frac{\partial}{\partial x^j}\bigg|_p (f) = \sum_{j=1}^n X_p^j \frac{\partial \hat{f}}{\partial x^j}(\hat{p}).$$

So, in particular, taking $\hat{f}(x) = x^j$ yields simply $X_p(f) = X_p^j$. Thus, we compute the components by evaluating X_p on the n smooth functions f for which $\hat{f} = f \circ \varphi^{-1}(x) = x^j$, $1 \le j \le n$. What are these functions? They are $f(p) = \varphi^j(p)$, where $\varphi = (\varphi^1, \dots, \varphi^n)$ (the n component functions of $\varphi \colon U \to \mathbb{R}^n$). So $X_p^j = X_p(\varphi^j)$; this is often (abusively) written as $X_p(x^j)$, writing $\varphi(p) = (x^1(p), \dots, x^n(p))$.

EXAMPLE 3.16. Following Example 3.15, consider the derivation $\frac{\partial}{\partial z}\Big|_p \in T_pU$ (where U is the $x > 0$ open half-plane). We want to express $\frac{\partial}{\partial z}\Big|_p$ as a linear combination of the coordinate vectors (3.4). To do this, we need expressions for the coordinates $(\varphi^1, \varphi^2, \varphi^3) = (\rho, \theta, \phi)$ (rather than the inverse $\varphi^{-1}(\rho, \theta, \phi) = (x, y, z)$). Once we have these, we then have

$$\frac{\partial}{\partial z}\bigg|_p = \frac{\partial \rho}{\partial z}\bigg|_{\hat{p}} \frac{\partial}{\partial \rho}\bigg|_p + \frac{\partial \theta}{\partial z}\bigg|_{\hat{p}} \frac{\partial}{\partial \theta}\bigg|_p + \frac{\partial \phi}{\partial z}\bigg|_{\hat{p}} \frac{\partial}{\partial \phi}\bigg|_p.$$

(For example, we know $\rho = \sqrt{x^2 + y^2 + z^2}$ so $\frac{\partial \rho}{\partial z} = \frac{z}{\rho} = \cos\phi$; so the coefficient of $\frac{\partial}{\partial \rho}\Big|_p$ is $\cos\phi_0$.) Of course, what we see is simply the chain rule again.

Now that we can express vectors in local coordinates, we can also write the differential in local coordinates. Let $F\colon M^m \to N^n$ be a smooth map between smooth manifolds. Let (U, φ) be a chart at p, and let (V, ψ) be a chart at $F(p)$. Then we have a coordinate representation of F:

$$\hat{F} = \psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \to \psi(V).$$

As above, let $\hat{p} = \varphi(p)$. We denote the components of φ as $(x^1, \ldots, x^m)$ and the components of ψ as $(y^1, \ldots, y^n)$.

PROPOSITION 3.17. *In terms of the coordinate bases* $\left\{ \frac{\partial}{\partial x^j}\Big|_p \right\}_{1 \le j \le m}$ *for* T_pM *and* $\left\{ \frac{\partial}{\partial y^j}\Big|_{F(p)} \right\}_{1 \le j \le n}$ *for* $T_{F(p)}N$*, we have*

$$dF_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right) = \sum_{k=1}^n \frac{\partial \hat{F}^k}{\partial x^j}(\hat{p})\, \frac{\partial}{\partial y^j}\bigg|_{F(p)}, \quad 1 \le j \le m.$$

In other words: in local coordinates, the matrix of dF_p is precisely the Jacobian matrix of the coordinate representation $\hat{F}$ of the function. The differential has been cooked up to be a coordinate-independent version of the Jacobian matrix.

PROOF. By definition,

$$\frac{\partial}{\partial x^j}\bigg|_p = d(\varphi^{-1})_{\hat{p}}\left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}}\right),$$

and so

$$dF_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right) = d(F \circ \varphi^{-1})_{\hat{p}}\left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}}\right).$$

Now, since $F \circ \varphi^{-1} = \psi^{-1} \circ \hat{F}$, we therefore have

$$d(F \circ \varphi^{-1})_{\hat{p}}\left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}}\right) = d(\psi^{-1} \circ \hat{F})_{\hat{p}}\left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}}\right) = d(\psi^{-1})_{\hat{F}(\hat{p})}\left(d\hat{F}_{\hat{p}}\left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}}\right)\right)$$

Now, the inside expression is the differential $d\hat{F}_{\hat{p}}$ of a map between Euclidean spaces, acting on a derivation on Euclidean space. By definition (3.1), this is the total derivative $D\hat{F}(\hat{p})$ acting on the vector corresponding (via $\vartheta_{\hat{p}}$) to $\frac{\partial}{\partial x^j}\Big|_{\hat{p}}$ (which is the unit basis vector e^j) expressed in terms of derivations in the coordinates $\frac{\partial}{\partial y^k}\Big|_{\hat{F}(\hat{p})}$. The matrix of the total derivative is just the Jacobian matrix, and so we have

$$d\hat{F}_{\hat{p}}\left(\frac{\partial}{\partial x^j}\bigg|_{\hat{p}}\right) = \vartheta_{\hat{F}(\hat{p})}([D\hat{F}(\hat{p})]e_j) = \sum_{k=1}^n \vartheta_{\hat{F}(\hat{p})}\left(\frac{\partial \hat{F}^k}{\partial x^j}(\hat{p})e_k\right) = \sum_{k=1}^n \frac{\partial \hat{F}^k}{\partial x^j}(\hat{p})\, \frac{\partial}{\partial y^k}\bigg|_{\hat{F}(\hat{p})}.$$

Using linearity and the fact (by definition) that

$$d(\psi^{-1})_{\hat{F}(\hat{p})}\left(\frac{\partial}{\partial y^k}\bigg|_{\hat{F}(\hat{p})}\right) = \frac{\partial}{\partial y^k}\bigg|_{F(p)}$$

proves the result. □

As a final exercise in our discussion of tangent vectors in local coordinates, let us consider a *change of coordinates*. That is: consider two charts (U, φ) and (V, ψ) at $p \in M$, and as usual denote the components of φ as $(x^1, \ldots, x^n)$ and the components of ψ as $(y^1, \ldots, y^n)$. Then the transition map $\psi \circ \varphi^{-1}$ is the map taking $(x^1, \ldots, x^n)$ to $(y^1, \ldots, y^n)$. This is a special case of the above discussion, where we take the identity function $F = \mathrm{Id}_M \colon M \to M$ and express it in these local coordinates: $\psi \circ \varphi^{-1} = \psi \circ \mathrm{Id}_M \circ \varphi^{-1} = \widehat{\mathrm{Id}_M}$. Hence, from Proposition 3.17, we have

$$\left.\frac{\partial}{\partial x^j}\right|_p = d(\mathrm{Id}_M)_p \left(\left.\frac{\partial}{\partial x^j}\right|_p \right) = \sum_{k=1}^n \frac{\partial(\widehat{\mathrm{Id}_M})^k}{\partial x^j}(\hat{p}) \left.\frac{\partial}{\partial y^k}\right|_{\mathrm{Id}_M(p)} = \sum_{k=1}^n \frac{\partial y^k}{\partial x^j}(\hat{p}) \left.\frac{\partial}{\partial y^k}\right|_p . \tag{3.5}$$

Again, this is just another statement of the chain rule.

EXAMPLE 3.18. Continuing Examples 3.15 and 3.16: our manifold is $M = \mathbb{R}^3$, and we have some base point p in the $x > 0$ half-plane. We have two sets of coordinates: $\varphi = (\rho, \theta, \phi)$ and $\psi = (x, y, z)$. From (3.5), we can express the coordinate vectors in the spherical coordinates in terms of the Euclidean ones, via

$$\begin{aligned}
\left.\frac{\partial}{\partial \rho}\right|_p &= \frac{\partial x}{\partial \rho}(\hat{p}) \left.\frac{\partial}{\partial x}\right|_p + \frac{\partial y}{\partial \rho}(\hat{p}) \left.\frac{\partial}{\partial y}\right|_p + \frac{\partial z}{\partial \rho}(\hat{p}) \left.\frac{\partial}{\partial z}\right|_p \\
\left.\frac{\partial}{\partial \theta}\right|_p &= \frac{\partial x}{\partial \theta}(\hat{p}) \left.\frac{\partial}{\partial x}\right|_p + \frac{\partial y}{\partial \theta}(\hat{p}) \left.\frac{\partial}{\partial y}\right|_p + \frac{\partial z}{\partial \theta}(\hat{p}) \left.\frac{\partial}{\partial z}\right|_p \\
\left.\frac{\partial}{\partial \phi}\right|_p &= \frac{\partial x}{\partial \phi}(\hat{p}) \left.\frac{\partial}{\partial x}\right|_p + \frac{\partial y}{\partial \phi}(\hat{p}) \left.\frac{\partial}{\partial y}\right|_p + \frac{\partial z}{\partial \phi}(\hat{p}) \left.\frac{\partial}{\partial z}\right|_p
\end{aligned}$$

where $\hat{p} = \varphi(p) = (\rho_0, \theta_0, \phi_p)$ is the base point's spherical coordinates. In this case, we have the explicit coordinate transformation $(x, y, z) = \psi \circ \varphi^{-1}(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$, and so we can explicitly write down the Jacobian yielding

$$\begin{aligned}
\left.\frac{\partial}{\partial \rho}\right|_p &= \cos\theta_0 \sin\phi_0 \left.\frac{\partial}{\partial x}\right|_p + \sin\theta_0 \sin\phi_0 \left.\frac{\partial}{\partial y}\right|_p + \cos\phi_0 \left.\frac{\partial}{\partial z}\right|_p \\
\left.\frac{\partial}{\partial \theta}\right|_p &= -\rho_0 \sin\theta_0 \sin\phi_0 \left.\frac{\partial}{\partial x}\right|_p + \rho_0 \cos\theta_0 \sin\phi_0 \left.\frac{\partial}{\partial y}\right|_p \\
\left.\frac{\partial}{\partial \phi}\right|_p &= \rho_0 \cos\theta_0 \cos\phi_0 \left.\frac{\partial}{\partial x}\right|_p + \rho_0 \sin\theta_0 \cos\phi_0 \left.\frac{\partial}{\partial y}\right|_p - \rho_0 \sin\phi_0 \left.\frac{\partial}{\partial z}\right|_p .
\end{aligned}$$

Note: by transforming these derivations first via $(d\psi^{-1})_{\psi(p)}$ into partial derivative operators in the (x, y, z) coordinates, and then into vectors in $T_{\psi(p)}\mathbb{R}^3 \cong \{\psi(p)\} \times \mathbb{R}^3$ via the isomorphism $\vartheta_{\psi(p)}$, we get the three vectors

$$e_{\rho_0} = \begin{bmatrix} \cos\theta_0 \sin\phi_0 \\ \sin\theta_0 \sin\phi_0 \\ \cos\phi_0 \end{bmatrix}, \quad e_{\theta_0} = \begin{bmatrix} -\rho_0 \sin\theta_0 \sin\phi_0 \\ \rho_0 \cos\theta_0 \sin\phi_0 \\ 0 \end{bmatrix}, \quad e_{\phi_0} = \begin{bmatrix} \rho_0 \cos\theta_0 \cos\phi_0 \\ \rho_0 \sin\theta_0 \cos\phi_0 \\ -\rho_0 \sin\phi_0 \end{bmatrix}.$$

These vectors are routinely used by physicists and engineers. They are the standard basis vectors in spherical coordinates (at the point p whose spherical coordinates are $(\rho_0, \theta_0, \phi_0)$).

4. Velocity Vectors of Curves

Let M be a smooth manifold. A **smooth curve** in M (perhaps more accurately a smooth *parametrized* curve) is a smooth map $\alpha\colon (a,b) \to M$ for some open interval $(a,b) \subseteq \mathbb{R}$. Fix $t_0 \in (a,b)$, and let (U,φ) be a local chart at $p = \alpha(t_0)$ in M; then we can express the curve in local coordinates as $\hat{\alpha}(t) = \varphi \circ \alpha(t)$. The velocity vector to the curve $\hat{\alpha}(t) = (\hat{\alpha}^1(t), \dots, \hat{\alpha}^n(t))$ is the vector in $\mathbb{R}^n$

$$\left(\frac{d\hat{\alpha}^1}{dt}(t_0), \dots, \frac{d\hat{\alpha}^n}{dt}(t_0)\right) = \sum_{j=1}^n \frac{d\hat{\alpha}^j}{dt}(t_0) e^j.$$

Realizing the tangent space $T_{\hat{\alpha}(t_0)}\hat{U}$ as $\mathrm{Der}_{\hat{\alpha}(t_0)}(\mathbb{R}^n)$, the velocity vector is

$$\sum_{j=1}^n \frac{d\hat{\alpha}^j}{dt}(t_0) \left.\frac{\partial}{\partial x^j}\right|_{\hat{\alpha}(t_0)} \tag{3.6}$$

whose action on $\hat{f} \in C^\infty(\mathbb{R}^n)$ is (by the chain rule)

$$\sum_{j=1}^n \frac{d\hat{\alpha}^j}{dt}(t_0) \left.\frac{\partial}{\partial x^j}\right|_{\hat{\alpha}(t_0)} (\hat{f}) = \left.\frac{d}{dt}(\hat{f} \circ \hat{\alpha})\right|_{t_0}.$$

Now, $\hat{\alpha} = \varphi \circ \alpha$, and if we take $f \in C^\infty(M)$, restricted to U, and let $\hat{f} = f \circ \varphi^{-1}$ as usual, then this means the velocity vector to the local coordinate representation $\hat{\alpha}$ of α at t_0 acts as a derivation on $C^\infty(M)$ by

$$\dot{\alpha}(t_0)(f) \equiv \left.\frac{d}{dt}(f \circ \alpha)\right|_{t_0}.$$

This is our definition of the velocity vector to the curve α at time t_0, which is manifestly coordinate independent. In fact, noting that $\alpha\colon (a,b) \to M$ is a smooth map between manifolds, we can interpret the velocity vector as

$$\dot{\alpha}(t_0)(f) \equiv \left.\frac{d}{dt}(f \circ \alpha)\right|_{t_0} = d\alpha_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right)(f).$$

Here we write $\left.\frac{d}{dt}\right|_{t_0}$ instead of $\left.\frac{\partial}{\partial t}\right|_{t_0}$, as is common when there is only one variable; it is the standard basis vector for $T_{t_0}(a,b)$. So, to summarize:

DEFINITION 3.19. *Let $(a,b) \subseteq \mathbb{R}$ be a nonempty open interval, let M be a smooth manifold, and let $\alpha\colon (a,b) \to M$ be a smooth curve. For any $t_0 \in (a,b)$, the* **velocity vector** *of α at time t_0 is defined to be the derivation*

$$\dot{\alpha}(t_0) \equiv d\alpha_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right) \in T_{\alpha(t_0)}M.$$

In local coordinates, it is given by the usual expression (3.6).

In fact, we can use velocity vectors of curves to naturally characterize all tangent vectors. First note that all tangent vectors are velocity vectors to (lots of) curves.

LEMMA 3.20. *Let M be a smooth manifold, $p \in M$, and $X_p \in T_pM$. There is some smooth curve $\alpha\colon (-\epsilon,\epsilon) \to M$ for some $\epsilon > 0$ with $\alpha(0) = p$ so that $\dot{\alpha}(0) = X_p$.*

PROOF. Let (U, φ) be a chart at p, and with $\varphi = (x^1, \ldots, x^n)$, write X_p in coordinates $X_p = \sum_{j=1}^n X_p^j \left.\frac{\partial}{\partial x^j}\right|_p$. Because $\hat{U} = \varphi(U)$ is open and $\hat{p} = \varphi(p) \in \hat{U}$, there is some $\epsilon > 0$ so that the straight curve $\hat{\alpha}(t) = t(X_p^1, \ldots, X_p^n)$ for $: |t| < \epsilon\}$ is contained in $\hat{U}$. Define $\alpha(t) = \varphi^{-1}(t(X_p^1, \ldots, X_p^n))$. The above computations (cf. e.velocity.local) show that $\dot{\alpha}(0) = X_p$, as desired. □

What's more: velocity vectors of curves transform naturally under smooth maps.

LEMMA 3.21. *Let M, N be smooth manifolds, and let $F\colon M \to N$ be a smooth map. If $\alpha\colon (a, b) \to M$ is a smooth curve in M, then $\beta = F \circ \alpha\colon (a, b) \to N$ is a smooth curve in N, and for any $t_0 \in (a, b)$,*

$$\dot{\beta}(t_0) = dF_{\alpha(t_0)}(\dot{\alpha}(t_0)).$$

PROOF. This is just the chain rule for differentials:

$$\dot{\beta}(t_0) = d\beta_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right) = d(F \circ \alpha)_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right) = dF_{\alpha(t_0)}\left(d\alpha_{t_0}\left(\left.\frac{d}{dt}\right|_{t_0}\right)\right) = dF_{\alpha(t_0)}(\dot{\alpha}(t_0)).$$

□

Lemmas 3.20 and 3.21 give a method for computing differentials that is often the most effective in practice.

COROLLARY 3.22. *Let M, N be smooth manifolds, and $F\colon M \to N$ a smooth map. Fix $p \in M$, and $X_p \in T_pM$. Then*

$$dF_p(X_p) = (F \circ \alpha)'(0)$$

for any smooth curve $\alpha\colon (-\epsilon, \epsilon) \to M$ such that $\alpha(0) = p$ and $\dot{\alpha}(0) = X_p$.

(Here we have used the "prime" notation instead of the "dot" notation, since it is hard to put a dot over top the composition. We will generally use these two interchangeably for velocity vectors of curves.) If F is presented in a form other than an explicit coordinate representation, Corollary 3.22 is often the best way to actually calculate the differential dF_p. We will use this frequently in what comes.

REMARK 3.23. Corollary 3.22 shows us yet another equivalent construction of the tangent space: we can define T_pM to consist of equivalence classes of smooth curves $\alpha\colon (-\epsilon, \epsilon) \to M$ with $\alpha(0) = p$. We want the equivalence relation to somehow say that all elements should have the same $\dot{\alpha}(0)$; not having defined the tangent space yet, we instead follow the above discussion and say that two curves α, β are equivalent if, for every smooth function $f\colon M \to \mathbb{R}$, the two *real*-valued curves $f \circ \alpha$ and $f \circ \beta$ have the same velocity vector at $t = 0$: i.e. $(f \circ \alpha)'(0) = (f \circ \beta)'(0)$. Thus, we think of tangent vectors based at p as "infinitesimal curves" through p. The preceding corollary then shows that there is a bijection from this construction to the derivations we've chosen to use. What's more, as Lemma 3.21 hints at, in this construction of T_pM, the differential dF_p is simply the map which sends a vector $[\alpha]$ (the equivalence class of some curve α) to $[F \circ \alpha]$; the chain rule then shows this is well-defined. This is a very nice geometric way to picture the tangent space; it has the significant disadvantage that there is no obvious vector space structure (indeed, the best way to see that T_pM is a vector space is to use the bijection to our tangent space to import it).

5. The Tangent Bundle

We now have a good understanding of the tangent space at a point, T_pM. We can then put all of them together to form the *tangent bundle*.

DEFINITION 3.24. *Let M be a smooth manifold. The* **tangent bundle** *TM is the disjoint union*

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

It carries a natural projection map $\pi\colon TM \to M$, defined by $\pi(X_p) = p$ for $X_p \in T_pM$.

For example, as we can identify $T_p\mathbb{R}^n \cong \{p\}\times\mathbb{R}^n$, we have $T\mathbb{R}^n \cong \bigsqcup_{p\in\mathbb{R}^n}\{p\}\times\mathbb{R}^n = \mathbb{R}^n\times\mathbb{R}^n$. In some sense, this should work for any manifold, since $T_pM \cong \mathbb{R}^n$ for every p. However, this is purely set-theoretic; there is more interesting structure we can put on TM, and it typically will not amount to just $M \times \mathbb{R}^n$. Indeed, TM can be (naturally) given the structure of a smooth manifold of dimension $2n$.

To make TM into a smooth manifold, we refer to Proposition 1.20: we can define the topology and smooth structure all in one shot. We need to define the charts. To that end, begin by selecting some covering set of charts (U, φ) of M. For each such U, let $\widetilde{U} = \pi^{-1}(U)$: the collection of all vectors X_p anchored at all points $p \in U$. Now, let $\varphi(p) = (x^1(p), \dots, x^n(p))$ for $p \in U$. Then any element of $\pi^{-1}(U)$ can be written uniquely in the form

$$X_p = \sum_{j=1}^n X_p^j \left.\frac{\partial}{\partial x^j}\right|_p.$$

We therefore define a coordinate chart $\widetilde{\varphi}\colon \widetilde{U} \to \mathbb{R}^n \times \mathbb{R}^n$ by

$$\widetilde{\varphi}\left(\sum_{j=1}^n X_p^j \left.\frac{\partial}{\partial x^j}\right|_p\right) = (x^1(p), \dots, x^n(p), X_p^1, \dots, X_p^j) = (\varphi(p), X_p^1, \dots, X_p^n). \tag{3.7}$$

Notice that the image $\widetilde{\varphi}(\widetilde{U}) = \varphi(U) \times \mathbb{R}^n$ is an open subset of $\mathbb{R}^n \times \mathbb{R}^n$, and the map is a bijection here – its inverse is given by

$$\widetilde{\varphi}^{-1}(x, v^1, \dots, v^n) = \sum_{j=1}^n v^j \left.\frac{\partial}{\partial x^j}\right|_{\varphi^{-1}(x)}, \quad x \in \varphi(U).$$

To apply Proposition 1.20, we need to verify the last three conditions for the charts $(\widetilde{U}, \widetilde{\varphi})$ for TM. First, by selecting a countable set of the (U, φ) that cover M, the corresponding countable collection of $(\widetilde{U}, \widetilde{\varphi})$ cover TM. Moreover, if X_p and Y_q are distinct elements of TM, then either $p = q$ and so X_p and $Y_q = Y_p$ both lie in a single chart $(\widetilde{U}, \widetilde{\varphi})$ (for any U containing p), or $p \neq q$ in which case (by the Hausdorff assumption on M) there are disjoint charts U, V on M with $p \in U$ and $q \in V$, in which case $\widetilde{U}$ and $\widetilde{V}$ are disjoint charts containing X_p and Y_q.

Thus, it remains only to show that the transition maps are smooth. Let (U, φ) and (V, ψ) be charts on M with $U \cap V \neq \varnothing$, and let $(\widetilde{U}, \widetilde{\varphi})$ and $(\widetilde{V}, \widetilde{\psi})$ be the corresponding charts on TM. Then $\widetilde{\varphi}(\widetilde{U} \cap \widetilde{V}) = \varphi(U \cap V) \times \mathbb{R}^n$ and $\widetilde{\psi}(\widetilde{U} \cap \widetilde{V}) = \psi(U \cap V) \times \mathbb{R}^n$ are open subsets of $\mathbb{R}^n \times \mathbb{R}^n$. We can compute the transition map

$$\widetilde{\psi} \circ \widetilde{\varphi}^{-1}\colon \varphi(U \cap V) \times \mathbb{R}^n \to \psi(U \cap V) \times \mathbb{R}^n$$

explicitly. Denote the components of φ as $x = (x^1, \dots, x^n)$ and the components of ψ as $y = (y^1, \dots, y^n)$. Then for any point $p \in U \cap V$, and any $v \in \mathbb{R}^n$,

$$\widetilde{\varphi}^{-1}(x, v) = \sum_{j=1}^{n} v^j \left.\frac{\partial}{\partial x^j}\right|_p = \sum_{j,k=1}^{n} v^j \frac{\partial y^k}{\partial x^j}(x) \left.\frac{\partial}{\partial y^k}\right|_p$$

from (3.5). Hence

$$\widetilde{\psi} \circ \widetilde{\varphi}^{-1}(x, v) = \left(y(p), \sum_{j=1}^{n} v^j \frac{\partial y^1}{\partial x^j}(x), \dots, \sum_{j=1}^{n} v^j \frac{\partial y^n}{\partial x^j}(x) \right).$$

This is evidently a smooth function of (x, v). Thus, we have satisfied all the conditions of Proposition 3.7, and so TM is a smooth manifold. What's more: in the chart $(\widetilde{U}, \widetilde{\varphi})$, we have $\hat{\pi} = \pi \circ \widetilde{\varphi}^{-1}$ has the action $\hat{\pi}(x, v) = x$, which is smooth.

Let us summarize this discussion as follows.

PROPOSITION 3.25. *Let M be a smooth manifold. The charts in* (3.7) *define a smooth atlas that makes TM into a smooth manifold, and the projection map $\pi\colon TM \to M$ is a smooth map.*

As noted, while TM is *locally* isomorphic to $M \times \mathbb{R}^n$, this is typically not the case globally. (We will see some examples a little later on on the course.) If $TM \cong M \times \mathbb{R}^n$, we say the tangent bundle is *trivial*. One case where the tangent bundle is trivial is when the manifold itself has global coordinates.

PROPOSITION 3.26. *Let M be a smooth manifold, and suppose there is a single chart covering M (i.e. there exists a diffeomorphism $\varphi\colon M \to \hat{U}$ for some open subset $\hat{U} \subseteq \mathbb{R}^n$). Then TM is diffeomorphic to $M \times \mathbb{R}^n$.*

PROOF. In the discussion preceding Proposition 3.25, we showed that for any chart (U, φ), the bundle chart $(\widetilde{U}, \widetilde{\varphi})$ is a diffeomorphism onto $\varphi(U) \times \mathbb{R}^n$. Since $\widetilde{M} = \pi^{-1}(M) = TM$, and since $\varphi(M) = \hat{U}$, this proves the proposition. □

REMARK 3.27. It can happen that TM is trivial even if there is no global chart for M (indeed, this is true for all *Lie groups*, which we will study in the following two quarters). But typically TM is non-trivial. For example, it is known (due to deep work of Adams) that $T\mathbb{S}^n$ is trivial iff $n \in \{1, 3, 7\}$. The tangent bundle $T\mathbb{S}^2$ is not diffeomorphic to $\mathbb{S}^2 \times \mathbb{R}^2$ (we will be able to prove this, at least in a restricted form, in a few more pages). What 4-manifold is it? Think of it this way, for general n: at each point $p \in \mathbb{S}^n$, we can identify $T_p\mathbb{S}^n \cong \mathbb{R}^n$. We can then further identify this with $\mathbb{S}^n \setminus \{-p\}$, via stereopgraphic projection from the point p. Gluing these together, then, we have

$$T\mathbb{S}^n \cong \mathbb{S}^n \times \mathbb{S}^n \setminus \{(p, -p)\colon p \in \mathbb{S}^n\}.$$

It is the product of two spheres with the anti-diagonal removed. When $n = 1$, this is a torus with one complete (twisted) circle "unzipped", which just gives the cylinder $\mathbb{S}^1 \times \mathbb{R}$ as expected. For $n = 2$, this unzipping does not work, and the bundle is not trivial.

Now, suppose $F\colon M \to N$ is a smooth map. We have studied its differential $dF_p\colon T_pM \to T_{F(p)}N$ pointwise. Now that we know how to glue the tangent spaces together to form a manifold, we can glue these maps together as well to form a smooth map between tangent bundles.

PROPOSITION 3.28. *Let M^m, N^n be smooth manifolds, and let $F\colon M \to N$ be a smooth map. Define a map $dF\colon TM \to TN$ as follows: for any $p \in M$ and any $X_p \in T_pM$, $dF(X_p) \equiv dF_p(X_p) \in T_{F(p)}N \subset TN$. Then dF is a smooth map (called the* global differential *of F).*

PROOF. Fix a chart (U, φ) in M, and look at the action of dF on the neighborhood $\widetilde{U} \subset TM$: in local coordinates, we have

$$\widetilde{\varphi} \circ dF \circ \widetilde{\varphi}^{-1}(x^1, \ldots, x^m, v^1, \ldots, v^m) = \left(F^1(x), \ldots, F^n(x), \sum_{j=1}^m \frac{\partial \hat{F}^1}{\partial x^j}(x) v^j, \ldots, \sum_{j=1}^m \frac{\partial \hat{F}^n}{\partial x^j}(x) v^j \right).$$

This is smooth (since F is smooth). □

We end this section by stating properties of the global differential, which follow immediately from the same properties of the pointwise differential (cf. Proposition 3.10).

PROPOSITION 3.29. *Let M, N, P be smooth manifolds, with $F\colon M \to N$ and $G\colon N \to P$ smooth maps.*

(a) $d(\mathrm{Id}_M) = \mathrm{Id}_{TM}$.
(b) $d(G \circ F) = dG \circ dF$.
(c) *If F is a diffeomorphism, then $dF\colon TM \to TN$ is a diffeomorphism, and $d(F^{-1}) = (dF)^{-1}$.*

CHAPTER 4

Vector Fields

1. Definitions and Examples

Let M be a smooth manifolds. A **vector field** on M is simply a choice of a tangent vector at each point. In other words, it is a map $X\colon M \to TM$, with the property that $X(p) \in T_pM$ for each $p \in M$. This can be written as follows: let $\pi\colon TM \to M$ be the projection $\pi(X_p) = p$ for any $X_p \in T_pM$. Then a vector field is a function $X\colon M \to TM$ with the property that

$$\pi \circ X = \mathrm{Id}_M.$$

That is: X is a right-inverse to the projection π. Such a map is generally called a **section**, so a vector field is a section of the tangent bundle. While $X\colon M \to TM$ is not valued in $\mathbb{R}^n$, we can still talk about it roughly in those terms since, for each p, the value $X(p) = X_p$ of X is in T_pM which is a vector space. In particular, we define as usual the **support** $\mathrm{supp}X$ of a vector field to be the closure of the set $\{p \in M\colon X_p \neq 0\}$.

We will largely be concerned with *smooth* vector fields (in the sense of being a smooth map between the two smooth manifolds M and TM). There is a simple characterization of smoothness here (in local coordinates).

LEMMA 4.1. *Let X be a vector field on a smooth manifold M; so X is a function $p \mapsto X_p \in T_pM$. Then X is smooth if and only if for any coordinate chart (U, φ) on M with $\varphi = (x^1, \dots, x^n)$, in the coordinate representation*

$$X_p = \sum_{j=1}^{n} X_p^j \left.\frac{\partial}{\partial x^j}\right|_p \tag{4.1}$$

the functions $p \mapsto X_p^j$ are smooth functions $U \to \mathbb{R}$.

PROOF. In the natural coordinate chart $(\widetilde{U}, \widetilde{\varphi})$ on TM, the representation of the vector field at any point $p \in U$ is given by

$$\widetilde{\varphi}\left(\sum_{j=1}^{n} X_p^j \left.\frac{\partial}{\partial x^j}\right|_p\right) = (x^1(p), \dots, x^n(p), X_p^1, \dots, X_p^n).$$

Thus, if we write X in local coordinates (U, φ) on M and $(\widetilde{U}, \widetilde{\varphi})$ on TM, this gives $\hat{X} = \widetilde{\varphi} \circ X \circ \varphi^{-1}$ is the map $\hat{U} = \varphi(U) \to \mathbb{R}^{2n}$ given by

$$\hat{X}(x^1, \dots, x^n) = (x^1, \dots, x^n, \hat{X}^1(x), \dots, \hat{X}^n(x))$$

where $\hat{X}^j(x) = X^j_{\varphi^{-1}(x)}$. By definition of smoothness, X_p^j is smooth iff $\hat{X}_p^j$ is smooth (for every chart), and the map $\hat{X}$ between open subsets of Euclidean space is smooth iff its components are smooth. Thus, X is smooth iff $\hat{X}$ is smooth iff $\hat{X}^j$ are smooth for $1 \leq j \leq n$ iff $p \mapsto X_p^j$ are smooth for $1 \leq j \leq n$, as claimed. $\square$

NOTATION 4.2. We denote the set of smooth sections of TM, i.e. smooth vector fields on M, as $\mathscr{X}(M)$. A vector field that is not in $\mathscr{X}(M)$ is often called a *rough vector field*.

EXAMPLE 4.3. Any vector field in $\mathscr{X}(\mathbb{R}^n)$ has a representation of the form $X = \sum_{j=1}^n X^j \frac{\partial}{\partial x^j}$, for functions $X^1, \dots, X^n \in C^\infty(\mathbb{R}^n)$. These are the kinds of vector fields discussed in vector calculus. With the identification $X \sim (X^1, \dots, X^n)$, one can think of such a vector field as a transformation $\mathbb{R}^n \to \mathbb{R}^n$. This is the *wrong picture* in general, since it means abandoning the structure of having the vector X_p anchored at the point p.

An important specific example is the **Euler field** $E \in \mathscr{X}(\mathbb{R}^n)$, given by

$$E_x = x^1 \left.\frac{\partial}{\partial x^1}\right|_x + \cdots + x^n \left.\frac{\partial}{\partial x^n}\right|_x .$$

It points radially outward, with magnitude equal to the distance of the point from the origin. Thought of as a transformation, it is just the identity map; this bears no relation to its behavior. We will come back to this example later through the chapter.

EXAMPLE 4.4. Realize $\mathbb{S}^1$ as the unit circle in $\mathbb{C}$: $\mathbb{S}^1 = \{u \in \mathbb{C} \colon |u| = 1\}$. Let $\exp \colon \mathbb{R} \to \mathbb{S}^1$ denote the map $\exp(\theta) = e^{i\theta}$, which is a smooth map. If $a < b \in \mathbb{R}$ and $b - a < 2\pi$, then $\exp$ is a diffeomorphism from (a, b) onto its image $\exp(a, b) \subset \mathbb{S}^1$. This gives us charts $(U, \theta) = (\exp(a, b), (\exp|_U)^{-1})$. In the associated chart coordinates $(\widetilde{U}, \widetilde{\theta})$ on $T\mathbb{S}^1$, any vector field has the form $\hat{X}(\theta) = f(\theta)\frac{d}{d\theta}$. Let us take $f \equiv 1$, so we have just the coordinate vector field $\frac{d}{d\theta}$ on U.

Now, let (c, d) be any other open subset of $\mathbb{R}$, and let $(V, \phi) = (\exp(c, d), (\exp|_V)^{-1})$. If $U \cap V \neq \varnothing$, then for any point $u \in U \cap V$ we have $\theta = \theta(u)$ and $\phi = \phi(u)$ defined so that $e^{i\theta} = \exp(\theta) = \exp(\phi) = e^{i\phi}$. It follows that $\phi = \theta + 2n\pi$ for some integer n (and since both θ and ϕ are smooth functions, n is constant on $U \cap V$). Hence $\frac{d}{d\theta}|_u = \frac{d}{d\phi}|_u$. This means we can define a global vector field this way: it is defined to be $\frac{d}{d\theta}|_u$ in any coordinate patch (U, θ) for U an open strict subset of $\mathbb{S}^1$ where θ is a right-inverse to $\exp$. It is clearly smooth in any such patch, and the above argument shows it is well defined. Even though there is no global coordinate chart to define it, we still commonly denote it $\frac{d}{d\theta}$.

Since a vector field is a smooth map $M \to TM$, all our tools for smooth maps readily apply. Additionally, since we can (almost) think of a vector field as a function (taking values in a vector space), some of those relevant tools also apply, such as Proposition 2.29.

PROPOSITION 4.5. *Let $A \subset M$ be a closed set, and let X be a smooth vector field on A (meaning that the map $X \colon A \to TM$ is smooth in the sense of Definition 2.28: for any $p \in A$, there is an open neighbhoorhood V and a smooth extension of $X|_{V \cap A}$ to U). If V is any open neighborhood of A in M, there is a smooth vector field $\widetilde{X} \in \mathscr{X}(M)$ such that $\widetilde{X}|_A = X$ and* $\operatorname{supp}\widetilde{X} \subseteq U$.

PROOF. This is an exercise on Homework 3. □

COROLLARY 4.6. *Let M be a smooth manifold and $p \in M$. Let $v \in T_pM$. There is a smooth vector field $X \in \mathscr{X}(M)$ such that $X_p = v$.*

PROOF. The set $\{p\}$ is closed. Let (U, φ) be any chart at p; then $v = \sum_{j=1}^n v^j \left.\frac{\partial}{\partial x^j}\right|_p$ for some coefficients v^j. The constant-coefficient vector field (given by the same sum) is smooth on U, and is equal to v at $\{p\}$. Hence, by Proposition 4.5, there is a global vector field $X \in \mathscr{X}(M)$ which agrees with X on $\{p\}$, i.e. $X(p) = v$ as required. □

Proposition 4.5 works here because we can use partitions of unity and "add and multiply in the codomain". That is, even though TM is not typically a vector space, it is still possible to use vector space operations there locally. Another consequence of this is that $\mathscr{X}(M)$, the space of smooth sections of TM, is a vector space – in fact, it is a $C^\infty(M)$-module.

PROPOSITION 4.7. *Let M be a smooth manifold. Given $X, Y \in \mathscr{X}(M)$, define $X + Y$ to be the vector field $(X+Y)_p = X(p) + Y(p)$; for $f \in C^\infty(M)$, define fX to be the vector field $(fX)_p = f(p)X_p$. These are smooth vector fields, and the operations so-defined make $\mathscr{X}(M)$ into a $C^\infty(M)$-module.*

PROOF. $X+Y$ and fX are defined to be vector fields (for each p the result is a vector vield at p because T_pM is a vector space, and they are defined to be based at p). Writing in local coordinates shows that they are smooth vector fields (because X, Y, and f are smooth). Verifying the module properties is a trivial exercise. □

So, for example, (4.1) could be thought of not only as a decomposition at each point, but as an equality between vector fields in $\mathscr{X}(U)$:

$$X = \sum_{j=1}^{n} X^j \frac{\partial}{\partial x^j}$$

where X^j is the smooth function $p \mapsto X^j_p$, and $\frac{\partial}{\partial x^j}$ is the smooth vector field $p \mapsto \frac{\partial}{\partial x^j}\Big|_p$.

2. Frames

Continuing the theme of vector space terminology applied to the tangent bundle, we have the following.

DEFINITION 4.8. *Let M be a smooth manifold. A collection $X_1, \ldots, X_k \in \mathscr{X}(M)$ of vector fields is called* **linearly independent** *if $\{X_1(p), \ldots, X_k(p)\}$ is a linearly independent set of vectors in T_pM for each $p \in M$. (This is, of course, only possible if $k \leq \dim M$.) If $\operatorname{span}\{X_1(p), \ldots, X_k(p)\} = T_pM$ for each p, we say that $X_1, \ldots, X_k$* **span the tangent bundle**. *A* **frame** *for M is a spanning set of linearly independent vector fields.*

EXAMPLE 4.9.
(1) The vector fields $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\}$ form a frame for $\mathbb{R}^n$. More generally, if (U, φ) is any chart on an n-manifold, the coordinate vector fields (with the same notation as above) form a frame for U.
(2) The vector field $\frac{d}{d\theta}$ on $\mathbb{S}^1$ is a frame. Here is a nice way to view it (acting on $\mathbb{S}^1 \subset \mathbb{C}$) that will be useful in the next example. Writing the vector field in the original framework (i.e. applying ϑ_u^{-1} at each point $u \in \mathbb{S}^1$), this vector field is simply $X_u = iu$ for $u \in \mathbb{S}^1$.
(3) The 3-sphere $\mathbb{S}^3$ is defined to be the set of unit-norm points in $\mathbb{R}^4$. There is a nice algebra structure on $\mathbb{R}^4$ called the *quaternions*: we realize $\mathbb{R}^4$ as $\operatorname{span}_{\mathbb{R}}\{1, i, j, k\}$ where i, j, k are linearly independent and satisfy $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$. Extending this product by linearity defines an associative, non-commutative division algebra, the quaternions $\mathbb{H}$ (which is an example of a Clifford algebra).

We define three vector fields X, Y, Z on $\mathbb{R}^4 \cong \mathbb{H}$ by $X(x) = ix$, $Y(x) = jx$, and $Z(x) = kx$ (where we suppress the "based at x" pair notation). These can be written

explicitly in terms of the standard $\mathbb{R}^4$ coordinates $x = (x^1, x^2, x^3, x^4)$ as

$$X(x) = \begin{bmatrix} -x^2 \\ x^1 \\ -x^4 \\ x^3 \end{bmatrix}, \quad Y(x) = \begin{bmatrix} -x^3 \\ x^4 \\ x^1 \\ -x^2 \end{bmatrix}, \quad Z(x) = \begin{bmatrix} -x^4 \\ -x^3 \\ x^2 \\ x^1 \end{bmatrix}.$$

These are linear functions of the coordinates, and so are smooth. One can readily check that the four vectors $\{u, X(u), Y(u), Z(u)\}$ form an orthonormal basis for $\mathbb{R}^4$ whenever $|u| = 1$. This shows that the vectors $X(u), Y(u), Z(u)$ are tangent to $\mathbb{S}^3$ at u, and are orthonormal, so linearly independent. Since $\dim T_u\mathbb{S}^3 = 3$, it follows that $X(u), Y(u), Z(u)$ span $T_u\mathbb{S}^3$. Thus, these three vector fields X, Y, Z form a frame for $\mathbb{S}^3$.

As part (1) above points out, it is always possible to find a *local* frame – i.e. a frame for an open neighborhood of any given point. But *global* frames are harder to come by. One might hope that something like the trick in part (3) of the example should work on any sphere, but this is decidedly false. While something like this works on $\mathbb{S}^7$, it doesn't work on any sphere $\mathbb{S}^n$ unless $n \in \{1, 3, 7\}$. For example, the *Hairy Ball Theorem* asserts that, if n is even, there does not exist a smooth (or even continuous) vector field on $\mathbb{S}^n$ that vanishes nowhere. (For an elementary proof of this fact, see the beautiful paper [**4**].)

A smooth manifold that admits a global (smooth) frame is called **parallelizable**. Parallelizable manifolds have trivial tangent bundles.

PROPOSITION 4.10. *Let M^n be parallelizable. Then there is a diffeomorphism $F\colon M \times \mathbb{R}^n \to TM$ such that $F(p, v) \in T_pM$ for every $p \in M$. In particular, TM is trivial.*

PROOF. Fix a frame $\{X_1, \dots, X_n\}$ for M, and define F by

$$F(p, (v^1, \dots, v^n)) = v^1 X_1(p) + \cdots v^n X_n(p).$$

Then $F(p, v) \in T_pM$ as claimed. Note that F is clearly a bijection: as $\{X_1, \dots, X_n\}$ is a frame, every vector $X_p \in T_pM$ can be written uniquely in the form $v^1 X_1(p) + \cdots + v^n X_n(p)$ for some $v = (v^1, \dots, v^n) \in \mathbb{R}^n$, and so $F^{-1}(X_p) = (p, v)$. We need only show that F and F^{-1} are smooth. This is an exercise in writing vector fields in local coordinates (and changing coordinates), and is left to the reader. □

The converse of Proposition 4.10 is also true in the following specific sense. We call TM a *trivial vector bundle* if there is a diffeomorphism $F\colon M \times \mathbb{R}^n \to TM$ with the properties

(1) $F(p, v) \in T_pM$ for all $p \in M$ and $v \in \mathbb{R}^n$, and
(2) for each $p \in M$, $F|_{\{p\}\times\mathbb{R}^n}$ is a vector space isomorphism $\mathbb{R}^n \to T_pM$.

Such a diffeomorphism is called a *trivilaization* of TM. Proposition 4.10 really shows that if M is parallelizable then TM is a trivial vector bundle, and the converse is also true: if F is a trivialization, then selecting any basis $\{e^1, \dots, e^n\}$ of $\mathbb{R}^n$, the vector fields $X_j(p) = F(p, e_j)$ are a frame for M.

So, in particular, the Hairy Ball Theorem shows that $T\mathbb{S}^n$ is not a trivial vector bundle whenever n is even, while Example 4.9 shows that $T\mathbb{S}^1$ and $T\mathbb{S}^3$ are trivial vector bundles. The question of whether it might still be true that $T\mathbb{S}^n \cong \mathbb{S}^n \times \mathbb{R}^n$ (via a diffeomorphism that does not preserve the bundle structure) is more delicate. The answer is still no for $n \notin \{1, 3, 7\}$, but we don't have the technology to prove it just now.

3. Derivations

A vector $X_p \in T_pM$ is a derivation at p. In particular, it is a certain kind of function $X_p\colon C^\infty(M) \to \mathbb{R}$. Denote this function space as $X_p \in \mathrm{Fun}(C^\infty(M), \mathbb{R})$. Thus, a (rough) vector field is a function X that takes values in $\mathrm{Fun}(C^\infty(M), \mathbb{R})$:

$$X \in \mathrm{Fun}\left(M, \mathrm{Fun}(C^\infty(M), \mathbb{R})\right).$$

For any sets A, B, C, there is a natural identification

$$\varsigma\colon \mathrm{Fun}\left(A, \mathrm{Fun}(B, C)\right) \to \mathrm{Fun}\left(B, \mathrm{Fun}(A, C)\right)$$

given by

$$\left(\varsigma(F)(b)\right)(a) = \left(F(a)\right)(b).$$

(It is immediate to check that this is a bijection; it will also preserve any reasonable extra structure on the spaces A, B, C, as we will see.) So, in our case, let X be a (rough) vector field on M. Then $\varsigma(X) \in \mathrm{Fun}\left(C^\infty(M), \mathrm{Fun}(M, \mathbb{R})\right)$: X can be viewed as a function which eats smooth functions $f \in C^\infty(M)$ and spits out real-valued functions on $\mathbb{R}$. The action is

$$\left(\varsigma(X)(f)\right)(p) = X_p(f).$$

A good notation for functions that allows us to omit the explicit reference to the variable p is to just represent it as an empty set of parentheses $(\,)$ when needed; so we have $\varsigma(X)(f) = X_{()}(f)$.

The defining properties of X_p are that it is linear and obeys the product rule at the point p. It then follows that, for $\alpha, \beta \in \mathbb{R}$ and $f, g \in C^\infty(M)$,

$$\varsigma(X)(\alpha f + \beta g) = X_{()}(\alpha f + \beta g) = \alpha X_{()}(f) + \beta X_{()}(g) = \alpha\varsigma(X)(f) + \alpha\varsigma(X)(g).$$

That is to say: both the domain $C^\infty(M)$ and codomain $\mathrm{Fun}(M, \mathbb{R})$ of $\varsigma(X)$ are linear spaces, and $\varsigma(X)$ is a linear map. As for the derivation property, we have

$$\varsigma(X)(fg)(p) = X_p(fg) = f(p)X_p(g) + g(p)X_p(f) = f(p)\left(\varsigma(X)(g)\right)(p) + g(p)\left(\varsigma(X)(f)\right)(p).$$

The expression on the right is a sum of products of functions $M \to \mathbb{R}$; so we have the identity

$$\varsigma(X)(fg) = f\varsigma(X)(g) + g\varsigma(X)(f).$$

DEFINITION 4.11. *A* **rough derivation** *on M is a linear function $D\colon C^\infty(M) \to \mathrm{Fun}(M, \mathbb{R})$ with the property that $D(fg) = fD(g) + gD(f)$ for all $f, g \in C^\infty(M)$. If D takes values in $C^\infty(M)$, we call it a* **smooth derivation***, or just a* **derivation***.*

Our discussion above thus shows that if X is any (possibly rough) vector field on M, then $\varsigma(X)$ is a rough derivation. From now on, we will abuse notation and drop the ς, and identify the vector field X as this rough derivation: the action of the vector field is $X(f)(p) = X_p(f)$ (where we now stop writing the second set of parentheses since we are identifying $X(f)(p) = X(p)(f)$).

Now, we know that each tangent vector X_p is locally-defined: the value $X_p(f)$ depends only on the germ of f at p. This gives us a restriction operation on vector fields that does not make sense in general for functions defined on $C^\infty(M)$. If $U \subseteq M$ is an open subset, then we have a vector field $X|_U$ which is a (rough) derivation $C^\infty(U) \to \mathrm{Fun}(U, \mathbb{R})$. Its action (cf. Proposition 3.11) is

$$X|_U\,(g)(p) = X_p(\tilde{g}) \tag{4.2}$$

where $\tilde{g}$ is any smooth function on M whose restriction to some neighborhood of p agrees with g there. (Here we witness the identification of T_pU with T_pM.) What this means is that, viewing X as a rough derivation, its domain is not really $C^\infty(M)$; it is the "sheaf of germs of f" (an object which formally encodes the notion of making two functions equivalent if they agree on a

neighborhood of each point, in a way that doesn't mean the two functions are equal). In particular, it follows that, for any $f \in C^\infty(M)$ and any open set $U \subseteq M$,

$$X|_U\,(f|_U) = X(f)|_U\,.$$

With this in hand, it becomes easy to characterize when a (possibly rough) vector field is actually smooth.

PROPOSITION 4.12. *Let $X\colon M \to TM$ be a possibly rough vector field, and identify it as a rough derivation as above. The following are equivalent.*

(a) *X is smooth.*
(b) *$X(f) \in C^\infty(M)$ for every $f \in C^\infty(M)$.*
(c) *For every open set $U \subseteq M$ and every $g \in C^\infty(U)$, the function $X|_U\,(g)$ is in $C^\infty(U)$.*

PROOF. We prove the chain of implications (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (a).

- (a) $\Longrightarrow$ (b): If X is a smooth vector field, then for any point $p \in M$, choose a chart (U, φ) and write $X_p = \sum_{j=1}^n X^j(p)\,\frac{\partial}{\partial x^j}\big|_p$. By Lemma 4.1, the coefficient functions X^j are smooth functions on U. Thus, interpreting X as a rough derivation, we have for $f \in C^\infty(M)$

$$X(f)(p) = X_p(f) = \sum_{j=1}^n X^j(p)\,\frac{\partial}{\partial x^j}\bigg|_p(f) = \sum_{j=1}^n X^j(p)\frac{\partial(f\circ\varphi^{-1})}{\partial x^j}(\varphi(p))$$

  which is a smooth function of $p \in U$. Thus $X(f)$ is smooth in a neighborhood of p, for any $p \in M$, proving that $X(f) \in C^\infty(M)$.
- (b) $\Longrightarrow$ (c): We assume X is a smooth derivation, so in particular $X(f) \in C^\infty(M)$ for all $f \in C^\infty(M)$. Now, let $U \subseteq M$ be open, and take any $g \in C^\infty(U)$. Fix any $p \in U$, and let $\tilde{g} \in C^\infty(M)$ be a function that agrees with g on some neighborhood p (such a smooth extension can be constructed with a bump function, cf. the proof of Proposition 3.11). Then by (4.2), we have

$$X|_U\,(g)(p) = X_p(\tilde{g}) = X(\tilde{g})(p)$$

  and since $X(\tilde{g})$ is smooth by assumption, this shows $X|_U\,(g)$ is smooth at p. This holds for every $p \in U$, so $X|_U\,(g) \in C^\infty(U)$.
- (c) $\Longrightarrow$ (a) Let $p \in M$, and let (U, φ) be a chart at p. By assumption, for any $g \in C^\infty(U)$, $X|_U\,(g)$ is smooth at p. In particular, denoting $\varphi(p) = (x^1(p), \ldots, x^n(p))$, the functions x^j are C^∞ on U, and so $X|_U\,(x^j)$ are smooth at p. But we can compute these functions from the coordinate representation of the vector field X:

$$X|_U\,(x^j)(p) = X_p(x^j) = \sum_{k=1}^n X_p^k\,\frac{\partial}{\partial x^k}\bigg|_p(x^j) = X^j(p).$$

  Hence, the conclusion is that the component functions X^j are smooth at p for each $p \in U$. By Lemma 4.1, it follows that X is a smooth vector field.

□

Thus, if $X \in \mathscr{X}(M)$ is a smooth vector field, viewed as a rough derivation $X\colon C^\infty(M) \to \mathrm{Fun}(M, \mathbb{R})$, it is actually a derivation $C^\infty(M) \to C^\infty(M)$. This turns out to be an equivalence. For the precise statement, it is useful to return to the ς notation one more time.

THEOREM 4.13. *Let M be a smooth manifold. A function $D\colon C^\infty(M) \to C^\infty(M)$ is a derivation if and only if $D = \varsigma(X)$ for some $X \in \mathscr{X}(M)$.*

PROOF. Proposition 4.12 shows that if $X \in \mathscr{X}(M)$ then $\varsigma(X)$ is a smooth derivation, leaving us only to prove the converse. Thus, given a smooth derivation $D\colon C^\infty(M) \to C^\infty(M)$, define a vector field $X\colon p \mapsto X_p$ the only way possible: $X_p(f) = (D(f))(p)$ (which is precisely to say that $\varsigma(X) = D$). It is immediate that, since D is a derivation, $f \mapsto (D(f))(p)$ is a derivation at p, so $X_p \in T_pM$, defining a rough vector field. Furthermore, since $X_p(f) = (D(f))(p)$ is (by assumption) a smooth function of p, it follows from Proposition 4.12 that X is a smooth vector field, concluding the proof. □

Henceforth, we will identify $\mathscr{X}(M)$ with the space of smooth derivations on M (sometimes denoted $\mathrm{Der}(M)$).

4. Push-Forwards

Let M, N be smooth manifolds, with a smooth map $F\colon M \to N$. If $p \in M$ and $v \in T_pM$, then we transport v to a vector $w = dF_p(v) \in T_{F(p)}N$. We might then expect that we can do the same to push forward a vector field $X \in \mathscr{X}(M)$ to a vector field on N. The näive approach doesn't quite work, however. The vector $dF_p(X(p))$ it tangent to $F(p) \in N$, so the only possible choice would be to define the pushed-forward vector field as $Y(q) = dF_p(X(p))$ where $q = F(p)$. But this is problematic for two reasons: (1) if F is not onto, then there are some points $q \in N$ that are not of the form $q = F(p)$, and so dF gives us no way to assign tangent vectors to those points. And (2) if F is not one-to-one, there are some distinct points $p_1 \neq p_2 \in M$ for which $q = F(p_1) = F(p_2)$. But in this case, there's no reason to expect that $dF_{p_1}(X(p_1)) = dF_{p_2}(X(p_2))$, and so this "push-forward" will, in general, not define a unique tangent vector at q.

This, the relation $Y(F(p)) = dF_p(X(p))$ does not generally define a vector field Y from X. But this relation may still hold between two vector fields. When it does, we say X and Y are **F-related**. Since we now know we can view a smooth vector field as a derivation, let us translate the notion of F-relation in that language.

PROPOSITION 4.14. *Let $F\colon M \to N$ be a smooth map, $X \in \mathscr{X}(M)$, and $Y \in \mathscr{X}(N)$. Then X and Y are F-related iff*

$$X(f \circ F) = Y(f) \circ F, \qquad \forall\, f \in C^\infty(N).$$

PROOF. Simply compute that, for any $p \in M$,

$$X(f \circ F)(p) = X_p(f \circ F) = dF_p(X_p)(f),$$

while

$$Y(f) \circ F(p) = Y(f)(F(p)) = Y_{F(p)}(f).$$

□

EXAMPLE 4.15. Let X be the (global) coordinate vector field on $\mathbb{R}$ (day $X = \frac{d}{dt}$ with global coordinate t on $\mathbb{R}$). Let $F\colon \mathbb{R} \to \mathbb{R}^2$ be the smooth map $F(t) = (\cos t, \sin t)$. Note that the image of F is just the unit circle $\mathbb{S}^1$, which amply shows how there can be no (unique) way to transport X to all of $\mathbb{R}^2$ via F. However, there are plenty of F-related vector fields Y to X. The most canonical one is

$$Y = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x},$$

the curl field. (Note that $Y|_{\mathbb{S}^1}$ is the vector field $\frac{d}{d\theta}$ of Example 4.4.) To see this, we just compute: for any $f \in C^\infty(\mathbb{R}^2)$,

$$X(f\circ F)(t) = \frac{d}{dt}(f(\cos t, \sin t)) = \frac{\partial f}{\partial x}(\cos t, \sin t)\cdot(-\sin t) + \frac{\partial f}{\partial y}(\cos t, \sin t)\cdot(\cos t) = (Yf)(F(t)).$$

So X and Y are F-related. Note, however, that since F can only "see" what happens to Y on $\mathbb{S}^1$, if we take *any* smooth vector Y' field on $\mathbb{R}^2$ that agrees with Y on $\mathbb{S}^1$, then X and Y' will also be F-related.

The one case where there is a unique F-related vector field to a given $X \in \mathscr{X}(M)$ is when F is a diffeomorphism.

PROPOSITION 4.16. *Let M, N be smooth manifolds and let $F\colon M \to N$ be a diffeomorphism. Let $X \in \mathscr{X}(M)$. Define $F_*(X) \in \mathscr{X}(N)$ by*

$$F_*(X)(q) = dF_{F^{-1}(q)}(X_{F^{-1}(q)}). \tag{4.3}$$

Then $F_*(X)$ is the unique smooth vector field on N that is F-related to X.

The vector field $F_*(X)$ is called the **push-forward** of X by F. One might think of Y as "a push forward of X by F" whenever F is a smooth map and X, Y are F-related, but this is not uniquely defined unless F is a diffeomorphism.

PROOF. Let $Y_q = F_*(X)(q)$ be the (a priori rough) vector field defined by (4.3), and let $p = F^{-1}(q)$. Then (4.3) the statement $Y_q = dF_p(X_p)$, which is precisely the statement that X and Y are F-related. This shows that Y is the unique (rough) vector field that is F-related to X. It remains only to show that Y is in fact smooth. But by definition it is the composition

$$N \xrightarrow{F^{-1}} M \xrightarrow{X} TM \xrightarrow{dF} TN$$

which is a composition of three smooth maps, and is hence smooth. □

COROLLARY 4.17. *Let $F\colon M \to N$ be a diffeomorphism, and let $X \in \mathscr{X}(M)$. Then for $f \in C^\infty(N)$,*

$$(F_*(X)(f)) \circ F = X(f \circ F).$$

Push forwards, and more generally F-relatedness, will be how we measure "naturality" of transformations of vector fields. Proposition 4.16 shows the unique way to push a vector field along a diffeomorphism, which shows that $\mathscr{X}(M)$ and $\mathscr{X}(N)$ are naturally isomorphic whenever M and N are diffeomorphic. More generally, if we have a smooth map $F\colon M \to N$ between manifolds, we will generally understand the behavior of a vector field on M to be "the same" as a vector field on N if they are F-related.

5. Lie Brackets

We now think of vector fields in $\mathscr{X}(M)$ as derivations: linear operators on $C^\infty(M)$ that satisfy the product rule. Locally, we know such operators can be expressed as first-order differential operators. Thus, the composition of two vector fields will, locally, be a second-order differential operator. Such operators are not derivations. Indeed, let us compute in general for two derivations $X, Y \in \mathrm{Der}(M)$: for $f \in C^\infty(M)$,

$$\begin{aligned} XY(fg) = X(f\cdot Y(g) + g\cdot Y(f)) &= [f\cdot XY(g) + Y(g)X(f)] + [g\cdot XY(f) + Y(f)X(g)] \\ &= f\cdot XY(g) + g\cdot XY(f) + X(f)Y(g) + X(g)Y(f). \end{aligned}$$

So, we see that the defect of XY from being a derivation is

$$XY(fg) - [f \cdot XY(g) + g \cdot XY(f)] = X(f)Y(g) + X(g)Y(f).$$

This is typically not 0; it happens only when (viewed as maps $M \to TM$) the supports $\operatorname{supp} X$ and $\operatorname{supp} Y$ are disjoint.

So we cannot usually compose derivations to get a derivation. But we can take their **Lie bracket**:

$$[X, Y] \equiv XY - YX. \tag{4.4}$$

LEMMA 4.18. *Let $X, Y \in \mathscr{X}(M)$. Then $[X, Y] \in \mathscr{X}(M)$.*

PROOF. For any linear operator $A\colon C^\infty(M) \to C^\infty(M)$, let $\Gamma_A\colon C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ measure the defect of A from being a derivation,

$$\Gamma_A(f, g) = A(fg) - [f \cdot A(g) + g \cdot A(f)].$$

Then $A \in \mathscr{X}(M)$ if and only if $\mathfrak{D}_A \equiv 0$. As computed above, we have

$$\Gamma_{XY}(f, g) = X(f)Y(g) + X(g)Y(f). \tag{4.5}$$

Now, $A \mapsto \Gamma_A(f, g)$ is linear, and so we have

$$\Gamma_{[X,Y]} = \mathfrak{D}_{XY} - \mathfrak{D}_{YX}.$$

Subbing in (4.5) yields, for all $f, g \in C^\infty(M)$,

$$\Gamma_{[X,Y]}(f, g) = [X(f)Y(g) + X(g)Y(f)] - [Y(f)X(g) + Y(g)X(f)] = 0.$$

□

EXAMPLE 4.19. Consider the following two vector fields on $\mathbb{R}^2$:

$$X = \frac{\partial}{\partial x}, \quad Y = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}.$$

We will call these two vector fields $X =$"drive east" and $Y =$"steer". Let's compute their bracket: for $f \in C^\infty(\mathbb{R}^2)$,

$$\begin{aligned}
[X, Y](f) &= \frac{\partial}{\partial x}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) f - \left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\frac{\partial}{\partial x} f \\
&= \frac{\partial}{\partial x}(xf_y - yf_x) - (xf_{yx} - yf_{xx}) \\
&= f_y + xf_{xy} - yf_{xx} - (xf_{yx} - yf_{xx}) = \frac{\partial}{\partial y} f.
\end{aligned}$$

That is: the bracket of "drive east" and "steer" is "drive north". This is why parallel parking works! More in this later.

The Lie bracket of vector fields is a genuinely *new* operation: it is not just a new coordinate-free version of some vector operation from vector calculus. Let us compute it in local coordinates.

PROPOSITION 4.20. *Let M be a smooth manifold, let $X, Y \in \mathscr{X}(M)$ be vector fields, and let (U, φ) be a coordinate chart with $\varphi = (x^1, \ldots, x^n)$. Express the vector fields in local coordinates (as usual)*

$$X|_U = \sum_{j=1}^n X^j \frac{\partial}{\partial x^j}, \qquad Y|_U = \sum_{j=1}^n Y^j \frac{\partial}{\partial x^j}.$$

Then, in local coordinates,

$$[X,Y]|_U = \sum_{j,k=1}^{n} \left(X^j \frac{\partial Y^k}{\partial x^j} - Y^j \frac{\partial X^k}{\partial x^j} \right) \frac{\partial}{\partial x^k}.$$

PROOF. This is an elementary (if tedious) computation. Fix $f \in C^\infty(M)$. To save space, denote $\frac{\partial}{\partial x^k} f = f_k$, and $\frac{\partial}{\partial x^j}\frac{\partial}{\partial x^k} f = f_{kj}$. Since f is smooth, $f_{jk} = f_{kj}$ as usual. Then (being clever about naming indices) we have

$$\begin{aligned}
[X,Y]|_U f &= \left(\sum_{j=1}^{n} X^j \frac{\partial}{\partial x^j} \right) \left(\sum_{k=1}^{n} Y^k \frac{\partial}{\partial x^k} \right) f - \left(\sum_{j=1}^{n} Y^j \frac{\partial}{\partial x^j} \right) \left(\sum_{k=1}^{n} X^k \frac{\partial}{\partial x^k} \right) f \\
&= \sum_{j,k=1}^{n} \left(X^j \frac{\partial}{\partial x^j}(Y^k f_k) - Y^j \frac{\partial}{\partial x^j}(X^k f_k) \right).
\end{aligned}$$

The term inside the brackets expands (using the product rule) to

$$X^j \frac{\partial}{\partial x^j}(Y^k f_k) - Y^j \frac{\partial}{\partial x^j}(X^k f_k) = X^j \frac{\partial Y^k}{\partial x^j} f_k + X^j Y^k f_{kj} - Y^j \frac{\partial X^k}{\partial x^j} f_k - Y^j X^k f_{kj}.$$

Grouping the first- and second-order terms into two sums, this gives

$$[X,Y]|_U f = \sum_{j,k=1}^{n} \left(X^j \frac{\partial Y^k}{\partial x^j} f_k - Y^j \frac{\partial X^k}{\partial x^j} f_k \right) + \sum_{j,k=1}^{n} (X^j Y^k f_{kj} - Y^j X^k f_{kj}).$$

Finally, break up the second sum and reverse the dummy indices:

$$\sum_{j,k=1}^{n} X^j Y^k f_{kj} - \sum_{j,k=1}^{n} Y^j X^k f_{kj} = \sum_{j,k=1}^{n} X^j Y^k f_{kj} - \sum_{j,k=1}^{n} Y^k X^j f_{jk} = 0,$$

where the final equality comes from the identity $f_{kj} = f_{jk}$. □

In particular, if all the coefficients X^j and Y^j are constant, then $[X,Y]|_U = 0$. (A special case of this is when $X = \frac{\partial}{\partial x^j}$ and $Y = \frac{\partial}{\partial x^k}$ for some fixed j,k; then the fact that $[X,Y] = 0$ is just the statement that $f_{jk} = f_{kj}$ for smooth f, which we used in the proof.) Whenever it happens that $[X,Y] = 0$, we say that the vector fields X and Y **commute**. Thus, coordinate vector fields (and in general constant coefficient vector fields) commute.

What does it mean for vector fields to commute? More generally: what does $[X,Y]$ really *mean*? Proposition 4.20 shows that it is a new vector field with components built out of the components of X and Y together with their derivatives. In a sense, the goal of the next chapter is to understand what this really means.

First, let us summarize some algebraic properties of the Lie bracket that will be very useful for computations.

PROPOSITION 4.21. *Let M be a smooth manifold, and let $X, Y, Z \in \mathscr{X}(M)$. The Lie bracket satisfies the following properties.*

(a) BILINEARITY*: for $a, b \in \mathbb{R}$,*

$$\begin{aligned}
[aX + bY, Z] &= a[X,Z] + b[Y,Z], \\
[Z, aX + bY] &= a[Z,X] + b[Z,Y].
\end{aligned}$$

(b) ANTISYMMETRY*:*

$$[X,Y] = -[Y,X].$$

(c) JACOBI IDENTITY*:*

$$[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]=0.$$

(d) *For* $f,g\in C^\infty(M)$*,*

$$[f\cdot X,g\cdot Y]=fg\cdot[X,Y]+(f\cdot Xg)Y-(g\cdot Yf)X.$$

PROOF. Items (a) and (b) are immediate from the definition. Item (c) is a simple if tedious computation that is left to the bored reader. The mysterious item (d) is left as a homework exercise. □

Let us take a moment to examine the Jacobi identity. In one sense, it is just a mnemonic for how to "associate" three elements (the Lie bracket is non-associative). To try to make a little more sense of it, we introduce some (important) notation. Given a vector field $X\in\mathscr{X}(M)$, denote by $\mathrm{ad}_X\colon\mathscr{X}\to\mathscr{X}$ the linear operator

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

Then we can rewrite the Jacobi identity as follows: first, using antisymmetry twice, write it as $[X,[Y,Z]]=[[X,Y],Z]+[Y,[X,Z]]$. This can be written as

$$\mathcal{L}_X([Y,Z])=[\mathcal{L}_X(Y),Z]+[Y,\mathcal{L}_X(Z)].$$

That is: if we take $\mathscr{X}(M)$ to be a (non-associative) algebra (called a **Lie algebra**), where the product is $[\cdot,\cdot]$, then the Jacobi identity says precisely that $\mathcal{L}_X$ is a derivation on $\mathscr{X}(M)$. This suggests that we can interpret $\mathcal{L}_X$ as a kind of "derivative with respect to X". We will make this precise in the next chapter. We will also return to Proposition 4.21(d) in the next chapter, where it will also make sense as a statement about the Lie bracket being a certain kind of derivative.

Let us conclude this section (and chapter) by noting that the Lie bracket is a "natural" object: it preserves F-relation for any smooth map F.

PROPOSITION 4.22. *Let* M,N *be smooth manifolds and let* $F\colon M\to N$ *be a smooth map. If* $X_1,X_2\in\mathscr{X}(M)$*,* $Y_1,Y_2\in\mathscr{X}(N)$*, and if* (X_j,Y_j) *are* F*-related for* $j=1,2$*, then* $[X_1,X_2]$ *and* $[Y_1,Y_2]$ *are* F*-related.*

PROOF. We use the F-relation twice to compute, for any $f\in C^\infty(N)$,

$$\begin{aligned}X_1X_2(f\circ F)&=X_1((Y_2(f)\circ F)=(Y_1Y_2(f))\circ F,\quad\text{and}\\ X_2X_1(f\circ F)&=X_2(Y_1(f)\circ F)=(Y_2Y_1(f))\circ F.\end{aligned}$$

Thus

$$[X_1,X_2](f\circ F)=X_1X_2(f\circ F)-X_2X_1(f\circ F)=(Y_1Y_2(f))\circ F-(Y_2Y_1(f))\circ F=([Y_1,Y_2](f))\circ F.$$

□

COROLLARY 4.23. *Let* M,N *be diffeomorphic manifolds via diffeomorphism* $F\colon M\to N$*. Let* $X_1,X_2\in\mathscr{X}(M)$*. Then* $F_*[X_1,X_2]=[F_*X_1,F_*X_2]$*.*

CHAPTER 5

Flows

1. Integral Curves

Let M be a smooth manifold, and let $X \in \mathscr{X}(M)$ be a smooth vector field. So $X(p)$ is a tangent vector at p; in particular, it is the tangent vector to lots of curves passing through that point. If α is a smooth curve with the property that, for every point p it passes through, its tangent vector *is* $X(p)$, it is called an **integral curve** of X. That is: the condition is that for $\alpha\colon (t_-, t_+) \to M$,

$$\dot\alpha(t) = X(\alpha(t)), \qquad t \in (t_-, t_+). \tag{5.1}$$

EXAMPLE 5.1. Let $M = \mathbb{R}^2$, and take $X = \frac{\partial}{\partial x}$. Then it is easy to check that any curve of the form $\alpha(t) = p + te^1$, for any $p \in \mathbb{R}^2$, is an integral curve: $\dot\alpha(t) = e^1 \sim \frac{\partial}{\partial x} = X(\alpha(t))$. It is also easy to see that $\alpha(t) = p + te^1$ is the *unique* integral curve passing through p at $t = 0$. Indeed, this is a statement about an ODE: for any curve $\alpha(t) = (x(t), y(t))$,

$$\left.\frac{\partial}{\partial x}\right|_{\alpha(t)} = X(\alpha(t)) = \dot\alpha(t) = d\alpha_t\left(\left.\frac{d}{dt}\right|_t\right) = \frac{dx}{dt}\left.\frac{\partial}{\partial x}\right|_{\alpha(t)} + \frac{dy}{dt}\left.\frac{\partial}{\partial y}\right|_{\alpha(t)}.$$

This sets up the ODE

$$\frac{dx}{dt} = 1, \frac{dy}{dt} = 0, \qquad (x(0), y(0)) = p$$

and it is elementary to verify that the unique solution is $\alpha(t) = (x(t), y(t)) = p + te^1$.

In particular, we see that, for this example, given any $p \in M$, there is a unique integral curve α^p of X that passes through p at time $t = 0$, $\alpha^p(0) = p$; moreover, for any $p, q \in \mathbb{R}^2$, the images of α^p and α^q are either identical (if $p^1 = q^1$) or disjoint (otherwise).

EXAMPLE 5.2. Let $M = \mathbb{R}^2$, and take $Y = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}$. Any integral curve $\alpha(t) = (x(t), y(t))$ is determined by

$$x(t)\left.\frac{\partial}{\partial y}\right|_{\alpha(t)} - y(t)\left.\frac{\partial}{\partial x}\right|_{\alpha(t)} = X(\alpha(t)) = \dot\alpha(t) = \frac{dx}{dt}\left.\frac{\partial}{\partial x}\right|_{\alpha(t)} + \frac{dy}{dt}\left.\frac{\partial}{\partial y}\right|_{\alpha(t)}.$$

In other words, we have the ODE

$$\frac{dx}{dt} = -y(t), \frac{dy}{dt} = x(t).$$

Subject to the constraint $\alpha(0) = (x(0), y(0)) = (x_0, y_0)$, the unique solution is

$$\alpha(t) = (x(t), y(t)) = (x_0 \cos t - y_0 \sin t, x_0 \sin t + y_0 \cos t).$$

When $(x_0, y_0) = (0, 0)$, this is the constant curve $\alpha(t) = 0$; otherwise, it is a counter-clockwise circle. Once again, we see that, for every point $p = (x_0, y_0)$, there is a unique integral curve α^p with $\alpha^p(0) = p$, and the images of any two integral curves are either identical or disjoint.

EXAMPLE 5.3. Let $M = \mathbb{R}^2$, and let $X = x^2 \frac{\partial}{\partial x}$. Then any integral curve $\alpha(t) = (x(t), y(t))$ satisfies the ODE

$$\dot{x}(t) = x(t)^2, \dot{y}(t) = 0.$$

The standard procedure is to "separate variables" and solve for x by $\frac{dx}{x^2} = dt$. This only works if $x \neq 0$, of course; if the initial condition is of the form $\alpha(0) = (0, y_0)$, then we can easily check that the constant curve $\alpha(t) = (0, y_0)$ is the unique solution. Otherwise, if $x_0 \neq 0$, the unique solution with $\alpha(0) = (x_0, y_0)$ is

$$x(t) = \frac{1}{1/x_0 - t}, \quad y(t) = y_0.$$

If $x_0 > 0$, the domain of this curve is $t \in (-\infty, 1/x_0)$. As $t \uparrow 1/x_0$, the curve accelerates off to ∞ along the line $y = y_0$. As $t \downarrow -\infty$, the curve approaches the point $(0, y_0)$. If $x_0 < 0$, the betavior is the same (in the opposite direction).

Integral curves always exist locally.

PROPOSITION 5.4. *Let M be a smooth manifold and $X \in \mathscr{X}(M)$. For each point $p \in M$, there exists an integral curve α for X defined on some time interval $(-\epsilon, \epsilon)$ such that $\alpha(0) = p$.*

PROOF. Let (U, φ) be a chart at p, with coordinate functions $\varphi = (x^1, \ldots, x^n)$. Write α in local coordinates as $\alpha(t) = (\alpha^1(t), \ldots, \alpha^n(t))$. (Technically we should write $\hat{\alpha}(t) = \varphi \circ \alpha(t)$, but at this point we will start dropping that extra notation.) Also write the vector field in local coordinates

$$X = \sum_{j=1}^{N} X^j \frac{\partial}{\partial x^j}.$$

Then (5.1) becomes the ODE

$$\frac{d\alpha^j}{dt} = X^j(\alpha^1(t), \ldots, \alpha^n(t)), \quad 1 \leq j \leq n.$$

By Theorem 0.22 (the Picard–Lindelöf Theorem), there is a unique smooth solution $\alpha(t)$ with $\alpha(0) = p$, for some time interval $(-\epsilon, \epsilon)$, proving the proposition. □

REMARK 5.5. The Picard-Lindelöf theorem shows that the integral curve is unique *in the chart* U. We will need more work to show that it is unique globally on M.

The next few lemmas investigate how integral curves respond to transformations of their vector field.

LEMMA 5.6 (Dilation). *Let M be a smooth manifold, and $X \in \mathscr{X}(M)$. Let $\alpha\colon (t_-, t_+) \to M$ be an integral curve for X. For any $a \in \mathbb{R}$, the curve $\delta_a(\alpha)\colon \{t\colon at \in (t_-, t_+)\} \to M$ defined by $\delta_a(\alpha)(t) = \alpha(at)$ is an integral curve for the vector field aX.*

PROOF. This is just a calculation: for any $t_0 \in \mathbb{R}$ with $at_0 \in (t_-, at_+)$, and any $f \in C^\infty(M)$,

$$\delta_a(\alpha)'(t_0)(f) = \frac{d}{dt}\bigg|_{t_0} f \circ \delta_a(\alpha)(t) = \frac{d}{dt}\bigg|_{t_0} (f \circ \alpha)(at) = a(f \circ \alpha)'(at_0) = a\alpha'(at_0)f = aX(\delta_a(\alpha))f.$$

□

LEMMA 5.7 (Translation). *Let M be a smooth manifold, and $X \in \mathscr{X}(M)$. Let $\alpha\colon (t_0, t_1) \to M$ be an integral curve for X. For any $b \in \mathbb{R}$, the curve $\tau_b(\alpha)\colon (t_0 - b, t_1 - b) \to M$ defined by $\tau_b(\alpha)(t) = \alpha(t + b)$ is also an integral curve of X.*

PROOF. On Homework 3. □

LEMMA 5.8 ("Integral curves are natural"). *Let M, N be smooth manifolds with a smooth map $F\colon M \to N$. Let $X \in \mathscr{X}(M)$ and $Y \in \mathscr{X}(N)$. Then X is F-related to Y iff F takes integral curves of X to integral curves of Y (i.e. for every integral curve $\alpha\colon (t_-, t_+) \to M$ of X, $F \circ \alpha\colon (t_-, t_+) \to N$ is an integral curve of Y.*

PROOF. Suppose X, Y are F-related, i.e. $dF_p(X(p)) = Y(F(p))$ for all $p \in M$. Let α be an integral curve of X; thus $\alpha'(t) = X(\alpha(t))$ for all t. So

$$(F \circ \alpha)'(t) dF_{\alpha(t)}(\alpha'(t)) = dF_{\alpha(t)}(X(\alpha(t))) = Y(F(\alpha(t))) = Y(F \circ \alpha(t)),$$

showing that $F \circ \alpha$ is an integral curve of Y.

Conversely, suppose F maps integral curves of X to integral curves of Y. For any $p \in M$, choose an integral curve α of X with $\alpha(0) = p$. Then

$$dF_p(X(p)) = dF_{\alpha(0)}(X(\alpha(0)) = dF_{\alpha(0)}(\alpha'(0)) = (F \circ \alpha)'(0).$$

By assumption, $F \circ \alpha$ is an integral curve of Y, which means that $(F \circ \alpha)'(0) = Y(F \circ \alpha(0)) = Y(F(p))$. This shows that X, Y are F-related. □

2. Flows

In Examples 5.1 and 5.2, we saw vector fields on a smooth manifold M whose integral curves had the following property:

Given any $p \in M$, there is a unique integral curve $\alpha_p\colon \mathbb{R} \to M$ with $\alpha_p(0) = p$.

As Example 5.3 shows, this property does not always hold: in particular, the curve may not exist for all time. Casting this aside for the moment, suppose X is a vector field with unique integral curves $\{\alpha_p\colon p \in M\}$ all existing for all time. Then we can define a map

$$\theta\colon \mathbb{R} \times M \to M, \quad \theta(t, p) = \alpha_p(t).$$

For fixed $t \in \mathbb{R}$, let $\theta_t\colon M \to M$ be the map $\theta_t(p) = \theta(t, p)$. Note that $\theta_0(p) = \theta(0, p) = \alpha_p(0) = p$ by definition; so $\theta_0 = \mathrm{Id}_M$. Now, let $s, t \in \mathbb{R}$, and consider $\theta(t + s, p)$. This is the point $\alpha_p(t + s)$: the point reached by the unique integral curve of X starting at p after time $t + s$. On the other hand, the translation Lemma asserts that $t \mapsto \alpha_p(t + s)$ is also an integral curve of X; it starts at $q = \alpha_p(s)$. That is, we have the relationship

$$\alpha_p(t + s) = \alpha_q(t), \quad \text{where} \quad q = \alpha_p(s).$$

Translating this into θ-language, this says that $\theta(t + s, p) = \theta(t, q) = \theta(t, \alpha_p(s)) = \theta(t, \theta(s, p))$; or, more succinctly

$$\theta_{t+s} = \theta_t \circ \theta_s. \tag{5.2}$$

We can derive a number of consequences immediately from this.

- For each $t \in \mathbb{R}$, $\theta_t\colon M \to M$ is a bijection. Indeed, $\theta_t \circ \theta_{-t} = \theta_{t-t} = \theta_0 = \mathrm{Id}_M$, so $(\theta_t)^{-1} = \theta_{-t}$.
- Given any two integral curves α_p, α_q, either $\alpha_p(\mathbb{R}) = \alpha_q(\mathbb{R})$, or $\alpha_p(\mathbb{R}) \cap \alpha_q(\mathbb{R}) = \varnothing$. After all, if there is some point $r \in \alpha_p(\mathbb{R}) \cap \alpha_q(\mathbb{R})$, then there are times $t_0, s_0 \in \mathbb{R}$ so that $r = \alpha_p(t_0) = \alpha_q(s_0)$. This means that $\theta_{t_0}(p) = \theta_{s_0}(q)$. Applying (5.2), we have, for *any* t,
$$\theta_t(p) = \theta_{t_0+t-t_0}(p) = \theta_{t-t_0}(\theta_{t_0}(p)) = \theta_{t-t_0}(\theta_{s_0}(q)) = \theta_{t+s_0-t_0}(q).$$

I.e. $\alpha_p(t) = \alpha_q(t + s_0 - t_0) = (\tau_{s_0-t_0}\alpha_q)(t)$, so $\alpha_p = \tau_{s_0-t_0}\alpha_q$. The two curves are translations of each other, and so (since both are defined for all time) they have the same image, as claimed.

Motivated by this discussion, we define a **flow** as follows.

DEFINITION 5.9. *A (smooth)* **global flow** *on M is a smooth function $\theta\colon \mathbb{R} \times M \to M$ for which, writing $\theta_t(p) = \theta(t, p)$, we have $\theta_0 = \mathrm{Id}_M$ and* (5.2) *holds.*

From the above discussion, we see that each map θ_t for fixed $t \in \mathbb{R}$ is a diffeomorphism of M: it is a smooth map with inverse θ_{-t}, which is (by definition) also smooth. So a smooth global flow is a 1-parameter group if diffeomorphisms.

Given a global flow, we get a collection of curves $\{\theta^p\colon p \in M\}$ defined by $\theta^p(t) = \theta_t(p)$ for all $t \in \mathbb{R}$. (So θ^p is what we formerly called α_p.) Since θ is smooth (in both variables), each curve α_p is a smooth curve. By the second item in the discussion above Definition 5.9, the images of these curves are either identical or disjoint, and they (of course) cover M.

Our next proposition shows that any smooth global flow comes from a vector field as per the discussion leading up to the definition.

PROPOSITION 5.10. *Let θ be a smooth global flow on M. For each $p \in M$, let $\theta^p(t) = \theta_t(p) = \theta(t, p)$, and define a vector $X_p \in T_pM$ by*

$$X_p = \dot{\theta^p}(0). \tag{5.3}$$

Then the rough vector field $X\colon p \mapsto X_p$ is smooth, and for each p, θ^p is an integral curve of X starting at p.

We call the vector field X above the **infinitesimal generator** of the flow θ.

PROOF. Let $f \in C^\infty(M)$. Then we calculate

$$X(p)(f) = X_p(f) = (\dot{\theta}^p(0))(f) = d(\theta^p)_p\left(\frac{d}{dt}\bigg|_0\right)(f) = \frac{d}{dt}\bigg|_0 f(\theta^p(t)) = \frac{\partial}{\partial t}\bigg|_{(0,p)} f(\theta(t,p)).$$

The function $f \circ \theta$ is smooth, and hence so are its partial derivatives. This shows that $X(f)$ is smooth, and so by Proposition 4.12 X is a smooth vector field.

Now, we must show that θ^p is an integral curve of X starting at p. We have $\theta^p(0) = \theta_0(p) = p$ by definition. Fix $t_0 \in \mathbb{R}$. Let $q = \theta^p(t_0)$. Then we can compute, for any $f \in C^\infty(M)$,

$$X(q)f = (\dot{\theta}^q(0))(f) = \frac{d}{dt}\bigg|_0 f \circ \theta^q(t) = \frac{d}{dt}\bigg|_0 f(\theta_t(q)).$$

But $\theta_t(q) = \theta_t(\theta^p(t_0)) = \theta_t \circ \theta_{t_0}(p) = \theta_{t+t_0}(p) = \theta^p(t + t_0)$, and so

$$X(q)f = \frac{d}{dt}\bigg|_0 f(\theta^p(t + t_0)) = (\dot{\theta}^p(t_0))(f).$$

That is: we have shown that, as derivations,

$$X(\theta^p(t_0)) = \dot{\theta}^p(t_0), \qquad t_0 \in \mathbb{R}.$$

This concludes the proof that θ^p is an integral curve of X. $\square$

EXAMPLE 5.11. Let $M = \mathbb{R}^n$, and define $\theta(t,p) = e^t p$, the dilation (by exponential time). This is a smooth map of both variables, with $\theta_0(p) = p$ and $\theta_{s+t}(p) = e^{s+t}p = e^s e^t p = \theta_s \circ \theta_t(p)$; so θ is a flow. Its infinitesimal generator E is the vector field

$$E(p)(f) = \frac{\partial}{\partial t}\bigg|_{(0,p)} f(\theta(t,p)) = \frac{\partial}{\partial t}\bigg|_{(0,p)} f(e^t p) = \sum_{j=1}^n p^j e^t \frac{\partial f}{\partial x^j}(e^t p)\bigg|_{t=0} = \sum_{j=1}^n p^j \frac{\partial}{\partial x^j}(p).$$

That is: $E(p) = \sum_{j=1}^n x^j \frac{\partial}{\partial x^j}\big|_p$ is the Euler field of Example 4.3. In other words, the Euler field is the infinitesimal generator of dilations.

As Example 5.3 shows, not every vector field is the infinitesimal generator of a global flow. When this happens, we call the vector field **complete**; more on complete vector fields later. It turns out that the only part that fails is the definition of the integral curves for all time. To deal with this, we weaken the definition of a flow as follows.

DEFINITION 5.12. *Let M be a manifold. A* **flow domain** *$\mathscr{D} \subseteq \mathbb{R} \times M$ is an open set for which each section $\mathscr{D}^p = \{t \in \mathbb{R} \colon (t,p) \in \mathscr{D}\}$ is an open interval in $\mathbb{R}$ containing 0. A (smooth)* **flow** *on M is a smooth function $\theta \colon \mathscr{D} \to M$ defined on some flow domain, with the properties:*

- *$\theta(0,p) = p$ for all $p \in M$, and*
- *for any $p \in M$, if $s \in \mathscr{D}^p$ and $t \in \mathscr{D}^{\theta(s,p)}$ satisfy $s + t \in \mathscr{D}^p$, then*

$$\theta(t, \theta(s,p)) = \theta(t+s, p).$$

Some authors call such an object a **local flow** (to distinguish it from a global flow). As usual, we let $\theta_t(p) = \theta^p(t) = \theta(t,p)$; so θ^p is a curve defined on the open interval $\mathscr{D}^p$. For fixed t, the set of p for which $\theta_t(p)$ makes sense is denoted M_t:

$$M_t = \{p \in M \colon (t,p) \in \mathscr{D}\}.$$

EXAMPLE 5.13. In Example 5.3, we found the integral curve α_p of $x^2 \frac{\partial}{\partial x}$ starting at $p = (x_0, y_0)$ was defined on the time interval $(-\infty, 1/x_0)$ of $x_0 > 0$, on the interval $(1/x_0, \infty)$ if $x < 0$, and on $\mathbb{R}$ if $x_0 = 0$. So we define this as our flow domain:

$$\begin{aligned}\mathscr{D} &= \{(t,(x,y)) \in \mathbb{R} \times \mathbb{R}^2 \colon x > 0, t < 1/x\} \cup \{(t,(x,y)) \in \mathbb{R} \times \mathbb{R}^2 \colon x < 0, t > 1/x\} \cup \mathbb{R} \times (\{0\} \times \mathbb{R}) \\ &= \{(t,(x,y)) \colon xt < 1, y \in \mathbb{R}\}.\end{aligned}$$

This is an open set. The definition as a union of three pieces shows that $\mathscr{D}^{(x,y)}$ for the three regions $x > 0$, $x < 0$, and $x = 0$. Drawing the picture, we see that $(\mathbb{R}^2)_t = \{(x,y) \colon x < 1/t, y \in \mathbb{R}\}$ if $t > 0$, $(\mathbb{R}^2)_t = \{(x,y) \colon x > 1/t, y \in \mathbb{R}\}$ if $t < 0$, and $(\mathbb{R}^2)_0 = \mathbb{R}^2$.

Putting the integral curves of the vector field together, the flow on this domain is

$$\theta(t,(x,y)) = \left(\frac{1}{1/x - t}, y\right) \text{ for } x \neq 0, \text{ and } \theta(t,(0,y)) = (0,y).$$

This is a smooth function on this domain. We can easily check that it satisfies the flow properties (but we don't need to, since this flow arose from the unique integral curves of a vector field).

Note that a smooth (local) flow defines a smooth vector field exactly as a global one does: if $\theta \colon \mathscr{D} \to M$ is a smooth flow, then for each p there is an open neighborhood (in $\mathbb{R} \times M$) of $(0,p)$ contained in $\mathscr{D}$. Thus, the proof of Proposition 5.10 carries through without alteration, and we find that any flow has an infinitesimal generator defined by (5.3).

Given a flow θ on a domain $\mathscr{D}$, we can always restrict it to a smaller domain $\mathscr{D}' \subset \mathscr{D}$ (provided $\mathscr{D}'$ is still a flow domain). For example, we could take the global flows of Examples 5.1 and 5.2

and restrict them to some flow domain smaller than $\mathbb{R}^2$. Then the infinitesimal generator would still be the same vector field in each case – it is determined only by the behavior of the flow in any small tubular neighborhood of the section $\{t = 0\}$ in the flow domain. This highlights that one cannot recover the flow domain from the infinitesimal generator, as we've defined things. But we can fix this by insisting the flow domain be as large as possible.

DEFINITION 5.14. *A* **maximal flow** *is a flow θ on a flow domain $\mathscr{D}$ with the property that the function θ has no smooth extension to* any *flow domain larger than $\mathscr{D}$. A* **maximal integral curve** *of a vector field is an integral curve which has no smooth extension to an integral curve on a strictly larger interval.*

With these definitions, vector fields and flows come into one-to-one correspondence.

THEOREM 5.15 (Fundamental Theorem on Flows). *Let M be a smooth manifold, and let $X \in \mathscr{X}(M)$. There is a unique maximal flow $\theta\colon \mathscr{D} \to M$ whose infinitesimal generator is X. Moreover, the flow satisfies the following properties.*

(a) *For each $p \in M$, the curve θ^p is the unique maximal integral curve of X starting at p.*
(b) *If $s \in \mathscr{D}^p$, then $\mathscr{D}^{\theta(s,p)} = \mathscr{D}^p - s = \{t - s\colon t \in \mathscr{D}^p\}$.*
(c) *For each $t \in \mathbb{R}$, the set M_t is open in M, and $\theta_t\colon M_t \to M_{-t}$ is a diffeomorphism, with inverse θ_{-t}.*

In order to prove Theorem 5.15, we will first show that every vector field has a unique *maximal* integral curve starting at any given point.

LEMMA 5.16. *Let M be a smooth manifold, and let $X \in \mathscr{X}(M)$. For each $p \in M$, there is a unique maximal integral curve α_p with $\alpha_p(0) = p$.*

PROOF. For each $p \in M$, by Proposition 5.4 there is an integral curve $\alpha\colon (-\epsilon_p, \epsilon_p) \to M$ that starts at p. Now, let α and $\tilde{\alpha}$ be two integral curves of X defined on intervals (t_-, t_+) and $(\tilde{t}_-, \tilde{t}_+)$ (not necessarily containing 0). Suppose the two time intervals intersect: $(t_-, t_+) \cap (\tilde{t}_-, \tilde{t}_+) = (a, b)$, and suppose there exists some $t_0 \in (a, b)$ with $\alpha(t_0) = \tilde{\alpha}(t_0)$. We will show that, in fact, $\alpha(t) = \tilde{\alpha}(t)$ for all $t \in (a, b)$. To do this, define $\mathscr{T} = \{t \in (a, b)\colon \alpha(t) = \tilde{\alpha}(t)\}$. Then $t_0 \in \mathscr{T}$, so the set is nonempty. Now, fix any metric d_M that metrizes the topology of M, and note that $\mathscr{T} = \{t \in (a, b)\colon d_M(\alpha(t), \tilde{\alpha}(t)) = 0\}$. The function $t \mapsto d_M(\alpha(t), \tilde{\alpha(t)})$ is continuous, and so $\mathscr{T}$ is the preimage of $\{0\}$ under a continuous map; it is therefore a closed set.

Now, let t_1 is any point in $\mathscr{T}$. In a chart at $\alpha(t_1) = \tilde{\alpha}(t_1) \in M$, the condition that α and $\tilde{\alpha}$ are integral curves passing through the same point at time t_1 shows that they are both solutions of the same ODE, and hence they are equal in this chart (by the Picard-Lindelöf Theorem 0.22). This shows that, for any $t_1 \in \mathscr{T}$, there is an open set of times containing t_1 at which the two curves agree, meaning that there is an open neighborhood of t_1 contained in $\mathscr{T}$. Hence, $\mathscr{T}$ is open. As (a, b) is connected, it follows that the clopen set $\mathscr{T} \subseteq (a, b)$ is equal to (a, b).

Hence, we have shown that any two integral curves of X that agree at one point must agree on their common domain. Now, for any $p \in M$, define $\mathscr{D}^p$ to be the union of all open intervals containing 0 on which there is an integral curve of X starting at p. By the above, for any $t \in \mathscr{D}^p$, all integral curves starting at p have the same value at t; so we may define $\alpha_p(t)$ to be this common value. By definition, α_p is an integral curve defined on this union of intervals. It is maximal: if it could be extended, then that extension would have been included in the union. It is unique: any other integral curve agrees with α_p at the time 0, and by the above, that means they agree everywhere. This concludes the proof. $\square$

PROOF OF THEOREM 5.15. For each $p \in M$, let $\mathscr{D}^p$ be the domain of the maximal integral curve of X starting at p, and define $\mathscr{D} = \{(t,p) \in \mathbb{R} \times M : t \in \mathscr{D}^p\}$. We then define $\theta \colon \mathscr{D} \to M$ by $\theta(t,p) = \alpha_p(t)$. As usual, we will write $\theta^p(t) = \theta_t(p) = \theta(t,p)$. So, in particular, $\theta^p = \alpha_p$, which gives part (a) by definition. We will verify the following things.

(1) For any $p \in M$, if $s \in \mathscr{D}^p$ and $t \in \mathscr{D}^{\theta(s,p)}$ satisfy $s + t \in \mathscr{D}^p$, then $\theta(t, \theta(s,p)) = \theta(t+s, p)$.
(2) Part (b).
(3) $\mathscr{D}$ is open, and θ is smooth.
(4) Part (c).

(1) Fix $p \in M$, let $s \in \mathscr{D}^p$, and set $q = \theta(s,p) = \theta^p(s)$. The curve $\alpha \colon \mathscr{D}^p - s \to M$ defined by $\alpha = \tau_s \theta^p$ is an integral curve (by the translation lemma) and starts at q. By Lemma 5.16, α agrees with θ^q on their common domain, which is precisely to say that if $t \in \mathscr{D}^{\theta(s,p)}$ with $s + t \in \mathscr{D}^p$ then $\theta^q(t) = \tau_s \theta^p(t) = \theta^p(t+s)$. I.e. $\theta(t, \theta(s,p)) = \theta(s+t, p)$, as required.

(2) In part (1), we saw that the domain of α was $\mathscr{D}^p - s$; this must be contained in the domain of the maximal integral curve, so $\mathscr{D}^p - s \subseteq \mathscr{D}^q = \mathscr{D}^{\theta(s,p)}$. On the other hand, since $0 \in \mathscr{D}^p$, and so $-s = 0 - s \in \mathscr{D}^p - s$, we have $-s \in \mathscr{D}^q$, and by the group law (verified in part (1)) we get $\theta^q(-s) = p$. Thus, applying the same argument with $(-s, q)$ in the place of (s,p), we see that $\mathscr{D}^q + s \subseteq \mathscr{D}^p$, proving the reverse containment, and giving part (b).

(3) This is the tricky one. We define a set $W \subseteq \mathscr{D}$ as follows: $(t,p) \in W$ iff there is some product neighborhood $J \times U$, where J is an open interval containing 0 and t, and U an open neighborhood of p in M, on which θ is defined and smooth. Note that, with $t = 0$, if we take a chart at p, then in local coordinates we may apply the Picard-Lindelöf theorem (part 3: smooth dependence on initial conditions) to conclude that θ is defined an smooth on a product neighborhood of $(0,p)$. Thus, W contains $\{0\} \times M$, and so is not empty. What's more, if $(t,p) \in W$, then all points in the neighborhood $J \times U$ presumed to exist are also in W (they all have neighborhood $J \times U$), and so W is open. We will show that $W = \mathscr{D}$; this will show that $\mathscr{D}$ is open, and (by definition of W) that θ is smooth on its domain. We prove this by contradiction: suppose there is some point $(\tau, p_0) \in \mathscr{D} \setminus W$. By the above, $\tau \neq 0$. Wlog, assume $\tau > 0$ (the $\tau < 0$ case is similar).

Let $t_0 = \inf\{t \in \mathbb{R} \colon (t, p_0) \notin W\}$. Since θ is defined and smooth in some product neighborhood of $(0, p_0)$, we must have $t_0 > 0$. Since $t_0 \leq \tau$ and $\mathscr{D}^{p_0}$ is an open interval containing 0 and τ, it follows that $t_0 \in \mathscr{D}^{p_0}$. Set $q_0 = \theta^{p_0}(t_0)$. By the above discussion, there is some neighborhood U_0 of q_0 and an interval $(-\epsilon, \epsilon)$ such that $(-\epsilon, \epsilon) \times U_0 \subseteq W$. The idea is to now use the group law to show that this neighborhood can be translated to give a smooth extension of θ to a neighborhood of (t_0, p_0), contradicting the definition of t_0.

To be precise: choose some $t_1 < t_0$ within distance ϵ (so $t_1 + \epsilon > t_0$) and such that $\theta^{p_0}(t_1) \in U_0$ (possible because $\theta^{p_0}(t_0) = q_0$, U_0 is a neighborhood of q_0, and θ^{p_0} is continuous). Now, by definition of t_0, since $t_1 < t_0$ we have $(t_1, p_0) \in W$. This means there is a product neighborhood $(t_1 - \epsilon, t_1 + \epsilon) \times U_1$ contained in W, which means that θ is defined and smooth there – and (since $t_1 - \epsilon < t_0$) also below: θ is smooth on $[0, t_1 + \epsilon) \times U_1$. By taking U_1 small, we may assume that $\theta(t_1, U_1) \subseteq U_0$ (since $\theta(t_1, p_0) \in U_0$). Now we define our extension: $\tilde{\theta} \colon [0, t_1 + \epsilon) \times U_1 \to M$ is defined to be

$$\tilde{\theta}(t,p) = \begin{cases} \theta_t(p), & p \in U_1, u \leq t < t_1, \\ \theta_{t-t_1} \circ \theta_{t_1}(p), & p \in U_1, t_1 - \epsilon < t < t_1 + \epsilon. \end{cases}$$

The group law for θ guarantees that this is well-defined (it agrees on the overlap), and by the preceding discussion $\tilde{\theta}$ is smooth. Now, by the translation lemma, each curve $t \mapsto \tilde{\theta}(t,p)$ is an integral curve, and so $\tilde{\theta}$ is a smooth extension of θ to a neighborhood of (t_0, p_0). This contradicts the choice of t_0. Hence, there cannot have been any point $(\tau, p_0) \in \mathscr{D} \setminus W$, concluding the proof that $W = \mathscr{D}$.

(4) Finally, we prove (c). Since $\mathscr{D}$ is open, $M_t = \{p \in M \colon (t,p) \in \mathscr{D}\}$ is open (by the definition of the product topology on $\mathbb{R} \times M$). Using part (b), we have

$$p \in M_t \implies t \in \mathscr{D}^p \implies \mathscr{D}^{\theta_t(p)} = \mathscr{D}^p - t \implies -t \in \mathscr{D}^{\theta_t(p)} \implies \theta_t(p) \in M_{-t}.$$

This shows that $\theta_t(M_t) \subseteq M_{-t}$. Now, the group law shows that $\theta_t \circ \theta_{-t}(p) = p$ for all $p \in M_{-t}$. What's more, the same argument as above with t and $-t$ reversed shows that $\theta_{-t}(M_{-t}) \subseteq M_t$. So any $p \in M_{-t}$ is equal to $\theta_t(q)$ for a point $q = \theta_{-t}(p) \in M_t$. This shows θ_t mapso M_t onto M_{-t}. The argument also shows that it is a bijection with inverse θ_{-t}. Since the flow is smooth, both θ_t and θ_{-t} are smooth, and hence they are diffeomorphisms. $\square$

Thus, every smooth vector field X is the infinitesmal generator of a unique maximal flow θ; we call it the **flow generated by** X or simply the **flow of** X.

To conclude our present discussion, let's note that Lemma 5.8, which says that integral curves are natural, implies that the collection of all integral curves together (the flow) is also a natural construction.

LEMMA 5.17 ("flows are natural"). *Let M and N be smooth manifolds, and let $F \colon M \to N$ be a smooth map. Let $X \in \mathscr{X}(M)$ and $Y \in \mathscr{X}(N)$, with flows θ and η respectively. If X, Y are F-related, then for each $t \in \mathbb{R}$ $F(M_t) \subseteq N_t$, and $\eta_t \circ F = F \circ \theta_t$ on M_t:*

$$\begin{tikzcd} M_t \arrow[r, "F"] \arrow[d, "\theta_t"'] & N_t \arrow[d, "\eta_t"] \\ M_{-t} \arrow[r, "F"'] & N_{-t} \end{tikzcd}$$

PROOF. By Lemma 5.8, given ant $p \in M$, the curve $F \circ \theta^p$ is an integral curve of Y, which starts at $F \circ \theta^p(p) = F(p)$. By uniqueness of (maximal) integral curves, the maximal integral curve $\eta^{F(p)}$ is defined on at least the interval $\mathscr{D}^p$, and $F \circ \theta^p = \eta^{F(p)}$ on this interval. This means that

$$p \in M_t \implies t \in \mathscr{D}^p \implies t \in \mathscr{D}^{F(p)} \implies F(p) \in N_t,$$

which is equivalent to $F(M_t) \subseteq N_t$. What's more, the statement $F(\theta^p(t)) = \eta^{F(p)}(t)$ for all $t \in \mathscr{D}^p$ is precisely the statement that $\eta_t \circ F(p) = F \circ \theta_t(p)$ for all $p \in M_t$, as desired. $\square$

COROLLARY 5.18. *Let M and N be a smooth manifolds, and let $F \colon M \to N$ be a diffeomorphism. If $X \in \mathscr{X}(M)$ with flow θ, then the flow of $F_*(X)$ is $\eta_t = F \circ \theta_t \circ F^{-1}$, with domain $N_t = F(M_t)$ for each $t \in \mathbb{R}$.*

As we mentioned just above Definition 5.12, a vector field is called **complete** if it is the infinitesimal generator of a global smooth flow. In light of Theorem 5.15, we can phrase this more accurately as: a vector field is complete if its flow is global, i.e. if its integral curves all exists for all time. It is generally quite hard to decide when this is the case. Example like 5.1 and 5.2 that can be solved completely explicitly are rare. Usually it is difficult to decide whether a given local integral curve can be extended. We do have the following tool.

LEMMA 5.19 (Uniform Time Lemma). *Let M be a smooth manifold and $X \in \mathscr{X}(M)$. Suppose there is a uniform $\epsilon > 0$ so that, for every $p \in M$, the integral curve θ^p of X starting at p is defined on $(-\epsilon, \epsilon)$. Then X is complete.*

This is a typical "bootstrapping" argument. Let $q = \theta^p(\epsilon/2)$; by assumption θ^q also extends past time $\epsilon/2$, so let $r = \theta^q(\epsilon/2)$; by assumption, θ^r extends past time $\epsilon/2$, so let $s = \theta^r(\epsilon/2)$; and so forth. Then the group law shows that $\theta^p(t)$ exists up to time $3\epsilon/2$: $\theta^p(t) = \theta^s(t-\epsilon)$ for $\epsilon \leq t < 3\epsilon/2$. We can continue this way extending $\epsilon/2$ time units at a time, out to ∞. In the proof, we give a shorter treatment of this idea by a contradiction proof.

PROOF. For a contradiction, we assume that there exists a $p \in M$ for which $\mathscr{D}^p$ is bounded above (a similar proof works in the case it is assumed to be bounded below). Fix some time t_0 with $\sup \mathscr{D}^p - \epsilon < t_0 < \sup \mathscr{D}^p$. Let $q = \theta^p(t_0)$. By assumption, θ^q is defined on $(-\epsilon, \epsilon)$. So we may define a curve $\alpha\colon (-\epsilon, t_0 + \epsilon)$ by

$$\alpha(t) = \begin{cases} \theta^p(t), & -\epsilon < t < \sup \mathscr{D}^p, \\ \theta^q(t - t_0), & t - \epsilon < t < t_0 + \epsilon. \end{cases}$$

By the group law, we have $\theta^q(t - t_0) = \theta(t - t_0, q) = \theta(t - t_0, \theta^p(t_0)) = \theta_{t-t_0} \circ \theta_{t_0}(p) = \theta_t(p) = \theta^p(t)$. This shows that the two pieces of the definition of α agree on their overlap. By the translation lemma, α is an integral curve, and it starts at p. But α is defined on a domain strictly larger than the maximal domain $\mathscr{D}^p$ of θ^p, which is a contradiction. □

This gives us at least one class of vector fields that are complete.

THEOREM 5.20. *Every compactly supported smooth vector field is complete.*

PROOF. Let $X \in \mathscr{X}(M)$ be compactly supported, with $K = \operatorname{supp} X$. Let $p \in K$; then, since the flow θ of X is defined on an open set, there is some product neighborhood $(-\epsilon_p, \epsilon_p) \times U_p$ of $(0, p)$ where θ is defined. By compactness of K, there are finitely many points $p_1, \ldots, p_k \in K$ so that the open sets $U_{p_1}, \ldots, U_{p_k}$ cover K. Set $\epsilon = \min\{\epsilon_{p_1}, \ldots, \epsilon_{p_k}\}$. Now, for any point $q \in K$, choose some p_j such that $q \in U_{p_j}$; then, since θ is defined on $(-\epsilon_{p_j}, \epsilon_{p_j}) \times U_{p_j}$, it follows that the integral curve θ^q is defined on $(-\epsilon_{p_j}, \epsilon_{p_j})$, and so therefore on the interval $(-\epsilon, \epsilon)$. Hence, every integral curve starting in K is defined on $(-\epsilon, \epsilon)$. If, on the other hand, $q \notin K = \operatorname{supp} X$, then $X(q) = 0$, and so the integral curve θ^q is constant and defined for all time (so in particular on $(-\epsilon, \epsilon)$. It follows from the uniform time lemma that X is complete. □

COROLLARY 5.21. *Every smooth vector field on a compact smooth manifold is complete.*

3. Lie Derivatives

Consider an old-fashioned vector field $\mathbf{X}\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$. We were able to take derivatives of such vector fields by treating them as functions: the directional derivative in some direction v was, as usual

$$D_v\mathbf{X}(p) = \frac{d}{dt}\bigg|_{t=0} \mathbf{X}(p + tv) = \lim_{t \to 0} \frac{\mathbf{X}(p + tv) - \mathbf{X}(p)}{t}.$$

Indeed, this simply means taking directional derivatives of the components independently. However, this heavily depends on the vector space structure of $\mathbb{R}^n$ (so that we can take $p + tv$).

We could try to fix this as follows. Let $X \in \mathscr{X}(M)$, and let $p \in M$. For a vector $v \in T_pM$, fix some curve $\alpha\colon (-\epsilon, \epsilon) \to M$ so that $\alpha(0) = p$ and $\dot{\alpha}(0) = p$. Then we could try to define

$$D_vX(p) = \lim_{t\to 0} \frac{X(\alpha(t)) - X(p)}{t}.$$

But this is still problematic. First, which curve should we choose? There is no reason to think that the answer will depend only on $\dot{\alpha}(0)$. But the bigger problem is: the difference quotient actually doesn't make sense: $X(\alpha(t)) \in T_{\alpha(t)}M$ while $X(p) \in T_pM$, different vector spaces. If $M = \mathbb{R}^n$, we can canonically identify these two spaces, but in general, there is no coordinate-free way to do so.

There is no solution to this problem: there is no way to take the directional derivative at some point p of a vector field in the direction of some vector living only in T_pM that is coordinate independent (i.e. well-defined). But there is a solution if we take the directional derivative in the direction of another *vector field*: the idea being that we should evaluate X at points along the flow of the other vector field in the difference quotient. Here is the definition.

DEFINITION 5.22. *Let M be a smooth manifold, and fix two vector fields $X, Y \in \mathscr{X}(M)$. Let θ be the flow of Y. The* **Lie derivative** *of X with respect to Y is defined to be the (rough) vector field $\mathscr{L}_Y(X)$ is defined by*

$$\mathscr{L}_Y(X)(p) = \frac{d}{dt}\bigg|_{t=0} d(\theta_{-t})_{\theta_t(p)}(X_{\theta_t(p)}) = \lim_{t\to 0} \frac{d(\theta_{-t})_{\theta_t(p)}(X_{\theta_t(p)}) - X_p}{t}. \tag{5.4}$$

Actually, we need to verify that this definition makes sense (i.e. that the limit exists), but at least the difference quotient itself makes sense. We evaluate X at the point $\theta_t(p)$ that p flows to under the flow of X. This gives us a vector in $T_{\theta_t(p)}M$, which we cannot compare to $X(p)$. So we transform that vector back into T_pM by using the backwards flow, or rather the differential of the backwards flow, which is its infinitesimal action on vectors, giving us a vector back in T_pM (since $\theta_{-t}\colon \theta_t(p) \mapsto p$). Note: although θ is not generally globally defined, for fixed p the flow domain $\mathscr{D}^p$ is an interval containing 0, and so for small enough t the difference quotient makes sense.

By Theorem 5.15, θ_{-t} is a diffeomorphism from $M_{-t} \to M_t$, where $M_{-t} = \{q \in M\colon -t \in \mathscr{D}^q\}$ is open. If we choose t small enough that $-t \in \mathscr{D}^p$, then $p \in M_{-t}$; then we can restrict X to the neighborhood M_{-t} of p, and we have (from Proposition 4.16)

$$d(\theta_{-t})_{\theta_t(p)}(X(\theta_t(p))) = (\theta_{-t})_*(X)(p).$$

Thus, we can write

$$\mathscr{L}_Y(X)(p) = \frac{d}{dt}\bigg|_{t=0} (\theta_{-t})_*(X)(p), \tag{5.5}$$

again, provided this limit makes sense. Note: we state this at a given point $p \in M$ to emphasize the point that the derivative being taken is of the function $t \mapsto (\theta_{-t})_*(X)(p)$, which is a map from some time interval into the tangent space T_pM: it is a regular calculus function.

Proposition 5.24 below shows that the Lie derivative does exists, and is in fact a familiar object. First we need the following lemma.

LEMMA 5.23. *Let M be a smooth manifold, let $\epsilon > 0$, and let $f \in C^\infty((-\epsilon, \epsilon) \times M)$. For each $t \in (-\epsilon, \epsilon)$, let $f_t(p) = f(t, p)$. If $X \in C^\infty(M)$, then $(t, p) \mapsto X(f_t)$ is smooth function on $(-\epsilon, \epsilon) \times M$.*

PROOF. Let (U, φ) be a chart in M, with coordinate functions $\varphi = (x^1, \dots, x^n)$. Then we can write the vector field X restricted to U in the coordinate basis: for any $p \in U$,

$$X_p = \sum_{j=1}^{n} X^j(p) \left.\frac{\partial}{\partial x^j}\right|_p .$$

Write f_t in local coordinates $\hat{f}_t = f_t \circ \varphi^{-1}$. Since φ is a diffeomorphism $U \to \hat{U} = \varphi(U)$, the composition $(-\epsilon, \epsilon) \times \hat{U} \ni (t, x) \mapsto \hat{f}_t = f(t, \varphi^{-1}(X))$ is C^∞, and therefore so are all its partial derivatives. Letting $\hat{p} = \varphi(p)$, we have

$$X(f_t)(p) = \sum_{j=1}^{n} X^j(p) \frac{\partial \hat{f}_t}{\partial x^j}(\hat{p}),$$

which is therefore C^∞ in both variables. As this holds true in any chart U, it holds globally. $\square$

PROPOSITION 5.24. *For any $X, Y \in \mathscr{X}(M)$, let θ be the flow of Y. Then the function $t \mapsto (\theta_{-t})_*(X)(p)$ is smooth for each $p \in M$, and its derivative at $t = 0$ is $\mathscr{L}_Y(X) = [Y, X]$.*

PROOF. We apply $(\theta_{-t})_*(X)|_p$ to a function $f \in C^\infty(M)$; this gives

$$(\theta_{-t})_*(X)f(p) = d(\theta_{-t})_{\theta_t(p)}(X_{\theta_t(p)})f = X_{\theta_t(p)}(f \circ \theta_{-t}).$$

So we are supposed to take the limit as $t \to 0$ of the difference quotient

$$\frac{(\theta_{-t})_*(X)_p - X_p}{t}(f) = \frac{1}{t}\left[X_{\theta_t(p)}(f \circ \theta_{-t}) - X_p f\right].$$

Now we play the usual trick of adding and subtracting a connecting term: the above is equal to

$$\frac{1}{t}\left[X_{\theta_t(p)}(f \circ \theta_{-t}) - X_{\theta_t(p)}(f) + X_{\theta_t(p)}(f) - X_p f\right]$$

The last two terms gives us $YXf(p)$ in the limit as follows:

$$\lim_{t \to 0} \frac{X_{\theta_t(p)}(f) - X_p f}{t} = \left.\frac{d}{dt}\right|_0 Xf(\theta_t(p)) = \left.\frac{d}{dt}\right|_0 (Xf) \circ \theta^p(t) = \dot{\theta}^p(0)(Xf) = Y_p(Xf) = YX(f)|_p$$

since θ^p is the maximal integral curve of Y that starts at p (so in particular $\dot{\theta}^p(0) = Y(\theta^p(0)) = Y(p)$). So, we have shown that

$$\mathscr{L}_Y(X)|_p(f) = \lim_{t \to 0} \frac{X_{\theta_t(p)}(f \circ \theta_{-t}) - X_{\theta_t(p)}(f)}{t} + YX(f)|_p \tag{5.6}$$

provided this limit exists.

The key here is Taylor's theorem. For $q \in M$ and $t \in -\mathscr{D}^q$, let $g(t, q) = f(\theta_{-t}(q))$; for fixed $q \in M$, this is a smooth function $-\mathscr{D}^q \to \mathbb{R}$. In particular, it is a smooth calculus function defined on some neighborhood of 0, and so Taylor's theorem gives us $g(t, q) = g(0, q) + t\partial_1 g(0, q) + O(t^2) = f(q) + t\partial_1 g(0, q) + O(t^2)$ (since $f(0, q) = f(\theta_0(q)) = f(q)$). We'll need more precise control over the q-dependence of the error term, so we use Theorem 0.9 which gives the error term as an integral. In fact, we need only expand to first order:

$$f(\theta_{-t}(q)) = g(t, q) = f(q) + t \int_0^1 \partial_1 g(ts, q)\, ds \equiv f(q) + t h_t(q).$$

As g is C^∞ in both variables in a neighborhood of $(0,p)$, the same is true of its partial derivatives and their integrals, so h_t is a smooth function (for all small enough t for which it is defined). We may apply the vector field

$$X_{\theta_t(p)}(f\circ\theta_{-t}) = X_{\theta_t(p)}(f) + tX_{\theta_t(p)}(h_t).$$

Hence, the remaining limit difference quotient in (5.6) simplifies:

$$\lim_{t\to 0}\frac{X_{\theta_t(p)}(f\circ\theta_{-t}) - X_{\theta_t(p)}(f)}{t} = \lim_{t\to 0} X_{\theta_t(p)}(h_t) = \lim_{t\to 0}(X(h_t))(\theta_t(p)).$$

By Lemma 5.23, $(t,p)\mapsto X(h_t)(p)$ is a smooth function, and since $\theta_t(p)$ is also a smooth function of (t,p), this limit is equal to $X(h_0)(\theta_0(p)) = X_p(h_0)$. But

$$h_0(p) = \int_0^1 \partial_1 g(0,p)\,ds = \partial_1 g(0,p) = \left.\frac{\partial}{\partial t}\right|_{t=0} f(\theta_{-t}(p)).$$

Applying the chain rule to the change of variables $s=-t$ gives

$$h_0(p) = -\left.\frac{\partial}{\partial s}\right|_{s=0} f(\theta_s(p)) = -\left.\frac{\partial}{\partial s}\right|_{s=0} f(\theta^p(s)) = -\dot\theta^p(0)(f) = -Y_p(f).$$

Hence, finally combining with (5.6), we have

$$\mathscr{L}_Y(X)|_p\,(f) = X(-Y(f))|_p + YX(f)|_p = [Y,X](f)|_p\,.$$

This concludes the proof. □

REMARK 5.25. The Lie bracket $[Y,X]$ and the Lie derivative $\mathscr{L}_Y(X)$ were both known to be coordinate independent for a long time, but the only known proof of their equality required computations in local coordinates (which is the way [**3**] approaches the computation). The above invariant proof is relatively new (less than half a century old).

While we are on the subject of Lie derivatives, we can give a similar definition of the Lie derivative of a smooth *function* with respect to a vector field: the result $\mathscr{L}_X f$ is a new smooth function,

$$\mathscr{L}_X f(p) = \left.\frac{d}{dt}\right|_{t=0} f\circ\theta_t(p) = \lim_{t\to 0}\frac{f\circ\theta_t(p) - f(p)}{t}.$$

This, again, is an appealing interpretation of what should replace $D_v f(p) = \left.\frac{d}{dt}\right|_{t=0} f(p+tv)$ in the linear case, if $v = X(p)$. Here we can see easily that

$$\mathscr{L}_X f(p) = \left.\frac{d}{dt}\right|_{t=0} f(\theta^p(t)) = \dot\theta^p(0)(f) = X(f).$$

This is consistent with our reinterpretation of tangent vectors as "directional derivative operators". So, in summary, we have Lie derivatives (so far) of functions and vector fields:

$$\mathscr{L}_X f = Xf, \qquad \mathscr{L}_X Y = [X,Y].$$

The formula for the Lie derivative (being simply a Lie bracket) gives rise to a whole host of properties that are not at all obvious from the definition. We summarize them here.

PROPOSITION 5.26. *Let M be a smooth manifold, and let $X,Y,Z\in\mathscr{X}(M)$.*

(a) $\mathscr{L}_X Y = -\mathscr{L}_Y X$.
(b) $\mathscr{L}_X[Y,Z] = [\mathscr{L}_X Y, Z] + [Y, \mathscr{L}_X Z]$.
(c) $\mathscr{L}_{[X,Y]}(Z) = \mathscr{L}_X\mathscr{L}_Y Z - \mathscr{L}_Y\mathscr{L}_X Z$.

(d) *If* $g \in C^\infty(M)$*, then* $\mathscr{L}_X(gY) = \mathscr{L}_X g \cdot Y + g\mathscr{L}_X Y = (Xg)Y + g\mathscr{L}_X Y$.
(e) *If* $F\colon M \to N$ *is a diffeomorphism, then* $F_*(\mathscr{L}_X Y) = \mathscr{L}_{F_*X}(F_*Y)$.

These are easy calculations left to the reader.

REMARK 5.27. We already saw Proposition 5.26(c) show up as the Jacobi identity of Proposition 4.21(c); now we see that it really makes sense as a statement that the Lie derivative is a *Lie* derivation. Similarly, the mysterious identity of Proposition 4.21(d) is just Proposition 5.26(d) applied to $[fX, gY] = \mathscr{L}_{fX}(gY) = -\mathscr{L}_{gY}(fX)$ in each variable separately. Item (e) shows that the Lie derivative is a natural construction.

4. Commuting Vector Fields

We say two vector fields $X, Y \in \mathscr{X}(M)$ **commute** if $[X, Y] = 0$. In light of the previous section, this means that $\mathscr{L}_X(Y) = 0$ (and also $\mathscr{L}_Y(X) = 0$), so that "X does not vary in the Y direction" and vice versa. This is a little cumbersome to understand since the Lie derivative is a complicated object. In fact, it means the following.

DEFINITION 5.28. *Let* $X \in \mathscr{X}(M)$*, and let* θ *be a smooth flow on* M*. Say that* X *is* **invariant under the flow** θ *if* X *is* θ_t*-related to itself for each* t*; more precisely, if* $X|_{M_t}$ *is* θ_t*-related to* $X|_{M_{-t}}$ *for each* $t \in \mathbb{R}$*. Equivalently, this says that*

$$d(\theta_t)_p(X_p) = X_{\theta_t(p)}, \qquad (t, p) \in \mathscr{D}(\theta).$$

So, to say X is invariant under θ means that pushing it forward by the diffeomorphism θ_t doesn't change X for any t. For example, it is clear that X is invariant under its own flow (this is precisely what it means for θ^p to be an integral curve of X). In fact, X is invariant under the flow of any vector field it commutes with.

THEOREM 5.29. *Let* $X, Y \in \mathscr{X}(M)$*. TFAE:*
(a) $[X, Y] = 0$.
(b) X *is invariant under the flow of* Y.
(c) Y *is invariant under the flow of* X.

PROOF. We will show that (a) $\iff$ (b); the equivalence (a) $\iff$ (c) is the same (up to a minus sign). Let θ be the flow of Y.

(b) $\implies$ (a): By assumption, $X_{\theta_t(p)} = d(\theta_t)_p(X_p)$ whenever $(t, p) \in \mathscr{D}(\theta)$. That is: $X_p = d(\theta_{-t})_{\theta_t(p)}(X_{\theta_t(p)})$ (since $\theta_{-t} = \theta_t^{-1}$). Referring the (5.4) defining the Lie derivative, it follows immediately that $\mathscr{L}_Y(X)|_p = 0$. By Proposition 5.24, it follows that $[X, Y]|_p = 0$. Since every $p \in M$ is in M_t and M_{-t} for all sufficiently small t, this shows $[X, Y] = 0$, confirming (a).

(a) $\implies$ (b): We are assuming that $\mathscr{L}_Y(X) = [Y, X] = 0$. Fix $p \in M$, and define a function $\alpha\colon \mathscr{D}^p \to T_pM$ by

$$\alpha(t) = d(\theta_{-t})_{\theta_t(p)}(X_{\theta_t(p)}) = (\theta_{-t})_*(X)|_p.$$

Then α is differentiable: it is clearly smooth on $\mathscr{D}^p \setminus \{0\}$, and by (5.5) its derivative at 0 is $\mathscr{L}_Y(X)|_p = 0$. In fact, let us compute $\alpha'(t_0)$ for any $t_0 \in \mathscr{D}^p$ (that is, the usual derivative of a curve in a vector space). We do this by translating t_0 to 0.

$$\alpha'(t_0) = \frac{d}{ds}\bigg|_{s=0} \alpha(t_0 + s) = \frac{d}{ds}\bigg|_{s=0} d(\theta_{-t_0-s})(X_{\theta_{t_0+s}(p)}).$$

Now, $\theta_{t_0+s} = \theta_{t_0} \circ \theta_s$ on a neighborhood of p, and since θ_{t_0} is independent of s, we therefore have

$$\alpha'(t_0) = d(\theta_{-t_0}) \left(\frac{d}{ds}\bigg|_{s=0} d(\theta_{-s})(X_{\theta_s(\theta_{t_0}(p))}) \right) = d(\theta_{-t_0})(\mathscr{L}_Y(X)|_{\theta_{t_0}(p)}) = 0.$$

Hence, α is constant, and since $\alpha(0) = X_p$, we have $\alpha(t) = X_p$ for all $t \in \mathscr{D}^p$. This is precisely to say that X is invariant under θ. □

EXAMPLE 5.30. We can use Theorem 5.29 to characterize what vector fields are invariant under the flows of Examples 5.1, 5.2, and 5.11.

- *Translation*: the flow $\tau_t(x, y) = (x + t, y)$ is generated by the vector field $Y = \frac{\partial}{\partial x}$. Let $X = X^1 \frac{\partial}{\partial x} + X^2 \frac{\partial}{\partial y}$ be a vector field on $\mathbb{R}^2$ that is invariant under translations in the e_1 direction, meaning invariant under the flow τ. This is the same as insisting that $[X, Y] = 0$. We can compute, for any $f \in C^\infty(\mathbb{R}^2)$,

$$[X, Y](f) = \left(X^1 \frac{\partial}{\partial x} + X^2 \frac{\partial}{\partial y} \right) \frac{\partial f}{\partial x} - \frac{\partial}{\partial x} \left(X^1 \frac{\partial f}{\partial x} + X^2 \frac{\partial f}{\partial y} \right) = -\frac{\partial X^1}{\partial x} \frac{\partial f}{\partial x} - \frac{\partial X^2}{\partial x} \frac{\partial f}{\partial y}.$$

  In order for this to be 0 for all f, it is necessary and sufficient that $\frac{\partial X^1}{\partial x} = \frac{\partial X^1}{\partial x} = 0$. Hence, for a vector field X to be invariant under translations in the x-direction, it is necessary and sufficient for its coefficients to be constant in x.
- *Rotation*: the flow $R_t(x, y) = (x \cos t - y \sin t, x \sin t + y \cos t)$ is generated by the vector field $Y = x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x}$. A vector field X is invariant under rotations (under the flow R) if and only if $[X, Y] = 0$. Here it will be convenient to work in polar coordinates (r, θ) (on the plane minus the negative x-axis, for example). Here $Y = \frac{\partial}{\partial \theta}$, and so expanding $X = X^1 \frac{\partial}{\partial r} + X^2 \frac{\partial}{\partial \theta}$, we have exactly the same calculation as above: $[X, Y] = 0$ iff X^1 and X^2 are constant functions of θ. That is: a vector field is rotationally-invariant iff its components depend only on the radial variable.
- *Dilation*: the Euler vector field $E = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$ generates the flow $\phi_t(p) = e^t p$. Here again, it will pay to convert to polar coordinates, where $E = \frac{\partial}{\partial r}$. Again by the same argument as above, it follows that a vector field X is invariant under the dilation flow ϕ_t if and only if $[E, X] = 0$, which means X (written in polar coordinates) has coefficients that only depend on θ, not on r.

The clearest and deepest way to understand what it means for two vector fields to commute is to say that it means their flows commute: if θ is the flow of X and ϕ is the flow of Y, then $[X, Y] = 0$ if and only if $\theta_t \circ \phi_s = \phi_s \circ \theta_t$ for all s, t. However, in general this is quite trickly to really make sense of: if X, Y are not complete, then the domains of the two flows may not be very compatible, and the näive guess for what it really means for them to commute turns out to be wrong: it is possible for $[X, Y] = 0$ and yet $\theta_t \circ \phi_s(p) \neq \phi_s \circ \theta_t(p)$ for some point p and times t, s for which both sides are defined. (The trouble is that $\mathscr{D}^p(\theta)$ and $\mathscr{D}^p(\phi)$ must be intervals containing 0, but when composing this recenters and one can be looking at the "wrong" integral curve – one that doesn't go through 0.) It is quite annoying to state and prove the right general theorem; we will content ourselves with the statement for complete vector fields presently.

THEOREM 5.31. *Let $X, Y \in \mathscr{X}(M)$ be complete vector fields. Then $[X, Y] = 0$ iff their (global) flows θ, ϕ commute in the sense that $\theta_t \circ \phi_s = \phi_s \circ \theta_t$ for all $s, t \in \mathbb{R}$.*

PROOF. First, assume $[Y, X] = \mathscr{L}_Y(X) = 0$. By Theorem 5.29, this means that X is invariant under ϕ, which means precisely that, for each $s \in \mathbb{R}$, $(\phi_s)_*(X) = X$. Now, by Corollary 5.18, the flow of $(\phi_s)_*(X)$ is $\phi_s \circ \theta_t \circ \phi_s^{-1}$; but since $(\phi_s)_*(X) = X$, this means that $\phi_s \circ \theta_t \circ \phi_s^{-1} = \theta_t$. This shows that θ and ϕ commute.

Conversely, suppose the flows commute. This can be written in the form

$$\phi^{\theta_t(p)}(s) = \theta_t(\phi^p(s)).$$

Now differentiate both sides with respect to s at $s = 0$. Since $\phi^q(s)$ is an integral curve, the left-hand side is $Y_{\theta_t(p)}$. The right-hand side becomes

$$\left.\frac{d}{ds}\right|_{s=0} \theta_t(\phi^p(s)) = d(\theta_t)_p(\dot{\phi}^p(0)) = d(\theta_t)_p(Y_p).$$

So $Y_{\theta_t(p)} = d(\theta_t)_p(Y_p)$, which means that $(\theta_t)_*(Y) = Y$ holds for all t. Replacing t with $-t$, this means that $\mathscr{L}_Y(X) = \left.\frac{d}{dt}\right|_{t=0} (\theta_{-t})_*(Y) = \left.\frac{d}{dt}\right|_{t=0} Y = 0$, as desired. $\square$

CHAPTER 6

The Cotangent Bundle and 1-Forms

We have thus far failed to mention what are arguably the most important vector fields from classical vector calculus: *conservative vector fields*, which have the form ∇f for some smooth function f. The reason they have not come up yet is because — spoiler alert—*they are not vector fields at all*. To be clear about what this means, let's take an example: let $M = \mathbb{R}^2_+$ be the right half-plane, $\{(x, y) \in \mathbb{R}^2 \colon x > 0\}$. Let $f \in C^\infty(M)$, and consider the vector field ∇f. In our present notation, this should be the section of TM whose expression in the (global) local coordinates (x, y) is given by

$$\nabla f|_p = \frac{\partial f}{\partial x}(\hat{p})\left.\frac{\partial}{\partial x}\right|_p + \frac{\partial f}{\partial y}(\hat{p})\left.\frac{\partial}{\partial y}\right|_p \tag{6.1}$$

where $\hat{p} = (x, y)$ is the coordinate expression for the point p in Cartesian coordinates. Now, if ∇f is really a vector field on the manifold M, this means (cf. Section 3) that we must have the same expression in *any* coordinates. For example, it must also be true that, using polar coordinates (r, θ) on M (which are also globally defined),

$$\nabla f|_p = \frac{\partial f}{\partial r}(\hat{p})\left.\frac{\partial}{\partial r}\right|_p + \frac{\partial f}{\partial \theta}(\hat{p})\left.\frac{\partial}{\partial \theta}\right|_p \tag{6.2}$$

where now $\hat{p}$ denotes the coordinate expression for p in polar coordinates. But we can check if this is really the case, by using the transformation law (again cf. Section 3). In particular, we have

$$\begin{aligned}
\frac{\partial}{\partial x} &= \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta} \\
\frac{\partial}{\partial y} &= \frac{\partial r}{\partial y}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y}\frac{\partial}{\partial \theta} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}
\end{aligned}$$

Thus, from (6.1), we have

$$\begin{aligned}
\nabla f|_p &= \frac{\partial f}{\partial x}\left(\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\right) + \frac{\partial f}{\partial y}\left(\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}\right) \\
&= \left(\cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}\right)\frac{\partial}{\partial r} + \frac{1}{r}\left(-\sin\theta\frac{\partial f}{\partial x} + \cos\theta\frac{\partial f}{\partial y}\right)\frac{\partial}{\partial \theta}
\end{aligned}$$

(where we implicitly write $\frac{\partial f}{\partial x}$ to mean $\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)$ here, and similarly with $\frac{\partial f}{\partial y}$.) Now we can also transform the derivatives in the coefficients to express them in terms of the vector fields $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \theta}$, in the same manner:

$$\begin{aligned}
\cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y} &= \cos\theta\left(\cos\theta\frac{\partial f}{\partial r} - \frac{\sin\theta}{r}\frac{\partial f}{\partial \theta}\right) + \sin\theta\left(\sin\theta\frac{\partial f}{\partial r} + \frac{\cos\theta}{r}\frac{\partial f}{\partial \theta}\right) \\
&= (\cos^2\theta + \sin^2\theta)\frac{\partial f}{\partial r} + \frac{1}{r}(-\cos\theta\sin\theta + \sin\theta\cos\theta)\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial r},
\end{aligned}$$

while

$$-\sin\theta\frac{\partial f}{\partial x}+\cos\theta\frac{\partial f}{\partial y}=-\sin\theta\left(\cos\theta\frac{\partial f}{\partial r}-\frac{\sin\theta}{r}\frac{\partial f}{\partial\theta}\right)+\cos\theta\left(\sin\theta\frac{\partial f}{\partial r}+\frac{\cos\theta}{r}\frac{\partial f}{\partial\theta}\right)$$
$$=(-\sin\theta\cos\theta+\cos\theta\sin\theta)\frac{\partial f}{\partial r}+\frac{1}{r}(\sin^2\theta+\cos^2\theta)\frac{\partial f}{\partial\theta}=\frac{1}{r}\frac{\partial f}{\partial\theta}.$$

Now combining, this shows that the gradient field ∇f, defined in Cartesian coordinates by (6.1), converted to polar coordinates, has the form

$$\nabla f|_p=\frac{\partial f}{\partial r}(\hat{p})\left.\frac{\partial}{\partial r}\right|_p+\frac{1}{r^2}\frac{\partial f}{\partial\theta}(\hat{p})\left.\frac{\partial}{\partial\theta}\right|_p.$$

This does *not* match (6.2) (there is a factor of $\frac{1}{r^2}$ discrepancy in the $\frac{\partial}{\partial\theta}$ term).

Thus, the gradient, viewed as a vector field in Cartesian coordinates, is most definitely coordinate dependent: it does not generalize to well-defined map from $M\to TM$. In older language, we would say it is not invariant, or coordinate-dependent. There was a strong clue that this would happen: in (6.1), there are $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ terms repeated; we know that, in our transformation laws (really just the chain rule), $\frac{\partial}{\partial x}$ should come with $\frac{\partial x}{\partial u}$ for some new coordinate (function) u; not with $\frac{\partial f}{\partial x}$.

This is not a failing, however. Instead, it introduces us to a new kind of vector field – one that transforms in the opposite way, with the $\frac{\partial}{\partial x}$ on the bottom – meaning that it transforms "in the same direction" as the derivatives. Such vector fields will be called *covariant* (transform the same direction) as opposed to *contravariant* (the vector fields we've been working with thus far are called contravariant in older literature). Note: these terms are *unrelated* to the words covariant and contravariant as they are used in category theory, so beware!

As we'll see, covariant vector fields are actually more natural and fundamental in many respects.

1. Dual Spaces and Cotangent Vectors

Let V be a finite-dimensional real vector space. The **dual space** V^* is defined to be $V^*=L(V,\mathbb{R})$, the set of linear maps from V to $\mathbb{R}$. This is a vector space in its own right, under the usual operations. It is, in fact, isomorphic to V. To see this, select a basis $\{e_1,\dots,e_n\}$ for V. The associated **dual vectors** $\{e_1^*,\dots,e_n^*\}$ in V^* are defined by

$$e_j^*\left(v^1e_1+\cdots v^ne_n\right)=v^j.$$

I.e. e_j^* are the linear extensions of $e_j^*(e_i)=\delta_j^i$. These are certainly linear functionals on V. They form a basis:

- They are linearly independent: if $\sum_j\alpha_je_j^*=0$ (as a linear functional) this means that $\sum_j\alpha_je_j^*(v)=0$ for all $v\in V$; in particular this means that $0=\sum_j\alpha_je_j^*(e_i)=\alpha_i$ for each i, so $\alpha_1=\cdots=\alpha_n=0$.
- They span V^*: if $\lambda\in V^*$, define $\omega=\lambda(e_1)e_1^*+\cdots+\lambda(e_n)e_n^*$. We compute for any $v=\sum_j v^je_j\in V$ that $\omega(v)=\sum_j\lambda(e_j)e_j^*(v)=\sum_j v^j\lambda(e_j)=\lambda(\sum_j v^je_j)=\lambda(v)$. Thus $\lambda=\omega\in\operatorname{span}\{e_1^*,\dots,e_n^*\}$.

Thus, $\dim V^*=\dim V$, and so V and V^* are isomorphic as vector spaces. Indeed, we could define an isomorphism $V\to V^*$ to be the linear extension of $e_j\mapsto e_j^*$, $1\le j\le n$.

There's nothing wrong with this, but it is not *natural*, which we can make sense of simply here by saying that it is *basis dependent*. Indeed, define this isomorphism $\phi\colon V \to V^*$ by $\phi(e_j) = e_j^*$ for the given chosen basis. Let's write this in terms of a different basis $\{f_1 \dots, f_n\}$ for V. Since it is also a basis, there is an invertible $n \times n$ matrix A so that $f_j = \sum_i A_j^i e_i$ for all j. Then we have

$$\phi(f_j) = \sum_i A_j^i \phi(e_i) = \sum_i A_j^i e_i^*. \tag{6.3}$$

We might hope that this is equal to f_j^*. To check this, we must express the e_i in terms of the f_j, which means taking the inverse matrix $B = A^{-1}$, so that $e_j = \sum_i B_j^i f_i$. Then we can write f_j^* in the basis $\{e_j^*\}$ as follows:

$$f_j^*(v) = f_j^*\left(\sum_i v^i e_i\right) = f_j^*\left(\sum_i v^i \sum_k B_i^k f_k\right) = \sum_{i,k} v^i B_i^k f_j^*(f_k) = \sum_i v^i B_i^j = \sum_i B_i^j e_i^*(v)$$

which shows that

$$f_j^* = \sum_i B_i^j e_i^* = \sum_i [B^\top]_j^i e_i^*$$

where $B^\top$ is the transpose of B (transposing rows and columns). Thus, we see that (6.3) does not in general simplify to $\phi(f_j) = f_j^*$ – this is only the case when the change of basis matrix happens to satisfy $A = B^\top = (A^{-1})^\top$.

REMARK 6.1. Note this says that if we restrict our basis-changes to *orthogonal transformations* – i.e. where the change of basis matrix A is orthogonal, $A^\top A = I$ – then this isomorphism *is* natural. In fact, if we fix an inner product $\langle \cdot, \cdot \rangle$ on V to begin with, and insist that all our bases be *orthonormal* bases with respect to the inner product, then the change of basis matrix will indeed always be orthogonal. So we see that, in the category of finite-dimensional inner product spaces, V and V^* really are canonically isomorphic. It is easy to check that the isomorphism $e_j \mapsto e_j^*$ in this restricted setting has the basis independent representation

$$v \mapsto \langle \cdot, v \rangle.$$

REMARK 6.2. The inspired reader might see similarities between the above calculations and the ones we did changing the gradient from Cartesian to polar coordinates. This is no accident, as we will soon explain. Let us note that the Jacobian of the polar coordinate map $(r, \theta) \mapsto (x, y)$ is *not* orthogonal: although its columns are orthogonal vectors, the second one is not unit length, but rather of length r. This is why the $\frac{1}{r^2}$ comes up in the calculation there. In fact, if one were to choose new coordinates (u, v) for which the Jacobian of the transformation $(u, v) \mapsto (x, y)$ is orthogonal at every point, then the expression for the gradient in the new coordinates would indeed still be $\frac{\partial f}{\partial u}\frac{\partial}{\partial u} + \frac{\partial f}{\partial v}\frac{\partial}{\partial v}$. This turns out to be a very strong condition: it only happens for global isometries: maps T for which $|T(x)| = |x|$ for all x. The set of isometries of $\mathbb{R}^n$ consists of the *Euclidean group*: transformations that are a composition of a translations, rotations, and reflections. (In terms of a basis, this just means $T(x) = Qx + v$ for some orthogonal matrix Q and some vector v.) These are the only coordinate transformations that preserve the expression for the gradient.

Note that the transpose of the change-of-basis matrix came into play above. In fact, the transpose of a matrix is closely related to the dual space. Let $T\colon V \to W$ be a linear map. Then there is an induced **dual map** $T^*\colon W^* \to V^*$, which is simply defined by

$$(T^*\lambda)(v) = \lambda(Av).$$

It is easy to verify that, for any $\lambda \in W^*$, $T^*\lambda \in V^*$, so this is well-defined, and moreover T^* is a linear map. We record some key properties of it in the next lemma, which is left as an exercise to prove.

LEMMA 6.3. *Let V, W be finite-dimensional real vector spaces, and let $S, T\colon V \to W$ be linear maps. Then we have the following.*

(a) $(\mathrm{Id}_V)^* = \mathrm{Id}_{V^*}$.
(b) $(S \circ T)^* = T^* \circ S^*$.
(c) *If $\{v_j\}$ is a basis for V and $\{w_j\}$ is a basis for W, and A is the matrix of T in terms of these bases, then the matrix of A^* in terms of the dual bases $\{v_j^*\}$ and $\{w_j^*\}$ is $A^\top$.*

In fact, the dual map T^* is often called the *transpose* of T. This is a nice way to see that transpose is a natural construction: it commutes with changing bases, because it is really given by a basis-free object.

Nevertheless, even though there is a natural transpose map, it is still the case (as we saw above) that the isomorphism $e_j \mapsto e_j^*$ is basis dependent. In fact, it is a theorem that there does not exist a basis-independent isomorphism. (Also: the arguments above all fail if V is not finite-dimensional. Although the map $e_j \mapsto e_j^*$ is still an injective linear map $V \to V^*$ in that case, it is never surjective. But that won't bother us presently.)

Now, we can repeat the process: having constructed V^*, we may take *its* dual space $(V^*)^* = V^{**}$, the space of linear functionals $L(V^*, \mathbb{R})$. Again, there will be no basis-independent isomorphism $V^* \to V^{**}$. Somewhat miraculously, though, we may compose two basis-*dependent* isomorphisms, and get a basis-*independent* one.

LEMMA 6.4. *The canonical map $\xi\colon V \to V^{**}$ defined by*

$$\xi v(\lambda) = \lambda(v)$$

is an isomorphism.

PROOF. Fix a basis $\{e_j\}$ of V, and its dual basis $\{e_j^*\}$ of V^*. Then we know $\alpha\colon V \to V^*$ defined by $\alpha(e_j) = e_j^*$ is an isomorphism. Similarly, we define $\beta\colon V^* \to V^{**}$ by $\beta(e_j^*) = e_j^{**}$ (using the dual basis to $\{e_j^*\}$), and this is also an isomorphism. Then $\beta \circ \alpha\colon V \to V^{**}$ is an isomorphism. We claim that $\xi = \beta \circ \alpha$, and is therefore an isomorphism.

To see this, first note that ξ is linear (by elementary computation), so it suffices to check that ξ and $\beta \circ \alpha$ agree on a basis of V; we will, of course, use the basis $\{e_j\}$. On the one hand, we have for any $\lambda \in V^*$

$$\xi(e_j)(\lambda) = \lambda(e_j).$$

On the other hand,

$$\beta \circ \alpha(e_j)(\lambda) = \beta(e_j^*)(\lambda) = e_j^{**}(\lambda).$$

Now expanding $\lambda = \sum_i \lambda(e_i) e_i^*$, we therefore have

$$\beta \circ \alpha(e_j)(\lambda) = e_j^{**}\left(\sum_i \lambda(e_i) e_i^*\right) = \sum_i \lambda(e_i)\delta_j^i = \lambda(e_j).$$

□

Thus ξ is an isomorphism. Although our proof used a basis, the definition of ξ is manifestly basis-independent. (Note: in the infinite-dimensional case, ξ is no longer an isomorphism, but it is still injective; it is called the *canonical embedding* of V into V^{**}.)

Now, let M be a smooth manifold, and let $p \in M$. The **cotangent space** at p, denoted T_p^*M, is the dual to the tangent space:

$$T_p^*M \equiv (T_pM)^*.$$

We refer to elements of T_p^*M as **cotangent vectors**, and frequently use lower-case greek letters like $\omega_p, \lambda_p \in T_p^*M$. Fix a chart (U, φ) at p, with coordinate functions $\varphi = (x^1, \ldots, x^n)$. Then the coordinate vectors $\{\frac{\partial}{\partial x^j}\big|_p\}_{1\le j\le n}$ form a basis for T_pM. To avoid messy notation, for now let us refer to the dual basis as $\{\lambda_j\big|_p\}_{1\le j\le n}$. From the proof above that the dual basis spans the dual space, we can then express *any* cotangent vector $\omega_p \in T_p^*M$ as

$$\omega_p = \sum_{j=1}^n \omega_p^j \, \lambda_j\big|_p, \quad \text{where} \quad \omega_p^j = \omega_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right).$$

Now, what happens if we change coordinates? Let (V, ψ) be another chart at p with coordinate functions $\psi = (y^1, \ldots, y^n)$. Denote the dual basis to $\{\frac{\partial}{\partial y^j}\big|_p\}_{1\le j\le n}$ as $\{\mu_j\big|_p\}_{1\le j\le n}$. We want to compute the components $\tilde{\omega}_p^j$ of ω in terms of this new dual basis. As above, these coefficients are given simply by

$$\tilde{\omega}_p^j = \omega_p\left(\frac{\partial}{\partial y^j}\bigg|_p\right).$$

To relate the ω_p^j to the $\tilde{\omega}_p^j$, first convert the coordinate vectors:

$$\frac{\partial}{\partial x^j}\bigg|_p = \sum_{i=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) \, \frac{\partial}{\partial y^i}\bigg|_p.$$

Hence, we have

$$\omega_p^j = \omega_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right) = \omega_p\left(\sum_{i=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) \, \frac{\partial}{\partial y^i}\bigg|_p\right) = \sum_{i=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p})\tilde{\omega}_p^i. \tag{6.4}$$

Now, compare this to how the components of vector change under the same coordinate change: if $T_pM \ni X_p = \sum_{j=1}^n X_p^j \, \frac{\partial}{\partial x^j}\big|_p$, then we have

$$X_p = \sum_{j=1}^n X_p^j \left(\sum_{i=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) \, \frac{\partial}{\partial y^i}\bigg|_p\right) = \sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) X_p^j\right) \frac{\partial}{\partial y^i}\bigg|_p$$

which is to say that the components $\tilde{X}_p^i$ of the vector X_p in terms of the new coordinate basis are given by

$$\tilde{X}_p^i = \sum_{j=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) X_p^j. \tag{6.5}$$

Let us restate (6.4) and (6.5) for direct comparison in the following proposition.

PROPOSITION 6.5. *Let M be a smooth manifold and $p \in M$. Fix two charts $(U, \varphi = (x^j)_{j=1}^n)$ and $(V, \psi = (y^i)_{i=1}^n)$ at p. Let $X_p \in T_pM$ and $\omega_p \in T_p^*M$. If X has coordinates X_p^j in terms of the φ coordinate basis, and ω_p has coordinates ω_p^j in terms of its dual basis, then the components $\tilde{X}_p^i$*

of X_p in terms of the ψ coordinate basis, and the components $\tilde{\omega}_p^i$ of ω_p in terms of its dual basis, are given by

$$\tilde{X}_p^i = \sum_{j=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) X_p^j, \qquad \omega_p^j = \sum_{i=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) \tilde{\omega}_p^i.$$

Thus, the two transform opposite to each other. Indeed, we could restate the ω coordinate transformation as follows: the matrix $\frac{\partial y^i}{\partial x^j}(\hat{p})$ is just the Jacobian of the transition map $\psi \circ \phi^{-1}$ at the point $\varphi(p)$. Its inverse is the Jacobian of the inverse map at $\psi(p)$ (which we also, somewhat confusingly in this case, denote $\hat{p}$), and so we can write the transformation law for the components of ω_p in the form

$$\tilde{\omega}_p^i = \sum_{j=1}^n \frac{\partial x^j}{\partial y^i}(\hat{p}) \omega_p^j.$$

Now the ∂x^js are on top. This is the key to understanding the invariant generalization of a gradient.

2. The Differential, Reinterpreted

Let $f \in C^\infty(M)$, fix $p \in M$, and let $q = f(p)$. The differential map df_p is a linear map $T_pM \to T_q\mathbb{R}$. Of course, the tangent space $T_q\mathbb{R}$ can be identified with $\mathbb{R}$ itself, in a natural way. In terms of the näive definition of $T_q\mathbb{R} = \{(q, v) \colon v \in \mathbb{R}\}$, we simply identify $(q, v) \cong v$. In our more invariant language, what does this mean? To see the answer, we work in local (global) coordinates: fix the usual Cartesian coordinate x on $\mathbb{R}$, also known as the identity map $\mathrm{Id}_\mathbb{R}$. Then any element $X_q \in T_q\mathbb{R}$ can be written uniquely as

$$X_q = X_q^1 \left.\frac{\partial}{\partial x}\right|_q$$

for some coefficient $X_q^1 \in \mathbb{R}$. Of course, the identification we seek is $X_q \cong X_q^1$. What does this mean on the invariant level? Evidently,

$$X_q^1 = X_q^1 \left.\frac{\partial}{\partial x}\right|_q (x) = X_q(x) = X_q(\mathrm{Id}_\mathbb{R}).$$

Because of the 1-dimensionality, the map $T_q\mathbb{R} \to \mathbb{R}$ given by $X \mapsto X(\mathrm{Id}_\mathbb{R})$ is an isomorphism. Hence, we can think of df_q as a map $T_qM \to \mathbb{R}$, so long as we post compose with this isomorphism. This gives us a new interpretation of the differential, which for the moment we call $\widetilde{df_p}$:

$$\widetilde{df_p}(X_p) = df_p(X_p)(\mathrm{Id}_\mathbb{R}) = X_p(\mathrm{Id}_\mathbb{R} \circ f) = X_p(f). \tag{6.6}$$

We now immediately drop the $\tilde{}$ and refer to this *also* as the differential, keeping in mind that, although it is a different object, it is the same as the old differential modulo the identification $T_q\mathbb{R} \cong \mathbb{R}$. And note, what kind of object is it? df_p is a linear map $T_pM \to \mathbb{R}$: it is a cotangent vector.

DEFINITION 6.6. *Let M be a smooth manifold, $p \in M$, and $f \in C^\infty(M)$. The* **differential of** *f at p, $df_p \in T_p^*M$ is the cotangent vector defined by*

$$df_p(X_p) = X_p(f).$$

Note that, in our invariant language, T_pM consists of derivations at p: a certain class of linear functionals on the vector space $C^\infty(M)$. That is: T_pM is a subspace of $C^\infty(M)^*$. Thus, the definition above is akin to the canonical embedding $\xi\colon C^\infty(M) \to C^\infty(M)^{**}$. We have to be a little careful, since we are mapping $C^\infty(M)$ to the dual space of a *subspace* of $C^\infty(M)^*$; as such, it is not actually one-to-one.

In local coordinates $(U, \varphi = (x^j)_{j=1}^n)$ at $p \in M$, the coordinate functions x^j are smooth, and so they have differentials dx_p^j. We can express them in terms of the dual basis to the coordinate vectors $\frac{\partial}{\partial x^i}\big|_p$:

$$dx_p^j\left(\frac{\partial}{\partial x^i}\bigg|_p\right) = \frac{\partial}{\partial x^i}\bigg|_p (x^j) = \delta_j^i.$$

This is precisely to say that

$$dx_p^j = \left(\frac{\partial}{\partial x^i}\bigg|_p\right)^*.$$

I.e. the dual basis to the coordinate vectors is $\{dx_p^1, \dots, dx_p^n\}$. Thus, we can locally express any covariant vector $\omega_p \in T_p^*M$ as

$$\omega_p = \sum_{j=1}^n \omega_p^j \, dx_p^j.$$

Using this language, we can restate Proposition 6.5 as follows: the transformation laws for contravariant and covariant vector fields are mediated by the following formulas for changing coordinate vectors:

$$\frac{\partial}{\partial x^j}\bigg|_p = \sum_{i=1}^n \frac{\partial y^i}{\partial x^j}(\hat{p}) \, \frac{\partial}{\partial y^i}\bigg|_p, \qquad dx_p^j = \sum_{i=1}^n \frac{\partial x^j}{\partial y^i}(\hat{p}) \, dy_p^i. \tag{6.7}$$

3. The Cotangent Bundle, and Covariant Vector Fields

We now want to glue together covariant vectors ω_p at different points into a continuous object. The procedure will mirror what we did in Sections 5, 1, and 3 quite closely, and so we will be a lot more brief with the exposition here.

DEFINITION 6.7. *Let M be a smooth manifold. Its* **cotangent bundle** *T^*M is the disjoint union of all cotangent spaces*

$$T^*M \equiv \bigsqcup_{p \in M} T_p^*M.$$

There is a natural projections $\pi\colon T^*M \to M$ given by $\pi(\omega_p) = p$ for any $\omega_p \in T_p^*M$.

As with the tangent bundle, we can imbue T^*M with the structure of a smooth $2n$-dimensional manifold in essentially the same way: given a chart (U, φ) for M, we define a chart $(\widetilde{U}, \widetilde{\varphi})$ for T^*M as follows: $\widetilde{U} = \pi^{-1}(U)$, and if the coordinate functions of φ are $(x^1, \dots, x^n)$, then we define $\widetilde{\varphi}$ on any covariant vector $\omega_p = \sum_{j=1}^n \omega_p^j \, dx_p^j$ by

$$\widetilde{\varphi}\left(\sum_{j=1}^n \omega_p^j \, dx_p^j\right) = (x^1(p), \dots, x^n(p), \omega_p^1, \dots, \omega_p^n) = (\varphi(p), \omega_p^1, \dots, \omega_p^n).$$

The proof that these charts define a smooth structure on T^*M is almost identical to the proof of Proposition 3.25 (the corresponding statement for the tangent bundle), simply using the transformation law (6.7) for covariant vector fields (instead of contravariant vector fields) to show that the transition maps are smooth. We leave the details to the reader.

REMARK 6.8. As differentiable manifolds, TM and T^*M are "the same": they are diffeomorphic. We do not have the tools needed yet to prove this, but the idea is just a global version of the proof that $V \cong V^*$. As in that case, there isn't a "natural" diffeomorphism, unless extra structure is present. Following Remark 6.1, if we endow V with a fixed inner product, then there is a natural isomorphism $V \mapsto V^*$ given by $v \mapsto \langle \cdot, v \rangle$. Similarly, suppose that, for each p, there is an inner product $g_p \colon T_pM \times T_pM \to \mathbb{R}$, which varies smoothly in p (meaning that the map $p \mapsto g_p(X_p, Y_p)$ is smooth for any smooth vector fields $X, Y \in \mathscr{X}(M)$). Such a function $g \colon \mathscr{X}(M) \times \mathscr{X}(M) \to \mathbb{R}$ is called a **Riemannian metric**. If such a g exist (which is true on any smooth manifold, though we won't prove that presently), then it yields a map $TM \to T^*M$ defined by $X_p \mapsto g_p(\cdot, X_p)$ which is easily seen to be a bijection, and the smoothness of g makes it into a diffeomorphism. In fact, it is more: it is a generalization of what we called a global trivialization, or a *bundle isomorphism*. The map commutes with the projections, and its restriction to each fibre is a linear isomorphism.

It is important to note that, although TM and T^*M is diffeomorphic, even in a strong bundle sense, they are still different objects. And, as with V and V^*, they are not *naturally* isomorphic. As we will shortly see, the cotangent bundle is fundamentally the better object to work with.

A **section** of T^*M is a continuous function $\omega \colon M \to T^*M$ with the property that $\pi \circ \omega = \mathrm{Id}_M$: i.e. it is a continuous choice of a covariant vector at each point. We call such sections (rough) **contravariant vector fields**. As with contravariant vector fields, covariant vector fields are a module over $C^\infty(M)$ (in fact over $\mathrm{Fun}(M, \mathbb{R})$, where $(f\omega)_p(X_p) = f(p)\omega_p(X_p)$ and $(\omega + \lambda)_p = \omega_p + \lambda_p$.

Since T^*M is a smooth manifold, we can insist that a section be smooth.

DEFINITION 6.9. *The set of smooth covariant vector fields on M is denoted by $\mathscr{X}^*(M)$, or alternatively by $\Omega^1(M)$. The second notation goes along with the more common name for such vector fields:* **differential** 1**-forms***, or simply* 1**-forms***.*

Given a chart $(U, \varphi = (x^j)_{j=1}^n)$ in M, the **coordinate** 1**-forms** *are $dx^1, \ldots, dx^n$ (i.e. the sections $p \mapsto dx^j_p$); it is immediate from the definition of the smooth structure of T^*M that these are smoooth.*

In general, any (rough) covariant vector field $\omega \in \mathscr{X}^*(M)$ can be expressed locally as

$$\omega|_U = \sum_{j=1}^n \omega_j \, dx^j$$

where $\omega_j \colon M \to \mathbb{R}$ are the functions $\omega_j(p) = \omega_p(\frac{\partial}{\partial x^j}\big|_p)$.

Given a (rough) covariant vector field ω, we can use it to eat a (rough) vector field X to get a function on M:

$$\omega(X) \colon M \to \mathbb{R}, \quad \omega(X)(p) = \omega_p(X_p).$$

As before, this allows us to think of covariant vector fields as certain kinds of functions. While a contravariant vector field $X \in \mathscr{X}(M)$ can be identified with a derivation $X \colon C^\infty(M) \to C^\infty(M)$, a covariant vector field $\omega \in \mathscr{X}^*(M)$ can be identified with a linear map $\omega \colon \mathscr{X}(M) \to \mathrm{Fun}(M, \mathbb{R})$, given by $(\omega(X))(p) = \omega_p(X_p)$ as above. In local coordinates, if we express $\omega = \sum_{j=1}^n \omega_j \, dx^j$ and

$X = \sum_{i=1}^n X^i \frac{\partial}{\partial x^i}$, then since $dx_p^j = (\frac{\partial}{\partial x^j}\big|_p)^*$ we have $\omega(X) = \sum_{j=1}^n \omega_j X^j$. This shows that, if the components of ω and X are smooth, then so is the function $\omega(X)$.

We can now characterize smoothness of a covariant vector field in terms of its coefficients, precisely mirroring Proposition 4.12 (the analogous statement for contraviariant vector fields). The proof is very similar, and is left as a homework exercise.

PROPOSITION 6.10. *Let M be a smooth manifold, and let $\omega\colon M \to T^*M$ be a (rough) covariant vector field. TFAE:*

(a) *ω is smooth.*
(b) *The component functions of ω in any chart are smooth.*
(c) *For every smooth contravariant vector field $X \in \mathscr{X}(M)$, the function $\omega(X)$ is smooth.*
(d) *Given any open subset $U \subseteq M$ and $X \in \mathscr{X}(U)$, the function $\omega(X)\colon U \to \mathbb{R}$ is smooth.*

The most important examples of smooth covariant vector fields are **gradient fields**, aka **gradient 1-forms**. Here we finally have the correct invariant generalization of the gradient from vector calculus.

DEFINITION 6.11. *Let $f \in C^\infty(M)$. Following Definition 6.6, we have a covariant vector $df_p \in T_p^*M$ for each p. The* **differential of** *f df is the covariant vector field defined by $df(p) = df_p$. Thinking of df as a linear function $\mathscr{X}(M) \to \mathrm{Fun}(M, \mathbb{R})$, its action is then $df(X) = X(f)$. If X is smooth, then $X(f)$ is smooth, which means that $df \in \Omega^1(M)$: it is a smooth covariant vector field.*

Again, let us connect df with the earlier notation, where $df\colon TM \to T\mathbb{R}$ is the map $df(X_p) = df_p(X_p)$. That is: the differential of the map $f\colon M \to \mathbb{R}$ is the linear map defined by $(df(X_p))(g) = X_p(g \circ f)$. Now we think of $df(X_p)$ as a real number rather than a vector in $\mathbb{R}$ (which is just a real number), and the way to do this is to evaluate the derivation at $\mathrm{Id}_\mathbb{R}$. Once again, that connects the two notations here: we have

$$df^{\text{new}}(X) = X(f) = X(\mathrm{Id}_R \circ f) = (df^{\text{old}}(X))(\mathrm{Id}_\mathbb{R}).$$

Referring to Definition 6.9, in local coordinates we can write

$$df = \sum_{j=1}^n (df)_j dx^j, \quad \text{where} \quad (df)_j(p) = df_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right) = \frac{\partial f}{\partial x^j}(\hat{p}),$$

meaning that

$$df = \sum_{j=1}^n \frac{\partial f}{\partial x^j}\, dx^j.$$

Here are some elementary properties of gradient 1-forms, whose proofs are left as exercises.

PROPOSITION 6.12. *Let M be a smooth manifold, and let $f, g \in C^\infty(M)$.*

(a) *For constants $a, b \in \mathbb{R}$, $d(af + bg) = a\, df + b\, dg$.*
(b) *$d(fg) = f\, dg + g\, df$.*
(c) *If $\mathrm{Im}(f) \subset (a, b)$ and $h\colon (a, b) \to \mathbb{R}$ is smooth, then $d(h \circ f) = (h' \circ f)\, df$.*
(d) *If f is constant, then $df = 0$.*

The converse of item (d) above is true, and it one important application of gradient 1-forms.

PROPOSITION 6.13. *Suppose $f \in C^\infty(M)$. Then $df = 0$ if and only if f is locally constant: it is constant on each component of M.*

PROOF. We will assume f is connected; then the statement is $df = 0$ iff f is constant. Proposition 6.12(d) shows the "if" part of this equivalence, so we now treat the "only if" part. Suppose $df = 0$. Fix a point $p \in M$, and let $\mathscr{C} = f^{-1}(f(p)) = \{q \in M \colon f(q) = f(p)\}$, which is closed by the continuity of f. For any point $q \in \mathscr{C}$, let $(U, \varphi = (x^j)_{j=1}^n)$ be a chart at q, and let $r \in U$. In this chart, $df = \sum_{j=1}^n \frac{\partial f}{\partial x^j}\, dx^j$. As $\{dx_r^j\}_{j=1}^n$ form a basis of T_r^*M, in order for $df_r = 0$, it therefore follows that $\frac{\partial f}{\partial x^j}(\hat{r}) = 0$ for $1 \le j \le n$. Thus, $\frac{\partial f}{\partial x^j} \equiv 0$ on U for all j; by elementary calculus, this means f is constant on U. This shows that $\mathscr{C}$ is open. Since M is connected, it follows that $\mathscr{C} = M$, concluding the proof. □

4. Pullbacks of Covariant Vector Fields

Let M, N be smooth manifolds, and $F \colon M \to N$ a smooth map. Given a point $p \in M$, the (ordinary) differential map dF_p is a linear transformation $T_pM \to T_{F(p)}N$. Thus, there is a dual (transpose) map

$$(dF_p)^* \colon T^*_{F(p)}N \to T^*_pM.$$

This is called the **pullback** by F at p, or the **cotangent map**, or the **codifferential**. By the definition of the transpose, its action is simply

$$[(dF_p)^*(\omega_{F(p)})](X_p) = \omega_{F(p)}(dF_p(X_p)), \qquad \omega_{F(p)} \in T^*_{F(p)}N,\ X_p \in T_pM.$$

That is: we pullback a covariant vector by letting it act on the push-forward of its argument. But it turns out that this notion is better, in the following sense. Recall that vector fields generally do not push-forward: although we can push forward any given vector, we cannot do it to a whole vector field consistently unless that map F is a diffeomorphism. Not so with covariant vector fields.

DEFINITION 6.14. *Let M, N be smooth manifolds, and let $F \colon M \to N$ be a smooth map. Let $\omega \colon N \to T^*N$ be a (rough) covariant vector field. The* **pullback** *of ω along F is the (rough) covariant vector field $F^*\omega \colon M \to T^*M$ defined by*

$$(F^*\omega)_p = dF_p^*(\omega_{F(p)}).$$

I.e. the action of $F^*\omega$ is

$$(F^*\omega)_p(X_p) = \omega_{F(p)}(dF_p(X_p)).$$

Note that pullback is a linear operation: $F^*(a\omega + b\lambda) = aF^*\omega + bF^*\lambda$ for $a, b \in \mathbb{R}$.

By pulling back a 1-form (which is already a dual object), we see there is no ambiguity about which point to evaluate the argument vector field at, and hence we can pull back globally even if F is not a diffeomorphism. Of course, we should expect that if $\omega \in \mathscr{X}^*(M)$ is smooth then the pullback $F^*\omega$ is smooth. This is true. To prove it, the following computational lemma will be useful.

LEMMA 6.15. *Let M, N be smooth manifolds and $F \colon M \to N$ a smooth map. Let $u \colon N \to \mathbb{R}$ be continuous, and let $\omega \colon N \to T^*N$ be a (rough) covariant vector field on N. Then*

$$F^*(u\omega) = (u \circ F)F^*\omega. \tag{6.8}$$

In addition, if $u \in C^\infty(M)$, then

$$F^*du = d(u \circ F). \tag{6.9}$$

PROOF. We compute from the definition: for any $p \in M$,

$$\begin{aligned}(F^*(u\omega))_p = dF_p^*((u\omega)_{F(p)}) &= dF_p^*\left(u(F(p))\omega_{F(p)}\right)\\ &= u(F(p))dF_p^*(\omega_{F(p)})\\ &= (u\circ F)(p)(F^*\omega)_p\end{aligned}$$

which proves (6.8). Similarly, for any $X_p \in T_pM$,

$$\begin{aligned}(F^*du)_p(X_p) = \left(dF_p^*(du_{F(p)})\right)(X_p) &= du_{F(p)}(dF_p(X_p))\\ &= dF_p(X_p)(u)\\ &= X_p(u\circ F)\\ &= d(u\circ F)_p(X_p)\end{aligned}$$

which proves (6.9) □

PROPOSITION 6.16. *Let M, N be smooth manifolds and $F\colon M \to N$ a smooth map. If $\omega\colon N \to T^*N$ is a continuous covariant vector field, then so is $F^*\omega$. Moreover, if $\omega \in \Omega^1(N) = \mathscr{X}^*(N)$ is a smooth 1-form, then so is $F^*\omega$.*

PROOF. Fix $p \in M$, and let $q = F(p)$. Let $(V, \psi = (y^j)_{j=1}^n)$ be a chart at q, and let $U = F^{-1}(V)$ (which is an open neighborhood of p). Then we may write $\omega|_U$ in local coordinates as $\sum_{j=1}^n \omega_j\, dy^j$. Applying (6.8), we then have

$$F^*\omega = F^*\left(\sum_{j=1}^n \omega_j\, dy^j\right) = \sum_{j=1}^n F^*(\omega_j\, dy^j) = \sum_{j=1}^n (\omega_j\circ F)F^*(dy^j).$$

Now, by (6.9), $F^*(dy^j) = d(y^j\circ F)$, which are smooth gradient forms. The coefficient functions $\omega_j\circ F$ are continuous if ω is continuous, and smooth of ω is smooth. This concludes the proof. □

EXAMPLE 6.17. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ be the map $(u,v) = F(x,y,z) = (yz, e^{xy}z^2)$. Let $\omega \in \Omega^1(\mathbb{R}^2)$ be the 1-form $\omega = u\,dv + 2v\,du$. Then

$$\begin{aligned}F^*\omega = F^*(u\,dv) + 2F^*(v\,du) &= (u\circ F)d(v\circ F) + 2(v\circ F)d(u\circ F)\\ &= (yz)d(e^{xy}z^2) + 2(e^{xy}z^2)d(yz)\\ &= yz\left[e^{xy}d(z^2) + z^2d(e^{xy})\right] + 2e^{xy}x^2\left[y\,dz + z\,dy\right]\\ &= yze^{xy}\cdot 2z\,dz + z^2e^{xy}(x\,dy + y\,dx) + 2e^{xy}x^2(y\,dz + z\,dy)\\ &= yz^2e^{xy}\,dx + (xz^2 + 2x^2z)e^{xy}\,dy + 2(yz^2 + x^2y)e^{xy}\,dz.\end{aligned}$$

EXAMPLE 6.18. Let $M = \mathbb{R}^2_+$ be the right half-plane $\{(x,y) \in \mathbb{R}^2\colon x > 0\}$. Consider the smooth map $\mathrm{Id}\colon M \to M$. Let us use Cartesian coordinates on M in the codomain of Id and polar coordinates on the domain. Taking $\omega = x\,dy - y\,dx$ on M, we then have

$$\begin{aligned}\omega = x\,dy - y\,dx &= \mathrm{Id}^*(x\,dy - y\,dx)\\ &= (r\cos\theta)d(r\sin\theta) - (r\sin\theta)d(r\cos\theta)\\ &= r\cos\theta\cdot(\sin\theta\,dr + r\cos\theta\,d\theta) - r\sin\theta\cdot(\cos\theta\,dr - r\sin\theta\,d\theta)\\ &= (r\cos\theta\sin\theta - r\sin\theta\cos\theta)\,dr + (r^2\cos^2\theta + r^2\sin^2\theta)\,d\theta = r^2\,d\theta.\end{aligned}$$

Now, recall the definition of the Lie derivative of a vector field Y with respect to X:

$$\mathscr{L}_X(Y) = \frac{d}{dt}\bigg|_{t=0} (\theta_{-t})_*(Y), \quad \text{where } \theta \text{ is the flow of } X.$$

To be clear: this is a pointwise definition: $\mathscr{L}_X(Y)(p) = \frac{d}{dt}\big|_{t=0} (\theta_{-t})_*(Y)(p)$, the latter being the derivative of a map from an interval in $\mathbb{R}$ into the vector space T_pM.

This suggests a way to define the Lie derivative of a 1-form with respect to a vector field: by pulling back along θ instead of pushing forward.

DEFINITION 6.19. *Let $\omega \in \Omega^1(M)$ be a 1-form, and let $X \in \mathscr{X}(M)$ be a smooth vector field. Define a new 1-form $\mathscr{L}_X\omega$ as follows:*

$$\mathscr{L}_X\omega = \frac{d}{dt}\bigg|_{t=0} (\theta_t)^*\omega$$

where θ is the flow of X.

To show this is well-defined, we need to show that the limit exists, and indeed defines a (smooth) covariant vector field on M. To be clear, the definition is

$$\mathscr{L}_X\omega\big|_p = \frac{d}{dt}\bigg|_{t=0} \theta_t^*\omega\big|_p = \lim_{t\to 0} \frac{d(\theta_t)_p^*(\omega_{\theta_t(p)}) - \omega_p}{t}.$$

Note that it makes sense to use t rather than $-t$ since we are pulling back instead of pushing forward (either way, the difference is just a minus sign convention).

PROPOSITION 6.20. *For any vector field $X \in \mathscr{X}(M)$ and any 1-form $\omega \in \Omega^1(M)$, $\mathscr{L}_X\omega$ is a 1-form, and its action on vector fields is*

$$(\mathscr{L}_X\omega)(Y) = \mathscr{L}_X(\omega(Y)) - \omega(\mathscr{L}_X Y) = X(\omega(Y)) - \omega([X,Y]). \tag{6.10}$$

PROOF. Similar to the calculations in Proposition 5.24. We will see a more general result later, called "Cartan's magic formula", that proves this as a special case. □

Note, rearranging (6.10) yields

$$\mathscr{L}_X(\omega(Y)) = (\mathscr{L}_X\omega)(Y) + \omega(\mathscr{L}_X Y).$$

That is: $\mathscr{L}_X$ is a derivation even with respect to the "product" $\Omega^1(M) \times \mathscr{X}(M) \to C^\infty(M)$ given by $(\omega, Y) \mapsto \omega(Y)$.

COROLLARY 6.21. *If $f \in C^\infty(M)$ and $X \in \mathscr{X}(M)$, then*

$$\mathscr{L}_X(df) = d(\mathscr{L}_X f).$$

PROOF. This is just a calculation: fixing another vector field Y, from Proposition 6.20 we have

$$\begin{aligned}[\mathscr{L}_X(df)](Y) = \mathscr{L}_X(df(Y)) - df([X,Y]) &= \mathscr{L}_X(Y(f)) - [X,Y](f) \\ &= XY(f) - (XY(f) - YX(f)) \\ &= YX(f) = [d(X(f))](Y).\end{aligned}$$

Since $\mathscr{L}_X f = X(f)$, this concludes the proof. □

Again, we will see later a very general result that Lie derivatives commute with "exterior" derivatives, of which $d\colon f \mapsto df$ is a special case.

5. Line Integrals

What are 1-forms? Primary, they are precisely those objects that can be integrated over curves (in a coordinate-independent fashion). Indeed, in the calculus integral

$$\int_a^b f(t)\,dt$$

we should really think of $f(t)\,dt$ as a single object, a 1-form on $[a,b]$. (Since $[a,b]$ is not a manifold, what we really mean here is a 1-form on a slightly larger open interval restricted to $[a,b]$.) So, we take $\omega = f(t)\,dt$ for some smooth f (which is a general 1-form), and define

$$\int_{[a,b]} \omega \equiv \int_a^b f(t)\,dt.$$

This is not just a new notation. For example, it gives a very nice form for the change of variables theorem.

PROPOSITION 6.22. *Let $\omega \in \Omega^1([a,b])$. Let $\varphi\colon [c,d] \to [a,b]$ be a homeomorphism, whose restriction to (c,d) is a diffeomorphism. Then φ is either strictly increasing or strictly decreasing. Moreover, we have*

$$\int_{[c,d]} \varphi^*\omega = \pm \int_{[a,b]} \omega,$$

$+$ if φ is increasing, and $-$ if φ is decreasing.

PROOF. Denote the (global) coordinate on $[c,d]$ as s. First, any continuous bijection from $[c,d]$ to $[a,b]$ is necessarily monotone; that is a diffeomorphism means $\phi'(s) \neq 0$ for any s, and so it is strictly monotone. Now, we have

$$(\varphi^*\omega)(s) = \varphi^*(f(t)\,dt) = f \circ \varphi(s)\,d(\varphi(s)) = f \circ \varphi(s)\varphi'(s)\,ds$$

by Lemma 6.15. Now, if φ is increasing then $\varphi(c) = a$ and $\varphi(d) = b$, and so by the change of variables theorem from calculus

$$\int_{[c,d]} \varphi^*\omega = \int_c^d f(\varphi(s))\varphi'(s)\,ds = \int_a^b f(t)\,dt = \int_{[a,b]} \omega.$$

If, on the other hand, φ is decreasing, then $\varphi(c) = b$ and $\varphi(d) = a$, and we get instead $\int_b^a f(t)\,dt = -\int_a^b f(t)\,dt$. $\square$

So, the change of variables theorem from calculus is really the statement

$$\int_C \varphi^*\omega = \int_{\varphi(C)} \omega$$

for sufficiently nice maps φ. We will now see how this generalizes beyond intervals in $\mathbb{R}$. In any smooth manifold, a **piecewise smooth curve** α is a continuous map $\alpha\colon [a,b] \to M$ for some nonempty interval $[a,b]$ with the property that there are finitely many points $a = a_0 < a_1 < \cdots < a_k = b$ such that $\alpha|_{[a_{j-1},a_j]}$ is smooth for $1 \leq j \leq k$. Note, this is slightly stronger than insisting that γ be smooth on (a_{j-1}, a_j); it means that γ has an extension to a smooth curve $\tilde{\gamma}_j\colon (a_{j-1}-\epsilon, a_j+\epsilon) \to M$ for some $\epsilon > 0$ (although the actual value of γ may well disagree with this extension off the interval $[a_{j-1}, a_j]$).

Now, let α be a piecewise smooth curve in M. If $\alpha\colon [a,b] \to M$ is actually smooth, then we can pullback along α: $\alpha^*\omega \in \Omega^1([a,b])$. This means that $\alpha^*\omega = f(t)\,dt$ for some smooth f, and so we know how to integrate it. We therefore define

$$\int_\alpha \omega \equiv \int_{[a,b]} \alpha^*\omega.$$

More generally, if α is any piecewise smooth curve on $[a,b]$ that is smooth on the intervals $[a_{j-1}, a_j]$ for $1 \leq j \leq k$, then we define

$$\int_\alpha \omega \equiv \sum_{j=1}^{k} \int_{[a_{j-1},a_j]} \alpha^*\omega.$$

This is called a **line integral**. Here are some basic properties that are elementary to prove from the definitions.

PROPOSITION 6.23. *Let M be a smooth manifold, and let $\alpha\colon [a,b] \to M$ be a piecewise smooth curve. Let $\omega, \omega_1, \omega_2 \in \Omega^1(M)$, and let $c_1, c_2 \in \mathbb{R}$.*

(a) *For any $c_1, c_2 \in \mathbb{R}$,*

$$\int_\alpha (c_1\omega_1 + c_2\omega_2) = c_1 \int_\alpha \omega_1 + c_2 \int_\alpha \omega_2.$$

(b) *If α is a constant curve, then $\displaystyle\int_\alpha \omega = 0$.*

(c) *If $a \leq c \leq b$ and $\alpha_1 = \alpha|_{[a,c]}$ and $\alpha_2 = \alpha|_{[c,b]}$, then*

$$\int_\alpha \omega = \int_{\alpha_1} \omega + \int_{\alpha_2} \omega.$$

(d) *Let N be a smooth manifold and $F\colon M \to N$ a smooth map. For any $\eta \in \Omega^1(N)$,*

$$\int_\alpha F^*\eta = \int_{F\circ\alpha} \eta.$$

PROOF. We prove (d), leaving the others as calculations to the reader. First, it suffices to prove in the case that α is smooth, since in general if α is smooth on the intervals $[a_{j-1}, a_j]$ then $F \circ \alpha$ is smooth on the same intervals. Thus, if $\alpha\colon [a,b] \to M$ is smooth, we have

$$\int_\alpha F^*\eta = \int_a^b \alpha^* F^*\eta.$$

Note that, for any $X \in \mathscr{X}([a,b])$, then for any $t \in [a,b]$

$$\begin{aligned}(\alpha^*F^*\eta)(X)(t) = (F^*\eta)(d\alpha_t(X_t))(\alpha(t)) &= \eta_{F(\alpha(t))}(dF_{\alpha(t)} \circ d\alpha_t(X_t)) \\ &= \eta_{F(\alpha(t))}(d(F\circ\alpha)_t(X_t)) \\ &= (F\circ\alpha)^*\eta(X)(t).\end{aligned}$$

I.e. $\alpha^*F^*\eta = (F\circ\alpha)^*\eta$, and so

$$\int_\alpha F^*\eta = \int_a^b (F\circ\alpha)^*\eta = \int_{F\circ\alpha} \eta.$$

$\square$

EXAMPLE 6.24. Let $M = \mathbb{R}^2 \setminus \{0\}$, and define $\omega \in \Omega^1(M)$ in local (global) coordinates by

$$\omega = \frac{x\,dy - y\,dx}{x^2 + y^2}.$$

Let $\alpha\colon [0, 2\pi] \to M$ be the standard parametrization of the ccw unit circle $\alpha(t) = (\cos t, \sin t)$. Then we compute

$$\alpha^*\omega = \frac{\cos t d \sin t - \sin t d \cos t}{\cos^2 t + \sin^2 t} = \cos t \cos t\, dt - \sin t(-\sin t)\, dt = dt.$$

Thus

$$\int_\alpha \omega = \int_0^{2\pi} dt = 2\pi.$$

We have defined curves to have a parametrization as part of the definition. But it turns out that the dependence on the parametrization is very mild. A curve $\beta\colon [c, d] \to M$ is called a **reparametrization** of $\alpha\colon [a, b] \to M$ if there is a homeomorphism $\varphi\colon [c, d] \to [a, b]$ so that $\beta = \alpha \circ \varphi$, and moreover if α is smooth on $[a_{j-1}, a_j]$ then $\varphi^{-1}|_{[a_{j-1},a_j]}$ is a diffeomorphism. Such a map is therefore monotone everywhere, and strictly monotone on each interval $\varphi[a_{j-1}, a_j]$. We call this a **forward reparametrization** if φ is increasing, and a **backward reparametrization** if φ is decreasing.

PROPOSITION 6.25. *Let M be a smooth manifold and $\omega \in \Omega^1(M)$. Let α be a piecewise smooth curve, and let β be a reparametrization. Then*

$$\int_\beta \omega = \pm \int_\alpha \omega,$$

$+$ if it is a forward reparametrization and $-$ if it is backward.

PROOF. Again, it suffices to assume α is smooth (by restricting to its smooth subintervals and adding up on both sides at the end). Then we have

$$\int_\beta \omega = \int_{[c,d]} (\alpha \circ \varphi)^*\omega = \int_{[c,d]} \varphi^*(\alpha^*\omega) = \pm \int_{[a,b]} \alpha^*\omega = \pm \int_\alpha \omega$$

by Proposition 6.22. □

Thus, the line integral does not really depend on the curve α by only its image $\alpha[a, b]$, together with an "orientation" (forward or backward).

Let us now connect our pullback definition of line integrals with the usual definition given in vector calculus.

PROPOSITION 6.26. *Let $\alpha\colon [a, b] \to M$ be a piecewise smooth curve, and let $\omega \in \Omega^1(M)$. Then*

$$\int_\alpha \omega = \int_a^b \omega_{\alpha(t)}(\dot\alpha(t))\, dt.$$

PROOF. First, note that $\alpha[a, b] \subset M$ is compact, and so there are finitely many charts (U_j, φ_j) that cover the image. Subdividing the domain further if necessary, we may assume that α is smooth on the intervals $[a_{j-1}, a_j]$ where $\alpha[a_{j-1}, a_j] \subset U_j$. It therefore suffices to assume that α is smooth,

and its image is contained in a single chart (U, φ). Let $\hat{\alpha}(t) = \varphi \circ \alpha(t) = (\alpha^1(t), \dots, \alpha^n(t))$, and write $\omega = \sum_{j=1}^n \omega_j \, dx^j$ in this chart. Then we have

$$\omega_{\alpha(t)}(\dot{\alpha}(t)) = \sum_{j=1}^n \omega_j(\alpha(t)) dx^j(\dot{\alpha}(t)) = \sum_{j=1}^n \omega_j(\alpha(t)) \dot{\alpha}^j(t).$$

Now $\alpha^j \colon [a, b] \to \mathbb{R}$ is a calculus function, and so $d\alpha_t^j = \dot{\alpha}_j(t) \, dt$. Thus we have

$$\omega_{\alpha(t)}(\dot{\alpha}(t)) \, dt = \sum_{j=1}^n (\omega_j \circ \alpha)(t) d(\alpha^j)_t.$$

On the other hand

$$(\alpha^* \omega)_t = \sum_{j=1}^n \alpha^*(\omega_j dx^j) = \sum_{j=1}^n (\omega_j \circ \alpha)(t) d(x^j \circ \alpha)(t) = \sum_{j=1}^n (\omega_j \circ \alpha)(t) d(\alpha^j)_t$$

by Lemma 6.15. Thus

$$\int_\alpha \omega = \int_{[a,b]} \alpha^* \omega = \int_a^b \omega_{\alpha(t)}(\dot{\alpha}(t)) \, dt.$$

□

This brings us to the Fundamental Theorem of Calculus.

THEOREM 6.27. *Let M be a smooth manifold. Let $f \in C^\infty(M)$ and let $\alpha \colon [a, b] \to M$ be a piecewise smooth curve. Then*

$$\int_\alpha df = f(\alpha(b)) - f(\alpha(b)).$$

PROOF. Let $a = a_0 < a_1 < \cdots < a_k = b$ be partitions points so that α is smooth on $[a_{j-1}, a_j]$ for $1 \leq j \leq k$. Let $\alpha_j = \alpha|_{[a_{j-1}, a_j]}$. By Proposition 6.26, we have

$$\int_{\alpha_j} df = \int_{a_{j-1}}^{a_j} df_{\alpha(t)}(\dot{\alpha}(t)) \, dt.$$

Now we compute

$$df_{\alpha(t)}(\dot{\alpha}(t)) = [\dot{\alpha}(t)](f) = \left. \frac{d}{ds} \right|_{s=t} (f \circ \alpha) = (f \circ \alpha)'(t).$$

Thus

$$\int_{\alpha_j} df = \int_{a_{j-1}}^{a_j} (f \circ \alpha)'(t) \, dt.$$

The function $t \mapsto f \circ \alpha$ is smooth, and hence we may apply the classical fundamental theorem of calculus to compute this is equal to

$$\int_{\alpha_j} df = f(\alpha(a_j)) - f(\alpha(a_{j-1})).$$

Finally, we then have

$$\int_\alpha df = \sum_{j=1}^k \int_{\alpha_j} df = \sum_{j=1}^n [f(\alpha(a_j)) - f(\alpha(a_{j-1}))] = f(\alpha(a_k)) - f(\alpha(a_0))$$

as this is a telescoping sum, concluding the proof. □

6. Exact and Closed 1-Forms

In classic vector calculus, a vector field is called **conservative** if it is the gradient of a smooth function. We might use the same word for 1-forms, but the more common term is **exact**: a 1-form $\omega \in \Omega^1(M)$ is exact if there is some $f \in C^\infty(M)$ for which $\omega = df$. By the Fundamental Theorem of Calculus (Theorem 6.27), if ω is exact then $\int_\alpha \omega$ only depends on α through its endpoints: i.e. if α and β are any two piecewise smooth curves connecting p to q, then $\int_\alpha \omega = \int_\beta \omega$. In particular, for any *closed* curve α, $\int_\alpha \omega = 0$. In fact, this is an equivalence. To see this, we first need a path-connectedness lemma.

LEMMA 6.28. *If M is a connected smooth manifold, and $p, q \in M$, there is a piecewise smooth curve $\alpha\colon [a,b] \to M$ with $\alpha(a) = p$ and $\alpha(1) = b$.*

We could always choose $\{a, b\} = \{0, 1\}$, but it is convenient to have the freedom to use other parameter domains.

PROOF. This is a (by now) standard local-to-global-connectedness proof. Fix $p \in M$, and let $\mathscr{C} \subseteq M$ be the set of points connected to p via some piecewise smooth curve:

$$\mathscr{C} = \{q \in M \colon \exists a < b, \text{ piecewise smooth } \alpha\colon [a,b] \to M \text{ s.t. } \alpha(a) = p, \alpha(b) = q\}.$$

The constant curve $\alpha(t) = p$ shows that $p \in \mathscr{C}$, so $\mathscr{C} \neq \varnothing$.

- *$\mathscr{C}$ is open*: Fix some $q \in \mathscr{C}$, and let (U, φ) be a chart at q. Fix $\delta > 0$ so that $B_\delta(\hat{q}) \subseteq \varphi(U)$, and let $V = \varphi^{-1}(B_\delta(\hat{q}))$. For any $r \in V$, the straight line curve $\hat{\alpha}(t) = (1-t)\hat{q} + t\hat{r}$ has image contained in $B_\delta(\hat{q})$ (since this ball is convex), and has a smooth extension to a slightly longer line on both sides. Then $\alpha = \varphi^{-1} \circ \hat{\alpha}$ is a smooth curve from q to r; concatenating it with the curve connecting p to q (presumed to exists since $q \in \mathscr{C}$) shows that $r \in \mathscr{C}$. Thus the neighborhood $V \ni q$ is contained in $\mathscr{C}$, showing that $\mathscr{C}$ is open.
- *$\mathscr{C}$ is closed*: let $q \in \partial\mathscr{C}$. Again fix a chart (U, φ) at q. Since $q \in \partial\mathscr{C}$, the open neighborhood U of q contains some point $r \in \mathscr{C} \cap U$. Constructing the straight-line path in local coordinates just as above thus shows that $r \in \mathscr{C}$. Thus $\partial\mathscr{C} \subseteq \mathscr{C}$, so $\mathscr{C}$ is closed.

Thus $\mathscr{C}$ is a clopen nonempty subset of the connected manifold M, and hence $\mathscr{C} = M$. □

THEOREM 6.29. *Let M be a smooth manifold, and let $\omega \in \Omega^1(M)$. Then ω is exact iff $\int_\alpha \omega = 0$ for any closed piecewise smooth curve α.*

PROOF. Using the discussion above, the Fundamental Theorem of Calculus proves the 'only if' direction of the theorem. For the converse, suppose integrals of ω around closed curves are always 0. This in fact implies that such integrals are path independent: for if α and β are two piecewise smooth curves connecting p to q, then the reversed curve $-\beta$ connects q to p, and so the concatenation $\alpha - \beta$ is a closed curve. Thus

$$0 = \int_{\alpha-\beta} \omega = \int_\alpha \omega + \int_{-\beta} \omega = \int_\alpha \omega - \int_\beta \omega.$$

Now, let us assume that M is connected. We may then define an "integral" operation for any two points $p, q \in M$:

$$\int_p^q \omega = \int_\alpha \omega \quad \text{for any piecewise smooth curve } \alpha \text{ connecting } p \text{ to } q.$$

This is well-defined and makes sense for any p, q by Lemma 6.28. Also, using Propositions 6.25 and 6.23(c), we have

$$\int_p^q \omega = -\int_q^p \omega, \quad \int_p^r \omega = \int_p^q \omega + \int_q^r \omega.$$

So: fix a base point $p_0 \in M$, and define a function $f\colon M \to \mathbb{R}$ by $f(p) = \int_{p_0}^p \omega$. We will show that $f \in C^\infty(M)$, and $df = \omega$. To accomplish this, let $q \in M$, and fix a chart $(U, \varphi = (x^j)_{j=1}^n)$ at q. In this chart we have $\omega = \sum_{j=1}^n \omega_j\, dx^j$. We will compute that $\omega_j(q) = \frac{\partial f}{\partial x^j}(q)$, which means that $df_q = \omega_q$ as required.

Fix $j \in \{1, \dots, n\}$ and let $\epsilon > 0$ be small enough that the curve $\alpha\colon (-\epsilon, \epsilon) \to U$ given by $\varphi \circ \alpha(t) = (0, \dots, t, \dots, 0)$ (with the t in the jth place) stays contained in U. Let $p_1 = \alpha(-\epsilon)$, and define a new function $\tilde{f}(p) = \int_{p_1}^p \omega$. Then

$$\tilde{f}(p) - f(p) = \int_{p_0}^p \omega - \int_{p_1}^p \omega = \int_{p_0}^p \omega + \int_p^{p_1} \omega = \int_{p_0}^{p_1} \omega$$

which is a constant. Hence, it suffices to show that $\tilde{f}$ is smooth and satisfies $\frac{\partial \tilde{f}}{\partial x^j}(q) = \omega_j(q)$. Well, by construction $\dot{\alpha}(t) = \frac{\partial}{\partial x^j}|_{\alpha(t)}$, and so

$$\omega_{\alpha(t)}(\dot{\alpha}(t)) = \sum_{i=1}^n \omega_i(\alpha(t))\, dx^i \left(\frac{\partial}{\partial x^j}\bigg|_{\alpha(t)} \right) = \omega_j(\alpha(t)).$$

Hence, we have

$$\tilde{f}(\alpha(t)) = \int_{p_1}^{\alpha(t)} \omega = \int_{-\epsilon}^t \omega_{\alpha(s)}(\dot{\alpha}(s))\, ds = \int_{-\epsilon}^t \omega_j(\alpha(s))\, ds$$

But $\omega \in \Omega^1(M)$, so its components are smooth. It then follows from the (classical) fundamental theorem of calculus that $\tilde{f} \circ \alpha$ is smooth, and

$$\frac{\partial \tilde{f}}{\partial x^j}(q) = \dot{\alpha}(0)\tilde{f} = \frac{d}{dt}\bigg|_{t=0} \tilde{f} \circ \alpha(t) = \frac{d}{dt}\bigg|_{t=0} \int_{-\epsilon}^t \omega_j(\alpha(s))\, ds = \omega_j(\alpha(0)) = \omega_j(q).$$

Since the components ω_j are smooth functions of q, this shows that $\tilde{f}$ is smooth, and moreover that $d\tilde{f}_q = df_q = \omega$ as claimed.

Finally, if M is not connected, then the above argument shows there are functions $f_i \in C^\infty(M_i)$ on the connected components M_i of M so that $df_i = \omega|_{M_i}$. Then the function f whose value on M_i is f_i is smooth and satisfies $df = \omega$. □

Now, not every 1-form is exact: as Example 6.24 shows, there are 1-forms with non-zero integrals around closed curves. So how can we tell if a 1-form ω *is* exact, without already knowing a function f for which $\omega = df$? To see the answer, we work in local coordinates: write $\omega = \sum_{j=1}^n \omega_j\, dx^j$ in some chart. If $\omega = df$, this means that $\omega_j = \frac{\partial f}{\partial x^j}$. Since f is smooth, its mixed partials commute, which means that

$$\frac{\partial \omega_j}{\partial x^i} = \frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i} = \frac{\partial \omega_i}{\partial x^j}.$$

DEFINITION 6.30. *A* 1*-form* ω *is called* **closed** *if, in every chart, the components* ω_j *satisfy*

$$\frac{\partial \omega_j}{\partial x^i} = \frac{\partial \omega_i}{\partial x^j}. \tag{6.11}$$

The preceding calculation shows that every exact 1-form is closed. This gives us a negative test for exactness: if we can find some chart in which (6.11) fails at some point, then ω is not exact. This is still a tall order, however: it, in principle, requires us to check *every* chart, which is impractical. In fact, as we will show next, it suffices to check this condition in *any one* chart; moreover, there is an equivalent invariant condition.

PROPOSITION 6.31. *For any $\omega \in \Omega^1(M)$, TFAE:*

(a) *ω is closed.*
(b) *ω satisfies* (6.11) *in some chart at each point.*
(c) *Given any open $U \subseteq M$ and any vector fields $X, Y \in \mathscr{X}(U)$, we have*

$$X(\omega(Y)) - Y(\omega(X)) = \omega([X, Y]). \tag{6.12}$$

PROOF. The implication (a) $\Longrightarrow$ (b) is immediate from the definition (which states that (6.11) holds in *every* chart). Now, assume (b) holds true, and fix any open set U and vector fields $X, Y \in \mathscr{X}(U)$. Fix a point $p \in U$, and let $(V, \varphi = (x^j)_{j=1}^n)$ be a coordinate chart at p, contained in U, in which (6.11) holds. Expand $\omega = \sum_{i=1}^n \omega_i \, dx^i$, $X = \sum_{j=1}^n X^j \frac{\partial}{\partial x^j}$, and $Y = \sum_{k=1}^n Y^k \frac{\partial}{\partial x^k}$, and compute

$$X(\omega(Y)) = X\left(\sum_{i=1}^n \omega_i Y^i\right) = \sum_{i=1}^n \left(\omega_i X(Y^i) + Y^i X(\omega_i)\right) = \sum_{i=1}^n \left(\omega_i X(Y^i) + \sum_{j=1}^n Y^i X^j \frac{\partial \omega_i}{\partial x^j}\right).$$

Exchanging the roles of X and Y shows that

$$Y(\omega(X)) = \sum_{i=1}^n \left(\omega_i Y(X^i) + \sum_{j=1}^n X^i Y^j \frac{\partial \omega_i}{\partial x^j}\right).$$

Subtracting, this gives

$$X(\omega(Y)) - Y(\omega(X)) = \sum_{i=1}^n \omega_i [X(Y^i) - Y(X^i)] + \sum_{i,j=1}^n Y^i X^j \left(\frac{\partial \omega_i}{\partial x^j} - \frac{\partial \omega_j}{\partial x^i}\right), \tag{6.13}$$

and the last term is 0 by assumption (of (6.11)). We also note that

$$[X, Y] = XY - YX = X\left(\sum_{i=1}^n Y^i \frac{\partial}{\partial x^i}\right) - Y\left(\sum_{i=1}^n X^i \frac{\partial}{\partial x^i}\right) = \sum_{i=1}^n [X(Y^i) - Y(X^i)] \frac{\partial}{\partial x^i},$$

and so $\sum_{i=1}^n \omega_i [X(Y^i) - Y(X^i)] = \omega([X, Y])$, concluding the calculation, and verifying that (b) $\Longrightarrow$ (c).

Finally, we proved (c) $\Longrightarrow$ (d). Fix any chart, and expand the 1-form ω and the vector fields X, Y in coordinate bases. Note that the calculation of (6.13) and the following equation show that, in general,

$$X(\omega(Y)) - Y(\omega(X)) = \omega([X, Y]) + \sum_{i,j=1}^n Y^i X^j \left(\frac{\partial \omega_i}{\partial x^j} - \frac{\partial \omega_j}{\partial x^i}\right).$$

Thus, assuming (c), we have for any smooth functions X^i and Y^j,

$$\sum_{i,j=1}^n Y^i X^j \left(\frac{\partial \omega_i}{\partial x^j} - \frac{\partial \omega_j}{\partial x^i}\right) = 0.$$

Choosing the functions so that $Y^i = X^j = 1$ and $Y^k = 0$ if $k \neq i$ and $X^k = 0$ if $k \neq j$ yields the result. $\square$

REMARK 6.32. We will soon talk about k-forms for $k > 1$, and extend the operator d to act on all forms. In particular, for a 1-form ω, $d\omega$ will be a 2-form (which eats two vector fields); its action will be defined as

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]).$$

So, the condition that ω is closed is really the condition that $d\omega = 0$; hence, we have if $\omega = df$ then $d\omega = 0$, which is to say that $d^2 = 0$. This generalizes the classical vector calculus statement that $\nabla \times \nabla f = 0$ for any smooth function (the curl of a gradient is 0).

So, we have a nice, easy to check, invariant condition that gives a half-decidable test for whether a given 1-form is exact. It is not, however, a fully-decidable test.

EXAMPLE 6.33. Consider again the 1-form of Example 6.24:

$$\omega = -\frac{y}{x^2 + y^2}\, dx + \frac{x}{x^2 + y^2}\, dy.$$

We showed (using the fundamental theorem of calculus) that ω is not exact. Nevertheless, we can quickly calculate that

$$\frac{\partial}{\partial y}\left(-\frac{y}{x^2 + y^2}\right) = \frac{-x^2 + y^2}{(x^2 + y^2)^2} = \frac{\partial}{\partial x}\left(\frac{x}{x^2 + y^2}\right).$$

Thus ω is closed, but not exact. More on this shortly.

Both closeness and exactness are natural conditions, in the following sense.

COROLLARY 6.34. *Let $F\colon M \to N$ be a smooth map between manifolds, and let $\omega \in \Omega^1(N)$. If ω is exact, then $F^*\omega$ is exact. Moreover, if F is a diffeomorphism, and ω is closed, then $F^*\omega$ is closed.*

PROOF. First, by Lemma 6.15, if ω is exact and so has the form $\omega = df$, then $F^*(\omega) = F^*(df) = d(f \circ F)$ is also exact. Now, suppose we only know ω is closed. One can compute that, for any vector fields $X, Y \in \mathscr{X}(M)$ and $\omega \in \Omega^1(N)$,

$$X(F^*\omega(Y)) = F_*(X)(\omega(F_*(Y))).$$

Thus, we simply have

$$X(F^*\omega(Y)) - Y(F^*\omega(X)) = F_*(X)(\omega(F_*(Y))) - F_*(Y)(\omega(F_*(X))).$$

Since ω is closed on N, it follows that

$$F_*(X)(\omega(F_*(Y))) - F_*(Y)(\omega(F_*(X))) = \omega([F_*(X), F_*(Y)]) = \omega(F_*[X, Y])$$

and one final calculation shows that this equals $F^*\omega([X, Y])$, proving that $F^*\omega$ is closed.

Alternatively, it is actually simpler to work in local coordinates here. Let (U, φ) be a chart for N. Then $(F^{-1}(U), \varphi \circ F)$ is a chart for M. In these coordinates, the map F is given by the identity, and so it is clear that ω satisfies (6.11) in the φ coordinates iff $F^*\omega$ satisfies (6.11) in the $\varphi \circ F$-coordinates. The result then follows by Proposition 6.31. □

REMARK 6.35. (1) In fact, in the local proof of Corollary 6.34, one only needs that F is a *local* diffeomorphism: in this case, an invariant form of the inverse function theorem will allow the local-coordinates argument to work on some sufficiently small open subset of the original chart U. We have yet to state and prove such an invariant inverse function theorem, so we leave it as is for now.

(2) One might wonder whether the local diffeomorphism condition is really necessary to preserved closedness, and in fact it is *not* necessary: like exactness, closedness is invariant under all pullbacks. The reason is the same: referring to Remark 6.32, ω is closed iff $d\omega = 0$, and then we will have $dF^*\omega = F^*(d\omega) = F^*(0) = 0$, as with functions. In the next section, we will develop the technology to prove this.

Now, the question of how far apart the conditions *exact* and *closed* are is a deep and interesting one. Returning once more to Example 6.24, although this form is not exact on $\mathbb{R}^2 \setminus \{0\}$, if we restrict it to $M = \mathbb{R}^2 \setminus \{(x,0)\colon x \in \mathbb{R}\}$, then the function $M = \tan^{-1}\frac{y}{x}$ is smooth on M, and $\omega = df$. (In fact, although the formula doesn't make sense, this function f has a smooth extension to any domain $\mathbb{R}^2 \setminus \gamma$ for any ray $\gamma = \{tv\colon t > 0\}$ where v is some nonzero vector in $\mathbb{R}^2$.) What this demonstrates is that exactness is not really a local condition (while closedness is): it depends on the global topology of the manifold where the form is defined. The main basic result in this direction is Poincaré's lemma, which shows that closed 1-forms on star-shaped regions in $\mathbb{R}^n$ are indeed exact.

THEOREM 6.36 (Poincaré Lemma). *Let $U \subseteq \mathbb{R}^n$ be open and star-shaped. Then every closed 1-form on U is exact.*

Note: star-shaped means there is a "center" point $c \in U$ so that, for every $p \in U$, the line segement $\{(1-t)c + tp\colon 0 \le t \le 1\}$ is contained in U. For example, convex sets are star-shaped.

PROOF. Let $\omega \in \Omega^1(U)$ be closed. First, let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be the translation $T(x) = x - c$; since T is a diffeomorphism, T^* preserves closedness and exactness, and so it suffices to prove that the closed 1-form $T^*\omega$ is exact. Henceforth we rename $T(U)$ as U and $T^*\omega$ as ω, and proceed under the assumption that $c = 0$ (which is just for notational convenience). Thus, for any $x \in U$, the straight-line $\alpha_x\colon [0,1] \to U$ given by $\alpha_x(t) = tx$ is a smooth curve in U, and we can define a function $f\colon U \to \mathbb{R}$ by

$$f(x) = \int_{\alpha_x} \omega.$$

We will show that f is smooth, and that $\omega = df$ – i.e. that $\frac{\partial f}{\partial x^j} = \omega_j$ for $1 \le j \le n$.

Expand $\omega = \sum_{j=1}^n \omega_j\, dx^j$. Use Proposition 6.26 to write f as

$$f(x) = \int_0^1 \omega_{\alpha_x(t)}(\dot\alpha_x(t))\, dt = \sum_{j=1}^n \int_0^1 \omega_j(tx)x^j\, dt.$$

The integrant is a smooth function of $t, x^1, \dots, x^n$, and so we may differentiate under the integral

$$\begin{aligned}\frac{\partial f}{\partial x^i}(x) &= \sum_{j=1}^n \int_0^1 \left(t\frac{\partial \omega_j}{\partial x^i}(tx)x^j + \omega_j(tx)\delta_i^j\right) dt \\ &= \int_0^1 \left(\sum_{j=1}^n \frac{\partial \omega_j}{\partial x^i}(tx)tx^j + \omega_i(tx)\right) dt.\end{aligned}$$

We now use the assumption that ω is closed, so that $\frac{\partial \omega_j}{\partial x^i} = \frac{\partial \omega_i}{\partial x^j}$, which gives

$$\frac{\partial f}{\partial x^i}(x) = \int_0^1 \left(\sum_{j=1}^n \frac{\partial \omega_i}{\partial x^j}(tx)tx^j + \omega_i(tx)\right) dt = \int_0^1 \frac{d}{dt}(t\omega_i(tx))\, dt = t\omega_i(tx)|_{t=0}^{t=1} = \omega_i(x).$$

Hence, the partial derivatives of f are the (smooth) functions ω_i. This shows f is smooth, and that $df = \omega$, as claimed. $\square$

REMARK 6.37. The precise condition for this to work is much weaker than star-shaped: it is *simply-connected*. That is: it must be true that any two piecewise smooth curves from p to q are homotopic. If this is the case, then one can define f as follows: choose a base point $p \in U$, and let α_x be some curve from p to x that is a finite collection of line segments in coordinate directions. Any two such curves are homotopic, and this (together with closedness of ω) shows that f is well-defined. A more involved version of the above calculation shows that $df(x) = \omega_x$.

COROLLARY 6.38. *If $\omega \in \Omega^1(M)$ is closed, then every point $p \in M$ has a neighborhood U so that $\omega|_U$ is exact.*

PROOF. Let $p \in M$, and choose a chart (V, φ) at p; let $\hat{U} \subseteq \varphi V)$ be a ball centered at $\varphi(p)$, and let $U = \varphi^{-1}(\hat{U})$. By assumption ω is closed and so, in the chart (U, φ) condition (6.11) holds true. Thus, since $\hat{U}$ is convex, the Poincaré Lemma implies that the components ω_j of ω in the φ-coordinates $(x^j)_{j=1}^n$ satisfy $\omega_j = \frac{\partial \hat{f}}{\partial x^j}$ for some smooth function $\hat{f}$ on $\hat{U}$. It then follows (pulling back along φ) that $\omega|_U = df$ where $f = \hat{f} \circ \varphi$. Thus $\omega|_U$ is exact. $\square$

CHAPTER 7

Tensors and Exterior Algebra

1. Mulitilinear Algebra

For a real vector space V, recall that V^* denotes the dual space $V^* = L(V, \mathbb{R})$. We now consider a multivariate generalization of this.

DEFINITION 7.1. *Let $V_1, \dots, V_k, W$ be real vector spaces. Denote by $L(V_1, \dots, V_k; W)$ the space of* **multilinear maps** $V_1 \times \cdots \times V_k \to W$. *This means that, for each $1 \le j \le k$, for $F \in L(V_1, \dots, V_k; W)$ we have*

$$F(v_1, \dots, v_{j-1}, av_j, v_{j+1}, \dots, v_k) = aF(v_1, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_k) \qquad a \in \mathbb{R}, \quad \textit{and}$$
$$F(v_1, \dots, v_{j-1}, v_j + v_j', v_{j+1}, \dots, v_k) = F(v_1, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_k) + F(v_1, \dots, v_{j-1}, v_j', v_{j+1}, \dots, v_k).$$

That is: F is linear in each slot separately.

EXAMPLE 7.2. Some common examples of multilinear maps are:

- The dot product $\langle v, w \rangle = \sum_{j=1}^n v_j w_j$ is a multilinear map in $L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R})$. We call it **bilinear**.
- The Lie bracket $[\cdot, \cdot]$ is a bilinear map $\mathscr{X}(M) \times \mathscr{X}(M) \to \mathscr{X}(M)$.
- The determinant $\det$ on $n \times n$ matrices, though of as a map on the columns of the matrix, is a multilinear map in $L(\mathbb{R}^n, \dots, \mathbb{R}^n; \mathbb{R})$.

It might be tempting to try to identify $L(V_1, \dots, V_k; \mathbb{R})$ with the dual space $L(V_1 \times \cdots \times V_k, \mathbb{R}) = (V_1 \times \cdots \times V_k)^*$, but these two spaces have little to do with each other when $k > 1$. Indeed, the two intersect trivially: if $F \in L(V, W; \mathbb{R})$ then $F(av, aw) = a^2 F(v, w)$, while if it is in $L(V \times W, \mathbb{R})$ then $F(av, aw) = F(a \cdot (v, w)) = aF(v, w)$. This must hold for all a, and so the only linear map in $L(V, W; \mathbb{R})$ is the zero map.

It turns out that there *is* a close connection between $L(V_1, \dots, V_k; \mathbb{R})$ and the space $V_1^* \times \cdots \times V_k^*$. We can define a map

$$\begin{gathered} \otimes \colon V_1^* \times \cdots \times V_k^* \to L(V_1, \dots, V_k; \mathbb{R}) \\ (\otimes(\lambda_1, \dots, \lambda_k))\,(v_1, \dots, v_k) = \lambda_1(v_1)\lambda_2(v_2) \cdots \lambda_k(v_k). \end{gathered} \tag{7.1}$$

It is easy to check that $\otimes(\lambda_1, \dots, \lambda_k)$ is indeed a multilinear map. In fact, it is a special case of the following: if $F \in L(V_1, \dots, V_k; \mathbb{R})$ and $G \in L(W_1, \dots, W_\ell; \mathbb{R})$, then define $F \otimes G$ to be the map $V_1 \times \cdots \times V_k \times W_1 \times \cdots \times W_\ell \to \mathbb{R}$ given by

$$(F \otimes G)(v_1, \dots, v_k, w_1, \dots, w_\ell) = F(v_1, \dots, v_k) G(w_1, \dots, w_\ell).$$

It is again easy to verify that $F \otimes G$ is multilinear. The map in (7.1) is simply the iteration of this map:

$$\otimes(\lambda_1, \dots, \lambda_k) = (\cdots((\lambda_1 \otimes \lambda_2) \otimes \lambda_3) \cdots \otimes \lambda_k).$$

It is easy to check that this operation is in fact associative, so the order of the parentheses doesn't matter. The operation is often called **tensor product**.

One might hope that the map $\otimes$ in (7.1) is an isomorphism identifying $L(V_1, \dots, V_k; \mathbb{R})$ with $V_1^* \times \cdots \times V_k^*$, but this is not true. For one thing, note that $0 \otimes v = 0$ for any v, and so it is never injective if $k > 1$. It is not surjective either: for example, the dot product on $\mathbb{R}^n \times \mathbb{R}^n$ is not of this form. It is, however, a linear combination of maps of this form. This will generally be true. Multilinear maps in the image of $\otimes$ are called **pure tensors**; every multilinear map is a linear combination of pure tensors. We can summarize this with the following useful lemma on finding a basis for $L(V_1, \dots, V_k; \mathbb{R})$ (in the finite-dimensional setting).

LEMMA 7.3. *Let $V_1, \dots, V_k$ be finite-dimensional vector spaces. Let $\beta_j = \{e_j^1, \dots, e_j^{n_j}\}$ be a basis for V_j. Then the collection*

$$\beta^* = \{(e_1^{i_1})^* \otimes \cdots \otimes (e_k^{i_k})^* \colon 1 \le i_1 \le n_1, \dots, 1 \le i_k \le n_k\}$$

is a basis for $L(V_1, \dots, V_k; \mathbb{R})$. In particular, this shows that every map in $L(V_1, \dots, V_k; \mathbb{R})$ is a linear combination of pure tensors, and moreover this space has dimension $\dim V_1 \cdot \dim V_2 \cdots \dim V_k$.

PROOF. For each $F \in L(V_1, \dots, V_k; \mathbb{R})$, define coefficients $F_{i_1,\dots,i_k}$ by

$$F_{i_1,\dots,i_k} = F(e_1^{i_1}, \dots, e_k^{i_k}). \tag{7.2}$$

Now, for any $v_j \in V_j$, we can expand in the basis $v_j = \sum_{i=1}^{n_j} v_j^i e_j^i$. Using multilinearity and the definition of the dual basis, this means that

$$\begin{aligned} F(v_1, \dots, v_k) &= \sum_{i_1,\dots,i_k} v_1^{i_1} \cdots v_k^{i_k} F(e_1^{i_1}, \dots, e_k^{i_k}) \\ &= \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k} v_1^{i_1} \cdots v_k^{i_k} \\ &= \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k} (e_1^{i_1})^*(v_1) \cdots (e_k^{i_k})^*(v_k) \\ &= \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k} (e_1^{i_1})^* \otimes \cdots \otimes (e_k^{i_k})^*(v_1, \dots, v_k). \end{aligned}$$

In other words, we've expanded

$$F = \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k} (e_1^{i_1})^* \otimes \cdots \otimes (e_k^{i_k})^*. \tag{7.3}$$

This shows that β^* is a spanning set for $L(V_1, \dots, V_k; \mathbb{R})$. What's more, suppose some linear combination as in (7.3) equals 0 (as a mutlilinear map). In particular, this means it yields 0 when applied to the vector $(e_1^{j_1}, \dots, e_k^{j_k})$, and this (by definition of dual basis) yields $F_{j_1,\dots,j_k} = 0$. As this holds for all $j_1, \dots, j_k$, it follows that all the coefficients are 0, and hence β^* is linearly independent. Thus it is a basis. $\square$

Note: we have shown, in the process, that the coefficients of any multilinear map in terms of this tensor basis are given by (7.2).

We have now shown that the space of multilinear maps $V_1 \times \cdots \times V_k \to \mathbb{R}$ has a special relationship to the space $V_1^* \times \cdots \times V_k^*$: there is a canonical mapping $\otimes \colon V_1^* \times \cdots \times V_k^* \to L(V_1, \dots, V_k; \mathbb{R})$. This map is not injective in general, but it is close; in the next section, we will see in what sense it "identifies" these spaces.

2. The Tensor Product Construction

The space $L(V_1, \dots, V_k; \mathbb{R})$ of multilinear maps $V_1 \times \cdots \times V_k \to \mathbb{R}$ has a special universal property in the category of all multilinear maps.

PROPOSITION 7.4. *Let $V_1, \dots, V_k$ be any real vector spaces. Let W be another vector space, and let $F\colon V_1^* \times \cdots \times V_k^* \to W$ be any multilinear map. Then F factors through a unique* linear *map $\widetilde{F}\colon L(V_1, \dots, V_k; \mathbb{R}) \to W$: $F = \widetilde{F} \circ \otimes$.*

$$\begin{tikzcd} V_1^* \times \cdots \times V_k^* \arrow[r, "F"] \arrow[d, "\otimes"'] & W \\ L(V_1, \dots, V_k; \mathbb{R}) \arrow[ur, "\widetilde{F}"'] & \end{tikzcd}$$

PROOF. The factoring condition tells us that, for any given dual vectors $\lambda_j \in V_j^*$, we must have

$$\widetilde{F}(\lambda_1 \otimes \cdots \otimes \lambda_k) = F(\lambda_1, \dots, \lambda_k). \tag{7.4}$$

Assuming this is well-defined (which is basically the content of the propositions), this tells us how to define $\widetilde{F}$ (and shows it is unique). Fix bases β_j for V_j, and the corresponding basis β^* for $L(V_1, \dots, V_k; \mathbb{R})$, as in Lemma 7.3. To define the linear map $\widetilde{F}$, it is necessary and sufficient to define its action on β^*. There is only one way to do this so the factoring condition holds: we must have

$$\widetilde{F}((e_1^{i_1})^* \otimes \cdots \otimes (e_k^{i_k})^*) = F((e_1^{i_1})^*, \dots, (e_k^{i_k})^*).$$

This definition shows that (7.4) holds true at least for elements $\lambda_j \in V_j^*$ that are dual basis vectors; that it holds in general is simply a matter of expanding the left-hand-side of (7.4) using multilinearity of $\otimes$, appling the definition, and then reassembling using multilinearity of F. $\square$

In fact, this universal property can be used to *define* the tensor product abstractly. For the moment, let us forget our concrete definition of $\otimes$ (as an operation on pairs of real-valued multilinear maps). Instead, we could make the following abstract algebraic definition.

DEFINITION 7.5. *Let $U_1, \dots, U_k$ be real vector spaces. A* **tensor product space** *for $U_1, \dots, U_k$ is a pair $(T, \otimes)$ where T is a real vector space and $\otimes\colon U_1 \times \cdots \times U_k \to T$ is a multilinear map, with the following property: if W is any real vector space, and $F\colon U_1 \times \cdots \times U_k \to W$ is a multilinear map, then F factors through a unique linear map $\widetilde{F}\colon T \to W$, i.e. $F = \widetilde{F} \circ \otimes$:*

$$\begin{tikzcd} U_1 \times \cdots \times U_k \arrow[r, "F"] \arrow[d, "\otimes"'] & W \\ T \arrow[ur, "\widetilde{F}"'] & \end{tikzcd} .$$

Proposition 7.4 shows that $L(V_1, \dots, V_k; \mathbb{R})$ is a tensor product space for $V_1^*, \dots, V_k^*$. Using the canonical isomorphism $V \cong V^{**}$ in the finite-dimensional setting, we therefore always have a tensor product space for $U_1, \dots, U_k$, given by $L(U_1^*, \dots, U_k^*; \mathbb{R})$.

We would like to talk about *the* tensor product space, while this definition only gives a property that anything called *a* tensor product space must have. In fact, this property uniquely defines the space T up to (unique) isomorphism.

EXERCISE 7.5.1. *Let $U_1, \dots, U_k$ be real vector spaces, and let $(T_1, \otimes_1)$ and $(T_2, \otimes_2)$ be two tensor product spaces for $U_1, \dots, U_k$. Show that there is a unique vector space isomorphism $\Phi\colon T_1 \to T_2$ such that $\otimes_2 = \Phi \circ \otimes_1$.*

Thus, we can (up to canonical isomorphism) refer to *the* tensor product space, and we denote it by $U_1 \otimes \cdots \otimes U_k$. *Caution*: this does not mean the image of $U_1 \times \cdots \times U_k$ under the multilinear map $\otimes$ (which is not even a vector space); it means the whole space. In the concrete context, we mean the entire space of multilinear maps, which is spanned by the image of $\otimes$.

Now, in the concrete case, we defined tensor product in such a way that associativity was immediate. That holds here in the abstract setting as well (although, since these spaces are only defined up to canonical isomorphism, we must be more careful about what we mean by "associative").

LEMMA 7.6. *Given any finite-dimensional real vector spaces U_1, U_2, U_3, there is are unique isomorphisms $(U_1 \otimes U_2) \otimes U_3 \cong U_1 \otimes U_2 \otimes U_3 \cong U_1 \otimes (U_2 \otimes U_3)$ identifying $(u_1 \otimes u_2) \otimes u_3 \cong u_1 \otimes u_2 \otimes u_3 \cong u_1 \otimes (u_2 \otimes u_3)$.*

PROOF. First, the three spaces have the same dimension (as can be seen by identifying with the concrete tensor product constructions of the last section and using Lemma 7.3). Now, define a map $\alpha\colon U_1 \times U_2 \times U_3 \to (U_1 \otimes U_2) \otimes U_3$ in the obvious fashion: $\alpha(u_1, u_2, u_3) = (u_1 \otimes u_2) \otimes u_3$. Since $\otimes$ is multilinear, this is a multilinear map, and hence α factors through a linear map $\widetilde{\alpha}\colon U_1 \otimes U_2 \otimes U_3 \to (U_1 \otimes U_2) \otimes U_3$, with the property that $\widetilde{\alpha}(u_1 \otimes u_2 \otimes u_3) = (u_1 \otimes u_2) \otimes u_3$. Since $(U_1 \otimes U_2) \otimes U_3$ is spanned by elements of the form $(u_1 \otimes u_2) \otimes u_3$, $\widetilde{\alpha}$ is surjective. Since the two spaces have the same dimension, it must be an isomorphism. The isomorphism with the space bracketed the other way is similar. $\square$

REMARK 7.7. Similarly appealing to Lemma 7.3, we see that in general a basis for the abstract tensor product space $U_1 \otimes \cdots \otimes U_k$ is given by the set of all vectors of the form $e_1^{i_1} \otimes \cdots \otimes e_k^{i_j}$, where each $\{e_j^i\}_i$ is a basis for U_j.

One final remark: our concrete construction of tensors is more precisely a construction of *covariant tensors*. The concrete realization of $V_1 \otimes \cdots \otimes V_k$ as $L(V_1^*, \dots, V_k^*; \mathbb{R})$ is the space of *contravariant tensors*; and we could mix and match some Vs with some V^*s to get *mixed rank tensors*. We will only have use for covariant tensors at present.

3. Symmetric and Antisymmetric Tensors

An argument much like the previous one shows that $V_1 \otimes V_2 \cong V_2 \otimes V_1$ (with canonical isomorphism extending $v_1 \otimes v_2 \mapsto v_2 \otimes v_1$). In general, the *space* $V_1 \otimes \cdots \otimes V_k$ is the same (up to canonical isomorphism) no matter what order the vector spaces are listed in. But the actual elements of these spaces are different: $v_1 \otimes v_2 \neq v_2 \otimes v_1$ in general. For our purposes, it is most useful to restrict attention to those tensors that do have symmetry (or antisymmetry) when changing the order of the vectors. This can only make sense if all the vector spaces V_j are equal.

DEFINITION 7.8. *Let S_k denote the symmetric group of permutations on $\{1, \dots, k\}$. For a given vector space V, define an action of S_k on $V \otimes \cdots \otimes V = V^{\otimes k}$ by (the linear extension of)*

$$\sigma \cdot v_1 \otimes \cdots \otimes v_k = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)}.$$

This is a *right* action: that is, for $\sigma, \tau \in S_k$,

$$\begin{aligned}(\sigma\tau) \cdot v_1 \otimes \cdots \otimes v_k &= v_{(\sigma\tau)(1)} \otimes \cdots \otimes v_{(\sigma\tau)(k)} \\ &= v_{\sigma(\tau(1))} \otimes \cdots \otimes v_{\sigma(\tau(k))} \\ &= \sigma \cdot (\tau \cdot v_1 \otimes \cdots \otimes v_k)\end{aligned}$$

Now, we define two projections on $V^{\otimes k}$:

$$\begin{aligned}\mathrm{Sym}(\psi) &= \frac{1}{k!} \sum_{\sigma \in S_k} \sigma \cdot \psi \\ \mathrm{Alt}(\psi) &= \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}\,(\sigma)\, \sigma \cdot \psi.\end{aligned}$$

Here $\mathrm{sgn}(\sigma)$ denotes the sign of a permutation σ: $+1$ if σ is even, and -1 if σ is odd. So, for example,

$$\mathrm{Sym}(u \otimes v) = \frac{1}{2}(u \otimes v + v \otimes u), \quad \mathrm{Alt}(u \otimes v) = \frac{1}{2}(u \otimes v - v \otimes u).$$

The images of these projections are known as

$$\begin{aligned}\Sigma^k(V) &= \mathrm{Sym}(V^{\otimes k}) = \textbf{symmetric } k\textbf{-tensors} \\ \Lambda^k(V) &= \mathrm{Alt}(V^{\otimes k}) = \textbf{antisymmetric } k\textbf{-tensors}.\end{aligned}$$

Antisymmetric tensors are also called **alternating**, ergo the name Alt for the projection. We can characterize these spaces more simply in terms of actions under any given permutation.

LEMMA 7.9. *Let $\psi \in V^{\otimes k}$. Then $\psi \in \Sigma^k(V)$ iff $\sigma \cdot \psi = \psi$ for any $\sigma \in S_k$. On the other hand, $\psi \in \Lambda^k(V)$ iff $\sigma \cdot \psi = \mathrm{sgn}\,(\sigma)\, \psi$ for any $\sigma \in S_k$.*

PROOF. If $\sigma \cdot \psi = \psi$ for each σ, then

$$\mathrm{Sym}(\psi) = \frac{1}{k!} \sum_{\sigma \in S_k} \sigma \cdot \psi = \frac{1}{|S_k|} \sum_{\sigma \in S_k} \psi = \psi.$$

Hence ψ is in the image of Sym, meaning it is in $\Sigma^k(V)$. Conversely, if $\psi \in \Sigma^k(V)$, then $\psi = \mathrm{Sym}(\varphi)$ for some tensor $\varphi \in V^{\otimes k}$, and we compute

$$\sigma \cdot \psi = \sigma \cdot \frac{1}{k!} \sum_{\tau \in S_k} \tau \cdot \varphi = \frac{1}{k!} \sum_{\tau \in S_k} \sigma \cdot (\tau \cdot \varphi) = \frac{1}{k!} \sum_{\tau \in S_k} (\sigma\tau) \cdot \varphi.$$

The map $\tau \mapsto \sigma\tau$ is a bijection of S_k, and so relabeling we just get

$$\sigma \cdot \psi = \frac{1}{k!} \sum_{\tau' \in S_k} \tau' \cdot \varphi = \psi$$

as desired. The argument for $\Lambda^k(V)$ is almost identical, with the only additional needed information being that $\mathrm{sgn}(\sigma\tau) = \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau)$. □

In the process, we proved the following (which is why it is reasonable to call Sym and Alt 'projections').

COROLLARY 7.10. *Let $\psi \in V^{\otimes k}$. Then $\psi \in \Sigma^k(V)$ iff $\mathrm{Sym}(\psi) = \psi$, and $\psi \in \Lambda^k(V)$ iff $\mathrm{Alt}(\psi) = \psi$.*

We will actually focus much more on $\Lambda^k(V)$ than $\Sigma^k(V)$. For starters, there is an even simpler equivalent condition for inclusion in $\Lambda^k(V^*)$, which we identify in the usual way as a subspace of $L(V, \dots, V; \mathbb{R})$.

PROPOSITION 7.11. *Let $\alpha \in L(V, \dots, V; \mathbb{R})$ be a covariant k-tensor. The following are equivalent.*

(a) $\alpha \in \Lambda^k(V^*)$.
(b) $\alpha(v_1, \dots, v_k) = 0$ *whenever* $\{v_1, \dots, v_k\}$ *are linearly dependent.*
(c) $\alpha(v_1, \dots, v_k) = 0$ *whenever there are some* $i \neq j$ *with* $v_j = v_j$.

PROOF. Suppose (a) holds true. Then we know from Lemma 7.9 that $(i\,j) \cdot \alpha = -\alpha$, where $(i\,j)$ is the transposition of i and j (which is odd, so has $\mathrm{sgn} = -1$). This means that

$$\alpha(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -\alpha(v_1, \dots, v_j, \dots, v_i, \dots, v_k).$$

But if $v_i = v_j$, this says this value is equal to its negative, and so (c) holds true. It is also clear that (b) $\Longrightarrow$ (c) (since any list of vectors with a repeat is linearly dependent).

Conversely, suppose (c) holds true. By polarization, if we put $v_i + v_j$ into both the i and j slot, then on the one hand we get 0 (by part (c)), and on the other hand this yields

$$\begin{aligned} 0 = \alpha(v_1, \dots, v_i + v_j, \dots, v_i + v_j, \dots, v_k) &= \alpha(v_1, \dots, v_i, \dots, v_i, \dots, v_k) + \alpha(v_1, \dots, v_i, \dots, v_j, \dots, v_k) \\ &\quad + \alpha(v_1, \dots, v_j, \dots, v_i, \dots, v_k) + \alpha(v_1, \dots, v_j, \dots, v_j, \dots, v_k) \end{aligned}$$

by multilinearity. The first and last terms are 0 by (c), and so the remaining terms add to 0, which shows that $(i\,j) \cdot \alpha = -\alpha$ for any transposition $(i\,j)$. Any permutation is a product of transpositions, and the homomorphism property of sgn shows that $\sigma \cdot \alpha = \mathrm{sgn}\,(\sigma)\,\alpha$ in general, proving that $\alpha \in \Lambda^k(V)$, showing (a).

Finally, suppose $v_1, \dots, v_k$ are linearly independent. Then there is some vector v_j that is a linear combination of the others, $v_j = \sum_{\ell \neq j} a_\ell v_\ell$. Then we have, by multilinearity,

$$\alpha(v_1, \dots, v_j) = \alpha(v_1, \dots, \sum_{\ell \neq j} a_\ell v_\ell, \dots, v_k) = \sum_{\ell \neq j} a_\ell \alpha(v_1, \dots, v_\ell, \dots, v_k).$$

Each term in the sum has a repeated vector, and so by (c), the sum is 0, proving (b). □

EXAMPLE 7.12. The determinant $\det_n$ on $n \times n$ matrices, though of as a map on the n columns, is multilinear, and has properties (b) $\Longleftrightarrow$ (c) of the proposition. Therefore $\det_n \in \Lambda^n((\mathbb{R}^n)^*)$.

Let us make one more observation: if $\dim V < k$, then *any* list of k vectors in V must be linearly dependent. The proposition then shows us:

COROLLARY 7.13. $\Lambda^k(V^*) = 0$ *if* $\dim V < k$.

Hence, for any given finite-dimensional vector space V, only finitely many of the spaces $\Lambda^k(V^*)$ are non-trivial. (This is certainly not true of $\Sigma^k(V^*)$, which only grows in dimension as k grows, as the reader can work out.) The question we now address is: what is the dimension of $\Lambda^k(V^*)$ for $k \leq \dim(V)$? More generally, we will construct a basis of this space.

In fact, we will see that anti-symmetric tensors are spanned by "determinants"; but we must be a little looser about what we mean by determinant, which is only defined on a square matrix (meaning $\det_n$ requires n-arguments in an n-dimensional space). But we can take *sub*determinants, by choosing which rows to keep.

DEFINITION 7.14. *Let V be an n-dimensional vector space, and fix a basis $e^1, \dots, e^n$ with dual basis $(e^1)^*, \dots, (e^n)^*$. Let $1 \le k \le n$, and let $I = (i_1, \dots, i_k)$ be a multi-index with $1 \le i_1, \dots, i_k \le n$. Define a function $\det^I \colon V^k \to \mathbb{R}$ as follows:*

$$\det{}^I(v_1, \dots, v_k) = \det \begin{bmatrix} (e^{i_1})^*(v_1) & \cdots & (e^{i_1})^*(v_k) \\ \vdots & \ddots & \vdots \\ (e^{i_k})^*(v_1) & \cdots & (e^{i_k})^*(v_k) \end{bmatrix} = \det \begin{bmatrix} v_1^{i_1} & \cdots & v_k^{i_1} \\ \vdots & \ddots & \vdots \\ v_1^{i_k} & \cdots & v_k^{i_k} \end{bmatrix}.$$

That is: we write the vectors $v_1, \dots, v_k$ as columns with respect to the basis $e_1, \dots, e_n$, producing an $n \times k$ matrix; we then select the $k \times k$ submatrix with rows $i_1, i_2, \dots, i_k$, and take the determinant of this square matrix.

REMARK 7.15. The full determinant is an invariant of a linear operator: it does not depend on the basis chosen. This is not so for these subdeterminants in general – they are basis dependent.

EXAMPLE 7.16. Taking $|I| = 1$, we have $\det^i(v) = (e^i)^*(v)$; i.e. $\det^i = (e^i)^*$. On the other hand, if $I = (1, 2, \dots, n)$, then $\det^I = \det$. For an intermediate example, we have

$$\det{}^{(4,2)}\left(\begin{bmatrix} v^1 \\ v^2 \\ v^3 \\ v^4 \end{bmatrix}, \begin{bmatrix} w^1 \\ w^2 \\ w^3 \\ w^4 \end{bmatrix} \right) = \det \begin{bmatrix} v^4 & w^4 \\ v^2 & w^2 \end{bmatrix} = v^4 w^2 - v^2 w^4.$$

It is easy to check that $\det^I \in \Lambda^k(V^*)$ for any multi-index I of length k (for any given basis for V defining the determinant). In fact, these will turn out to be our basis tensors of $\Lambda^k(V^*)$. However, not all of them are linearly independent.

LEMMA 7.17. *Fix an n-dimensional vector space V, with basis $e^1, \dots, e^n$. Let $k \le n$ and let $I = (i_1, \dots, i_k)$ be a multi-index. Denote by $\sigma \cdot I = (i_{\sigma(1)}, \dots, i_{\sigma(k)})$. Then $\det^{\sigma \cdot I} = \operatorname{sgn}(\sigma) \det^I$. Moreover, if there exists a repeated index $i_\ell = i_m$ for some $1 \le \ell < m \le k$, then $\det^I = 0$.*

PROOF. We have $\det^{\sigma \cdot I}(v_1, \dots, v_k) = \det^I(v_{\sigma(1)}, \dots, v_{\sigma(k)})$, and since $\det^I$ is anti-symmetric, the first result follows from Lemma 7.9. For the second part, if I has a repeated index $i_\ell = i_m$, then $\det^I(v_1, \dots, v_k)$ is the determinant of a matrix with identical ℓth and mth rows, which is 0. $\square$

To avoid these degeneracies, we will generally work with *increasing* multi-indices: $I = (i_1, \dots, i_k)$ with $i_1 < i_2 < \cdots < i_k$. For short, we write this as $I \uparrow$. These turn out to give our basis.

PROPOSITION 7.18. *Let V be an n-dimensional real vector space. Fix a basis $e^1, \dots, e^n$ for V, and let $\det^I$ be the corresponding subdeterminants for any multi-indices I of lengths $\le n$. Then*

$$\beta(\Lambda^k(V^*)) \equiv \{\det{}^I \colon |I| = k, I \uparrow\}$$

is a basis for $\Lambda^k(V^*)$. In particular,

$$\dim \Lambda^k(V^*) = \binom{n}{k}$$

when $0 \le k \le n$, and 0 otherwise.

Note, included in this definition is the trivial case of $\Lambda^0(V^*) = \Sigma^0(V^*) = (V^*)^{\otimes 0}$ which is, by convention, simply $\mathbb{R}$ (i.e. a multilinear function of *no* variables is just a constant).

PROOF. That the set of increasing multi-indices of length k is counted by $\binom{n}{k}$ is a simple combinatorial exercise. So we must show that $\beta(\Lambda^k(V^*))$ is indeed a basis. To that end, fix any $\omega \in \Lambda^k(V^*)$. As usual, its tensor components are $\omega_{i_1,\dots,i_k} = \omega(e^{i_1},\dots,e^{i_k})$. Let us restrict this to increasing multi-indices, and so define ω_I for $I\uparrow$. Indeed, by Lemma 7.17, for any multi-index I, if I has repeated indices then $\omega_I = 0$; otherwise, I is a permutation of an increasing multi-index, which is related by an appropriate sgn. Now, fix any multi-index $J = (j_1,\dots,j_k)$; then

$$\sum_{I\uparrow} \omega_I \det{}^I(e^{j_1},\dots,e^{j_k}) = \sum_{I\uparrow} \omega_I \det \begin{bmatrix} \delta^{i_1}_{j_1} & \cdots & \delta^{i_1}_{j_k} \\ \vdots & \ddots & \vdots \\ \delta^{i_k}_{j_1} & \cdots & \delta^{i_k}_{j_k} \end{bmatrix}$$

If J has a repeated index, this is 0. If J is a permutation of I, $J = \sigma \cdot I$, then this determinant is equal to $\mathrm{sgn}(\sigma)\det{}^J(e^{j_1},\dots,e^{j_k}) = \mathrm{sgn}(\sigma)\det I_k = \mathrm{sgn}(\sigma)$. Otherwise, if J is not a permutation of I, then this square matrix has a row of 0s, and the determinant is 0. Now, for fixed J without repeated indices, there is precisely one increasing multi-index I which is a permutation σ of J. Hence, precisely one term $I = \sigma \cdot J$ in the sum survives, and we have

$$\begin{aligned}\sum_{I\uparrow} \omega_I \det{}^I(e^{j_1},\dots,e^{j_k}) &= \omega_{\sigma\cdot J}\mathrm{sgn}(\sigma) \\ &= \mathrm{sgn}(\sigma)\omega(e^{j_{\sigma(1)}},\dots,e^{j_{\sigma(k)}}) = \omega(e^{j_1},\dots,e^{j_k}) = \omega_J.\end{aligned}$$

Now expanding any vectors $v_1,\dots,v_k \in V$ in terms of the basis e^j and using multilinearity on both sides and expanding shows that $\sum_{I\uparrow}\omega_I\det{}^I(v_1,\dots,v_k) = \omega(v_1,\dots,v_k)$. Thus $\beta(\Lambda^k(V^*))$ is a spanning set.

On the other hand, suppose $\sum_{I\uparrow}\alpha_I\det{}^I = 0$ for some coefficients α_I. Applying this to any k-tuple $(e_{j_1},\dots,e_{j_k})$ with $j_1 < \cdots < j_k$, we pick out precisely one term in the sum, and this shows that $\alpha_{j_1,\dots,j_k} = 0$. As this holds for all increasing multi-indices, this proves that $\beta(\Lambda^k(V^*))$ is linearly independent, concluding the proof. □

4. Wedge Product and Exterior Algebra

Let ω and η be antisymmetric tensors over V, say $\omega \in \Lambda^k(V^*)$ and $\eta \in \Lambda^\ell(V^*)$. We could combine them via tensor product: $\omega\otimes\eta \in \Lambda^k(V^*)\otimes\Lambda^\ell(V^*) \subset (V^*)^{\otimes(k+\ell)}$. But the resulting tensor is not antisymmetric. For example, if $k = \ell = 1$, then antisymmetry of ω,η is vacuous: they are just linear functionals in V^*, but $\omega\otimes\eta$ is then not antisymmetric (unless one of ω and η is 0). We can fix this by applying the Alt projection: so we could define an antisymmetric product by $(\omega,\eta)\mapsto \mathrm{Alt}(\omega\otimes\eta)$. Some authors do use this convention; for our purposes, it will be useful to modify it with a different normalization.

DEFINITION 7.19. *Let $\omega \in \Lambda^k(V^*)$ and $\eta \in \Lambda^\ell(V^*)$. Their* **wedge product** *$\omega\wedge\eta$ is the element of $\Lambda^{k+\ell}(V^*)$ defined by*

$$\omega\wedge\eta \equiv \frac{(k+\ell)!}{k!\ell!}\mathrm{Alt}(\omega\otimes\eta) = \frac{1}{k!\ell!}\sum_{\sigma\in S_{k+\ell}} \mathrm{sgn}(\sigma)\,\sigma\cdot(\omega\otimes\eta).$$

The reason for this convention is the following.

PROPOSITION 7.20. *Let V be an n-dimensional real vector space, with given basis defining the subdeterminant basis $\{\det^I : |I| = k, I \uparrow\}$ for $\Lambda^k(V^*)$ for $k \leq n$. Then for any multi-indices I, J of lengths k, ℓ,*

$$\det^I \wedge \det^J = \det^{IJ}$$

where IJ is the concatenation: if $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_\ell)$, then $IJ = (i_1, \ldots, i_k, j_1, \ldots, j_\ell)$.

The proof of Proposition 7.20 is elementary, but tedious and computational. The interested reader can consult [**3**, Lemma 14.10]. Note, if we had not changed the normalization and had simply defined the wedge product as $\mathrm{Alt}(\omega \otimes \eta)$, then we would have the less appealing formula

$$\mathrm{Alt}(\det^I \otimes \det^J) = \frac{k!\ell!}{(k+\ell)!}\det^{IJ}.$$

For computational purposes, the convention we've chosen is far more convenient.

Here are some immediate consequences, giving basic properties of the wedge product.

COROLLARY 7.21. *Let $\omega, \omega', \eta, \eta', \xi$ be antisymmetric tensors over V, and let $a, a' \in \mathbb{R}$. Fix a basis $\{e^i : 1 \leq i \leq n\}$ for V, giving associated subdeterminant tensors $\det^I$.*

(a) BILINEARITY*:*

$$(a\omega + a'\omega') \wedge \eta = a\omega \wedge \eta + a'\omega' \wedge \eta$$
$$\eta \wedge (a\omega + a'\omega') = a\eta \wedge \omega + a'\eta \wedge \omega'.$$

(b) ASSOCIATIVITY*:* $\omega \wedge (\eta \wedge \xi) = (\omega \wedge \eta) \wedge \xi$.

(c) ANTICOMMUTATIVITY*: For $\omega \in \Lambda^k(V^*)$ and $\eta \in \Lambda^\ell(V^*)$,*

$$\omega \wedge \eta = (-1)^{k\ell}\eta \wedge \omega.$$

(d) *For any indices $i_1, \ldots, i_k \in \{1, \ldots, n\}$, $(e^{i_1})^* \wedge \cdots \wedge (e^{i_k})^* = \det^{i_1,\ldots,i_k}$.*

(e) *More generally, if $\lambda^1, \ldots, \lambda^k$ are any elements of V^* and $v_1, \ldots, v_k \in V$, then*

$$\lambda^1 \wedge \cdots \wedge \lambda^k(v_1, \ldots, v_k) = \det[\lambda^j(v_i)]_{1 \leq i,j \leq n}.$$

The proof is, again, simple and computational (and not even very tedious). For example, (c) (anticommutativity) is immediate for $k = \ell = 1$ (by the definition of anticommutativity), and then follows in general by tracking how many transpositions are requires to shuffle IJ into JI for $|I| = k$ and $|J| = \ell$ (the answer is $k\ell$). It is also worth noting that properties (a)–(d) uniquely define $\wedge$; in fact, antisymmetric tensors are often built up axiomatically that way, without any mention of determinants. We have presented this in a more concrete manner, highlighting what these things really *are* (determinants); this will be important for understanding why and how we will use them.

The wedge product is also known as the **exterior product**. There is also an **interior product**, also known as **contraction**. Given a vector $v \in V$, the map $\iota_v : \Lambda^k(V^*) \to \Lambda^{k-1}(V^*)$ is given by

$$(\iota_v\omega)(v_1, \ldots, v_{k-1}) = \omega(v, v_1, \ldots, v_{k-1}).$$

Note: this makes sense for more general tensors, but it is usually only considered for antisymmetric (or symmetric) tensors, where plugging into the first slot is somehow not special. (We could define ι_v plugging into any given slot, and the only difference in the definition would be a ± 1.) Another common notation is

$$v \lrcorner\, \omega = \iota_v\omega.$$

Here are two computational properties for contractions.

LEMMA 7.22. *Let V be a finite-dimensional real vector space, and $v \in V$.*

(a) *For any $k \geq 2$, the operator $\iota_v \circ \iota_v \colon \Lambda^k(V^*) \to \Lambda^{k-2}(V^*)$ is identically 0.*
(b) *If $\omega \in \Lambda^k(V^*)$ and $\eta \in \Lambda^\ell(V^*)$, then*

$$v \lrcorner (\omega \wedge \eta) = (v \lrcorner \omega) \wedge \eta + (-1)^k \omega \wedge (v \lrcorner \eta).$$

(c) *More generally, if $\lambda^1, \dots, \lambda^k \in V^*$, then*

$$v \lrcorner (\lambda^1 \wedge \cdots \wedge \omega^k) = \sum_{j=1}^{k} (-1)^{j-1} \lambda^j(v) \lambda^1 \wedge \cdots \wedge \widehat{\lambda^j} \wedge \cdots \wedge \lambda^k$$

where the hat denotes a removed term from the product.

Again, the proof is just tedious computation. (Part (a) is immediate from the fact that antisymmetric tensors yield 0 when applied to repeated vectors.)

We have seen that $\Lambda^k(V^*)$ is only non-trivial when $0 \leq k \leq n = \dim V$. We can string these $n+1$ spaces together into a single object, the **exterior algebra**:

$$\Lambda(V^*) \equiv \bigoplus_{k=0}^{n} \Lambda^k(V^*).$$

It is an algebra under the operation $\wedge$; Corollary 7.21 shows that $\wedge$ distributes over $+$ and is associative. It is not a commutative algebra, but it is the next best thing. It is defined as a direct sum of subspaces Λ^k that have the property that $\Lambda^k \wedge \Lambda^\ell \subseteq \Lambda^{k+\ell}$. Such an algebra is called **graded**. A graded algebra is called **anticommutative** if $\omega \in \Lambda^k$ and $\eta \in \Lambda^\ell$ implies that $\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \lambda$, precisely property (c) of Corollary 7.21. Thus $\Lambda(V^*)$ is an anticommutative graded algebra. Its dimension is

$$\dim \Lambda(V^*) = \sum_{k=0}^{n} \dim \Lambda^k(V^*) = \sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Note that, for any covariant vector $\lambda \in V^* \subset \Lambda(V)$, we have $\lambda \wedge \lambda = 0$. As with our construction of tensors, we can use a property like this to abstractly define $\Lambda(V^*)$ via a universal property. In general, given a real vector space U, the **Grassmann algebra** $\Lambda(U)$ over U (together with the canonical inclusion map $U \hookrightarrow \Lambda(U)$) is the unique (up to canonical isomorphism) algebra with the following universal property: if A is any algebra, and $f \colon V \to A$ is any linear map with the property that $f(u)^2 = 0$ for all $u \in U$, then f factors through a unique algebra homomorphism $\widetilde{f} \colon \Lambda(U) \to A$:

$$\begin{tikzcd} U \arrow[r, "f"] \arrow[d, hook] & A \\ \Lambda(U) \arrow[ur, "\widetilde{f}"'] & \end{tikzcd} .$$

With this definition, our construction of the exterior algebra gives a concrete realization of the Grassmann algebra $\Lambda(V^*)$ (and is the best way to see that in general $\Lambda(U)$ has dimension $2^{\dim U}$).

CHAPTER 8

Tensor Fields and Differential Forms on Manifolds

1. Covariant Tensor Bundle, and Tensor Fields

Let M be a smooth manifold, and let k be a non-negative integer. The **rank-k covariant tensor bundle** over M, denoted $T^k(T^*M)$, is the union of all spaces of tensors over all cotangent spaces:

$$T^k(T^*M) = \bigsqcup_{p\in M} (T_p^*M)^{\otimes k} = \bigsqcup_{p\in M} L((T_pM)^k;\mathbb{R}).$$

In particular, $T^1(T^*M) = T^*M$. We have the usual projection $\pi\colon T^k(T^*M) \to M$ given by $\pi(F_p) = p$. As usual, this bundle is not just a set, but has a smooth structure, which we define in the same way we did for TM and T^*M. In particular, fix a chart $(U, \varphi = (x^j)_{j=1}^n)$ in M and some point $p \in U$. Then we know that $\{dx^j|_p\}_{j=1}^n$ is the basis dual to $\{\frac{\partial}{\partial x^j}|_p\}_{j=1}^n$ for T_pM. Then by Lemma 7.3

$$\{dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p\colon 1 \le i_1, \dots, i_k \le n\}$$

is a basis for $(T_p^*M)^{\otimes k}$, meaning that any tensor $F(p) \in (T_p^*M)^{\otimes k}$ can be expanded uniquely as

$$F(p) = \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k}(p)\, dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p,$$

where $F_{i_1,\dots,i_k}(p) = F_p\left(\frac{\partial}{\partial x^{i_1}}\big|_p, \dots, \frac{\partial}{\partial x^{i_k}}\big|_p\right)$. This lets us define a chart $(\widetilde{U}, \widetilde{\varphi})$ for $T^k(T^*M)$ in the usual way: $\widetilde{U} = \pi^{-1}(U)$, and $\widetilde{\varphi}\colon \widetilde{U} \to \mathbb{R}^n \times (\mathbb{R}^n)^k$ is given by

$$\widetilde{\varphi}\left(\sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k}(p)\, dx^{i_1}|_p \otimes \cdots \otimes dx^{i_k}|_p\right) = (\varphi(p), \{F_{i_1,\dots,i_k}(p)\}_{1\le i_1,\dots,i_k\le n}).$$

One can mimic the calculations of Section 1 to see how components change under change of variables, and easily prove that the transition maps are smooth, thus giving a smooth structure to $T^k(T^*m)$.

As usual, a **section** of the bundle $T^k(T^*M)$ is a map $F\colon M \to T^k(T^*M)$ with the property that $\pi \circ F = \mathrm{Id}_M$, i.e. $F(p) \in T^k(T_p^*M)$ for each p. Such sections are called (rough) **tensor fields**. As M and $T^k(T^*M)$ are smooth manifolds, we can insist that a rought tensor field F be smooth by insisting it is a smooth map. As we did with covariant vectors (and vectors before that), this notion of smoothness can be characterized more easily as follows.

PROPOSITION 8.1. *Let M be a smooth manifold, and let $F\colon M \to T^k(T^*M)$ be a (rough) tensor field. The following are equivalent.*

(a) *F is smooth.*
(b) *In every chart, the coordinate functions $p \mapsto F_{i_1,\dots,i_k}(p)$ are smooth functions.*
(c) *If $X_1, \dots, X_k \in \mathscr{X}(M)$ are smooth vector fields, then the function $F(X_1, \dots, X_k)\colon M \to \mathbb{R}$ defined by*

$$F(X_1, \dots, X_k)(p) = F_p(X_1|_p, \dots, X_k|_p)$$

is smooth.

This, we can think of a tensor field as a function from k vector fields to $\mathrm{Fun}(M, \mathbb{R})$, and it is smooth iff this function is in $C^\infty(M)$. The proof is routine and left to the reader. We denote the space of rank-k smooth tensor fields by $\mathscr{T}^k(M)$.

Tensor operations are also well-behaved with respect to smoothness.

PROPOSITION 8.2. *Let M be a smooth manifold, and let $F \in \mathscr{T}^k(M)$ and $G \in \mathscr{T}^\ell(M)$. Let $f\colon M \to \mathbb{R}$, and define new (rough) tensor fields fF and $F \otimes G$ by $(fF)(p) = f(p)F(p)$ and $(F \otimes G)(p) = F_p \otimes G_p$. Then if $f \in C^\infty(M)$, $fF \in \mathscr{T}^k(M)$, and $F \otimes G \in \mathscr{T}^{k+\ell}(M)$.*

The proofs are simple computations in local coordinates. Note that, in any local coordinates, we simply have

$$(F \otimes G)_{i_1,\dots,i_k,i_{k+1},\dots,i_{k+\ell}} = F_{i_1,\dots,i_k} G_{i_{k+1},\dots,i_{k+\ell}}.$$

Viewing a smooth tensor field $F \in \mathscr{T}^k(M)$ as a map from $\mathscr{X}(M)^k \to C^\infty(M)$, we note immediately that it is a multilinear map over $\mathbb{R}$. In fact, it is multilinear over $C^\infty(M)$: for $X_1, \dots, X_j, X_j', \dots, X_k \in \mathscr{X}(M)$ and $f, f' \in C^\infty(M)$,

$$\begin{aligned}
&F(X_1, \dots, fX_j + f'X_j', \dots, X_k)(p) \\
&= F(X_1|_p, \dots, f(p)X_j|p + f'(p)X_j'|_p, \dots, X_k|_p) \\
&= f(p)F(X_1|_p, \dots, X_j|_p, \dots, X_k|_p) + f'(p)F(X_1|_p, \dots, X_j'|_p, \dots, X_k|_p) \\
&= [fF(X_1, \dots, X_j, \dots, X_k) + f'(X_1, \dots, X_j', \dots, X_k)](p).
\end{aligned}$$

In fact, this characterizes smooth tensor fields in this dual picture.

LEMMA 8.3. *Let $\mathcal{F}\colon \mathscr{X}(M)^k \to C^\infty(M)$ be a function. It is induced by a smooth tensor field $F \in \mathscr{T}^k(M)$, $\mathcal{F}(X_1, \dots, X_k)(p) = F(p)(X_1|_p, \dots, X_k|_p)$, if an only if $\mathcal{F}$ is multilinear over $C^\infty(M)$.*

PROOF. The preceding discussion shows that 'only if' direction. So, assume $\mathcal{F}$ is multilinear over $C^\infty(M)$. The idea is to define a (rough) tensor field F as follows: for each $p \in M$ and vectors $v_1(p), \dots, v_k(p) \in T_pM$, let

$$F_p(v_1(p), \dots, v_k(p)) \equiv \mathcal{F}(X_1, \dots, X_k)(p)$$

where, for each i, X_i is *any* smooth vector field with $X_i|_p = v_i(p)$. Provided this is well defined, it follows that F_p is a covariant tensor since $\mathscr{F}$ is multilinear, and it is smooth by Proposition 8.1. So what we need to show is that this is well-defined: i.e. multilinearity over $C^\infty(M)$ must imply that $\mathscr{F}$ acts pointwise, meaning that if $X_i, Yij \in \mathscr{X}(M)$ satisfy $X_i|_p = Y_i|_p$, then $\mathcal{F}(X_1, \dots, X_i, \dots, X_k) = \mathscr{F}(X_1, \dots, Y_i, \dots, X_k)$ for each slot i).

First, suppose that X_i vanishes on some open neighborhood U of p. Let ψ be a smooth bump function supported in U with $\psi(p) = 1$. Then $\psi X_j \equiv 0$, and by linearity of $\mathcal{F}$ in the jth slot, we have

$$0 = \mathcal{F}(X_1, \dots, 0, \dots, X_k)(p) = \mathcal{F}(X_1, \dots, \psi X_j, \dots, X_k)(p) = \psi(p)\mathcal{F}(X_1, \dots, X_j, \dots, X_k)$$

which shows that $\mathcal{F}(X_1, \dots, X_k) = 0$. In particular, this shows that $\mathcal{F}(X_1, \dots, X_k)(p)$ only depends on the vector fields $X_1, \dots, X_k$ behaviour in any arbitrarily small neighborhood of the point p. This is not quite enough though: we need to show that it only depends on the behavior *at* p.

So, assume that $X_i|_p = 0$. Then in any coordinate chart $(U, (x^j))$, we can write $X_i = \sum_{j=1}^n X_i^j \frac{\partial}{\partial x^j}$ where all the components X_i^j vanish at p. By using a partition of unity subordinate to a cover including U, we can extend the coordinate vector fields $\frac{\partial}{\partial x^j}$ to global vector fields

E_j on M (that agree with $\frac{\partial}{\partial x^j}$ on U), and similarly we can extend the functions X_i^j on U to global smooth functions f_i^j on M. Thus $X_i|_U = \left(\sum_{j=1}^n f_i^j E_j\right)\Big|_U$. By the preceding argument, $\mathcal{F}(p)$ is determined by its action on U, and so, using multilinearity, we have

$$\begin{aligned}\mathcal{F}(X_1,\dots,X_i,\dots,X_k)(p) &= \mathcal{F}(X_1,\dots,\sum_{j=1}^n f_i^j E_j,\dots,X_k)(p)\\ &= \sum_{j=1}^n f_i^j(p)\mathcal{F}(X_1,\dots,E_j,\dots,X_k)(p) = 0.\end{aligned}$$

Hence, $\mathcal{F}$ truly acts pointwise, concluding the proof. □

2. Pullbacks and Lie Derivatives of Tensor Fields

Let M, N be smooth manifolds, with a smooth map $F\colon M \to N$. We have already discussed (in Section 4) how to pull-back a covariant vector field on N to one on M using F. The exact same procedure works for general covariant tensor fields of any rank. If $A \in \mathscr{T}^k(N)$ is a covariant tensor field, then we define a new (rough) tensor field $F^*A \in \mathscr{T}^k(M)$ as follows:

$$(F^*A)_p(X_1|_p,\dots,X_k|_p) = A_{F(p)}(dF_p(X_1|_p),\dots,dF_p(X_k|_p)).$$

This is called the **pull back** of A along F. Following almost exactly the same calculations we did in the $k = 1$ case in Section 4, and similar computations, we have the following basic properties for pull backs.

PROPOSITION 8.4. *Let M, N, P be smooth manifolds and let $M \xrightarrow{F} N \xrightarrow{G} P$ be a smooth maps. Let A and B be (rough) covariant tensor fields on N, and let $f\colon N \to \mathbb{R}$.*

(a) $F^*(fA) = (f \circ F)F^*A$.
(b) $F^*(A \otimes B) = F^*A \otimes F^*B$.
(c) $F^*(A + B) = F^*A + F^*B$.
(d) *If A is a smooth tensor field on N, then F^*A is a smooth tensor field on M.*
(e) $(G \circ F)^*A = F^*(G^*A)$.
(f) $(\mathrm{Id}_N)^*A = A$.

The elementary proofs are left to the reader.

REMARK 8.5. Consistent with defining $V^{\otimes 0} \equiv \mathbb{R}$, we think of functions $f\colon N \to \mathbb{R}$ as rank-0 covariant tensor fields. In that context, the only reasonable definition for pull back is $F^*f = f \circ F$ (a convention we have already used). Moreover, We should extend the tensor notation to make sense of $f \otimes A$ by $f \otimes A = fA$. Thus, in the above Proposition, (a) is really jusy a special case of (b).

This proposition tells us everything we need to know to compute pull backs, in local coordinates. For example, together with Lemma 6.15, it gives us the following immediate corollary.

COROLLARY 8.6. *Let $F\colon M \to N$ be smooth, and let A be a rank-k covariant tensor field on N. Let $p \in M$, and let $(U, (y^j))$ be a coordinate chart in N around $F(p)$. Then we can write A in U as $A|_U = \sum_{i_1,\dots,i_k} A_{i_1,\dots,i_k} dy^{i_1} \otimes \cdots \otimes dy^{i_k}$. The pull back along F is then given by*

$$F^*(A|_U) = \sum_{i_1,\dots,i_k} (A_{i_1,\dots,i_k} \circ F) d(y^{i_k} \circ F) \otimes \cdots \otimes d(y^{i_k} \circ F).$$

EXAMPLE 8.7. Let $M = \{(r,\theta)\colon r > 0, -\frac{\pi}{2} < \theta < \frac{\pi}{2}\}$, and let $N = \{(x,y)\colon x > 0\}$. Then the polar coordinate map $F(r,\theta) = (r\cos\theta, r\sin\theta)$ is a smooth map (in fact diffeomorphism) from M onto N. Let $A = \frac{1}{x^2}dy \otimes dy$, a smooth rank-2 covariant tensor field on N. We compute the pull back:

$$\begin{aligned} F^*A &= \frac{1}{(r\cos\theta)^2} d(r\sin\theta) \otimes d(r\sin\theta) \\ &= \frac{1}{(r\cos\theta)^2}(\sin\theta\, dr + r\cos\theta\, d\theta) \otimes (\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= \frac{1}{(r\cos\theta)^2}(\sin^2\theta\, dr \otimes dr + r\sin\theta\cos\theta(dr\otimes d\theta + d\theta \otimes dr) + r^2\cos^2\theta\, d\theta\otimes d\theta) \\ &= \frac{\tan^2\theta}{r^2}\, dr\otimes dr + \frac{\tan\theta}{r}(dr\otimes d\theta + d\theta\otimes dr) + d\theta\otimes d\theta. \end{aligned}$$

Having pull backs in hand, we can now define Lie derivatives of covariant tensor fields, as in Section 4. Let $X \in \mathscr{X}(M)$, and let θ be the flow of X. If X is complete, then θ_t is a diffeomorphism of M for each $t \in \mathbb{R}$, and so we can define θ_t^*A as above for any covariant tensor field A. If X is not complete, we can still do this near any point: we know that, for any given $p \in M$, for all t sufficiently close to 0, θ_t is a diffeomorphism from a neighborhood of p onto a neighborhood of $\theta_t(p)$, and so it makes perfect sense to define

$$(\theta_t^*A)_p(X_1|_p, \dots, X_k|_p) = A_{\theta_t(p)}(d(\theta_t)_p(X_1), \dots, d(\theta_t)_p(X_k))$$

as in the complete case, for small enough t. The size of allowed t might get arbitrarily small as p varies, but that doesn't matter: our definition will be for each fixed p. We define, as usual,

$$(\mathscr{L}_X A)_p = \left.\frac{d}{dt}\right|_{t=0} (\theta_t^*A)_p = \lim_{t\to 0} \frac{d(\theta_t)_p^*(A_{\theta_t(p)}) - A_p}{t}.$$

To be clear: this is an ordinary calculus derivative in the vector space $(T_p^*M)^{\otimes k}$ where $(\theta_t^*A)_p$ lives for all small t. This limit, should it exist, therefore defines an element of $(T_p^*M)^{\otimes k}$, and we therefore have a rough tensor field (of the same rank), provided the limit exists for all p. In fact it does, as a (somewhat involved) calculation like the one in Proposition 5.24 shows. We have the following.

PROPOSITION 8.8. *Let $A \in \mathscr{T}^k(M)$ be a smooth covariant tensor field on a smooth manifold M. Then the following product rules hold for the Lie derivative $\mathscr{L}_X$ with respect to some smooth vector field $X \in \mathscr{X}(M)$.*

(a) *For $f \in C^\infty(M)$, $\mathscr{L}_X(fA) = (\mathscr{L}_X f)A + f\mathscr{L}_X A = X(f)A + f\mathscr{L}_X A$.*
(b) *If B is another smooth covariant tensor field, then $\mathscr{L}_X(A\otimes B) = (\mathscr{L}_X A)\otimes B + A\otimes \mathscr{L}_X B$.*
(c) *If $X_1, \dots, X_k \in \mathscr{X}(M)$, then*

$$\mathscr{L}_X\left(A(X_1,\dots,X_k)\right) = (\mathscr{L}_X A)(X_1,\dots,X_k) + A(\mathscr{L}_X X_1, \dots, X_k) + \dots + A(X_1,\dots,\mathscr{L}_X X_k).$$

Item (c) shows us how to actually calculate the Lie derivative of a tensor field. The left-hand term is just X applied to the smooth function $A(X_1,\dots,X_k)$; all the Lie derivatives inside on the right are Lie brackets; and so we have the formula

$$(\mathscr{L}_X A)(X_1,\dots,X_k) = X(A(X_1,\dots,X_k)) - A([X,X_1],\dots,X_k) - \dots - A(X_1,\dots,[X,X_k]).$$

This is consistent with Proposition 6.20 (which is the $k = 1$ special case).

EXAMPLE 8.9. Let $A \in \mathscr{T}^2(M)$ and $X \in \mathscr{X}(M)$ be arbitrary. In local coordinates $(U, (x^j))$, write $A = \sum_{i,j} A_{ij} dx^i \otimes dx^j$, and $X = \sum_i X^i \frac{\partial}{\partial x^i}$. Using Corollary 6.21, we have

$$\mathscr{L}_X(dx^i) = d(\mathscr{L}_X x^i) = d(X(x^i)) = dX^i.$$

Thus, applying Proposition 8.8, we have

$$\begin{aligned}
\mathscr{L}_X A &= \sum_{i,j} \mathscr{L}_X(A_{ij} dx^i \otimes dx^j) \\
&= \sum_{i,j} \left(\mathscr{L}_X(A_{ij}) dx^i \otimes dx^j + A_{ij} dX^i \otimes dx^j + A_{ij} dx^i \otimes dX^j \right) \\
&= \sum_{i,j} \left(X(A_{ij}) dx^i \otimes dx^j + A_{ij} \sum_k \left(\frac{\partial X^i}{\partial x^k} dx^k \otimes dx^j + \frac{\partial X^j}{\partial x^k} dx^i \otimes dx^k \right) \right) \\
&= \sum_{i,j,k} \left(X^k \frac{\partial A_{ij}}{\partial x^k} dx^i \otimes dx^j + A_{ij} \frac{\partial X^i}{\partial x^k} dx^k \otimes dx^j + A_{ij} \frac{\partial X^j}{\partial x^k} dx^i \otimes dx^k \right).
\end{aligned}$$

Reindexing the three sums accordingly, we can express this a little more simply as

$$\mathscr{L}_X A = \sum_{i,j,k} \left(X^k \frac{\partial A_{ij}}{\partial x^k} + A_{kj} \frac{\partial X^k}{\partial x^i} + A_{ik} \frac{\partial X^k}{\partial x^j} \right) dx^i \otimes dx^j.$$

3. Differential Forms

A **differential k-form** on a manifold M is a smooth covariant k-tensor field $\omega \in \mathscr{T}^k(M)$ that is antisymmetric at each point. I.e., for each $p \in M$, $\omega_p \in \Lambda^k(T^*M)$. The set of all differential k-forms is denoted $\Omega^k(M)$. As usual, we can do vector space operations pointwise on sections, and so we can take the "direct sum" of these to get a total space

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

where $\Omega^0(M) = \mathscr{T}^0(M) = C^\infty(M)$. We can also define the wedge product on the space $\Omega^*(M)$ pointwise:

$$(\omega \wedge \eta)(p) = \omega_p \wedge \eta_p.$$

This makes $\Omega^*(M)$ into an anticommutative graded algebra (this time infinite dimensional). If $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$, then $\omega \wedge \eta \in \Omega^{k+\ell}(M)$. If $k + \ell > \dim M$ then it follows that $\omega \wedge \eta = 0$.

Let $(U, \varphi = (x^j)_{j=1}^n)$ be a chart at $p \in M$. We know that $\{dx^{i_1}|_p \wedge \cdots \wedge dx^{i_k}|p \colon 1 \leq i_1 < \cdots < i_k \leq n\}$ is a basis for $\Lambda^k(T_p^*M)$, and that the components of the tensor ω in this basis are

$$\omega_p = \sum_{i_1,\ldots,i_k} \omega_p \left(\left.\frac{\partial}{\partial x^{i_1}}\right|_p, \ldots, \left.\frac{\partial}{\partial x^{i_k}}\right|_p \right) dx^{i_1}|_p \wedge \cdots \wedge dx^{i_k}|p.$$

To simplify this notation, we use multi-indices $I = (i_1, \ldots, i_k)$, and so we have

$$\omega|_U = \sum_{I\uparrow} \omega_I dx^I = \sum_{I\uparrow} \omega \left(\frac{\partial}{\partial x^I} \right) dx^I$$

where

$$\frac{\partial}{\partial x^I} = \left(\frac{\partial}{\partial x^{i_1}}, \dots, \frac{\partial}{\partial x^{i_k}}\right), \quad \text{and} \quad dx^I = dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

EXAMPLE 8.10. $\omega = \sin(xy)dy \wedge dz$ is a 2-form on $\mathbb{R}^2$. More generally, *any* 2-form on $\mathbb{R}^2$ can be written in the form

$$\eta = f_3\, dx \wedge dy - f_2\, dx \wedge dz + f_1\, dy \wedge dz, \qquad f_1, f_2, f_3 \in C^\infty(\mathbb{R}^3).$$

The index and sign convention above will be explained in a little bit. As for 3-forms, since $\dim\Lambda^3((\mathbb{R}^3)^*) = \binom{3}{3} = 1$, one can always take $dx \wedge dy \wedge dz$ as the basis element, so any 3-form looks like $f\, dx \wedge dy \wedge dz$ for some $f \in C^\infty(\mathbb{R}^3)$.

As tensor fields, differential forms pullback to tensor fields; in fact, they pullback to differential forms.

LEMMA 8.11. *Let $F\colon M \to N$ be smooth. Then $F^*\colon \Omega^k(N) \to \Omega^k(M)$. Moreover, if $\omega, \eta \in \Omega^*(N)$, then*

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

In a chart $(U, (y^j))$ in N,

$$F^*\left(\sum_{I\uparrow} \omega_I dy^I\right) = \sum_{I\uparrow} (\omega_I \circ F) d(y^{i_1} \circ F) \wedge \cdots \wedge d(y^{i_k} \circ F).$$

All of these properties follow from the linearity of F^*, and the fact that the the projection from $(T^*M)^{\otimes k} \to \Lambda^k(T^*M)$ is linear. Boring details are left to the reader.

EXAMPLE 8.12. Let $M = N = \mathbb{R}^2$, and let $F\colon M \to N$ be the smooth map $F(r, \theta) = (r\cos\theta, r\sin\theta)$. This time, consider the form $\omega = dx \wedge dy$. Then

$$\begin{aligned} F^*\omega &= d(r\cos\theta) \wedge d(r\sin\theta) \\ &= (\cos\theta\, dr - r\sin\theta\, d\theta) \wedge (\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= (\cos\theta)(r\cos\theta)\, dr \wedge d\theta + (-r\sin\theta)(\sin\theta)\, d\theta \wedge dr \\ &= r\, dr \wedge d\theta. \end{aligned}$$

Note: if we restrict to the right half-plane for N as in Example 8.7, then we could think of F is the identity map written in terms of local coordinates (r, θ) in the domain and (x, y) in the range. In this interpretation, the statement is $dx \wedge dy = r\, dr \wedge d\theta$.

This should be very reminiscent of the change of variables formula for integrating in polar coordinates. That's no accident. In fact, said change of variables formula is actually a statement about pullbacks of top-degree forms.

PROPOSITION 8.13. *Let $F\colon M \to N$ be a smooth map between n-dimensional manifolds. Let $(U, (x^i))$ be a chart in M and $(V, (y^j))$ a chart in N. Let $f\colon V \to \mathbb{R}$ be continuous. Then on $U \cap F^{-1}(V)$, we have*

$$F^*(f\, dy^1 \wedge \cdots \wedge dy^n) = (f \circ F)(\det[DF])\, dx^1 \wedge \cdots \wedge dx^n$$

where $[DF]$ denotes the Jacobian matrix of F in local coordinates.

PROOF. By multilinearity and antisymmetry, it suffices to show that both sides agree when applied to $(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n})$. From Lemma 8.11, we have

$$F^*(f\, dy^1 \wedge \cdots \wedge dy^n) = (f \circ F)\, dF^1 \wedge \cdots \wedge dF^n$$

where $F^j = y^j \circ F$ are the components of F in the y-coordinates. Now, from Corollary 7.21,

$$dF^1 \wedge \cdots \wedge dF^n \left(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right) = \det\left[dF^j\left(\frac{\partial}{\partial x^i}\right)\right]_{1\le i,j\le n} = \det\left[\frac{\partial F^j}{\partial x^i}\right]_{1\le i,j\le n} = \det[DF].$$

This proves the result, since, on the left-hand-side, $dx^1 \wedge \cdots \wedge dx^n(\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n}) = 1$. □

This shows us how top-forms change under smooth change of coordinates.

COROLLARY 8.14. *Let $(U, (x^i))$ and $(V, (y^j))$ be overlapping charts in M. Then on $U \cap V$,*

$$dy^1 \wedge \cdots \wedge dy^n = \det\left[\frac{\partial y^j}{\partial x^i}\right]_{1\le i,j\le n} dx^1 \wedge \cdots \wedge dx^n.$$

Finally, let's note that interior multiplication also preserves $\Omega^*(M)$.

LEMMA 8.15. *Let $\omega \in \Omega^k(M)$ and $X \in \mathscr{X}(M)$. Then the contraction $X \lrcorner \omega$ defined pointwise by*

$$(X \lrcorner \omega)_p = X_p \lrcorner \omega_p$$

is in $\Omega^{k-1}(M)$.

PROOF. We know that $X_p \lrcorner \omega_p \in \Lambda^{k-1}(T_p^*M)$ for each $p \in M$, following the definition (preceding Lemma 7.22). So it only remains to check that $X \lrcorner \omega$ is smooth. For any $k-1$ vector fields $X_1, \dots, X_k \in \mathscr{X}(M)$,

$$(X \lrcorner \omega)(X_1, \dots, X_{k-1}) = \omega(X, X_1, \dots, X_{k-1})$$

and this is smooth since ω is smooth, cf. Lemma 8.3. □

EXAMPLE 8.16. Let $X \in \mathscr{X}(\mathbb{R}^3)$, $X = \sum_{j=1}^3 X^j \frac{\partial}{\partial x^j}$. Then the contraction $X \lrcorner (dx^1 \wedge dx^2 \wedge dx^3)$ is a 2-form on $\mathbb{R}^3$. We compute, for any vector fields $Y, Z \in \mathscr{X}(M)$,

$$X \lrcorner (dx^1 \wedge dx^2 \wedge dx^3)(Y, Z) = \sum_{j=1}^3 X^j (dx^1 \wedge dx^2 \wedge dx^3)\left(\frac{\partial}{\partial x^j}, Y, Z\right).$$

Writing Y and Z in terms of their components in these coordinates, each of these three expressions is a determinant

$$\det\begin{bmatrix} & Y^1 & Z^1 \\ e^j & Y^2 & Z^2 \\ & Y^3 & Z^3 \end{bmatrix}.$$

expanding along the first column, we therefore have

$$\begin{aligned} &X^1 \det\begin{bmatrix} Y^2 & Z^2 \\ Y^3 & Z^3 \end{bmatrix} - X^2 \det\begin{bmatrix} Y^1 & Z^1 \\ Y^3 & Z^3 \end{bmatrix} + X^3 \det\begin{bmatrix} Y^1 & Z^1 \\ Y^2 & Z^2 \end{bmatrix} \\ =&X^1\, dx^2 \wedge dx^3(Y, Z) - X^2\, dx^1 \wedge dx^3 + X^3\, dx^1 \wedge dx^2(Y, Z). \end{aligned}$$

Thus, we have

$$X \lrcorner (dx^1 \wedge dx^2 \wedge dx^3) = X^3\, dx^1 \wedge dx^2 - X^2 dx^1 \wedge dx^3 + X^1\, dx^2 \wedge dx^3.$$

This explains our seemingly odd choice of index and sign convention in Example 8.10. Note this shows that *any* 2-form on $\mathbb{R}^3$ can be uniquely represented in the form $X \lrcorner (dx^1 \wedge dx^2 \wedge dx^3)$ for some vector field; so $(\cdot)\lrcorner(dx^1 \wedge dx^2 \wedge dx^3)\colon \mathscr{X}(\mathbb{R}^3) \to \Omega^2(\mathbb{R}^3)$ is actually an isomorphism. (You can also check that it is linear over $C^\infty(\mathbb{R}^3)$, not just $\mathbb{R}$.) This will be important in our understanding of how the technology we're currently developing relates to the usual technology of vector calculus.

4. Exterior Derivatives

Let us briefly return to Section 6, where we discussed exact vs. closed 1-forms. Recall that a 1-form λ is *exact* if it is of the form $\lambda = df$ for some smooth function f. The form is *closed* if, in any chart $(U, (x^j))$, writing $\lambda = \sum_j \lambda_j dx^j$, we have

$$\frac{\partial \lambda_j}{\partial x^i} = \frac{\partial \lambda_i}{\partial x^j}, \qquad 1 \le i, j \le n.$$

This was a necessary condition for exactness (whether it is also sufficient depends entirely on the topology of the manifold). We also gave an invariant characterization of closedness, in Proposition 6.31: λ is closed iff

$$X(\lambda(Y)) - Y(\lambda(X)) = \lambda([X, Y]), \qquad \forall\, X, Y \in \mathscr{X}(M).$$

As we discussed in Remark 6.32, this can be phrased in the form $d\lambda = 0$ for an appropriate operator d. We will now formally define this. In fact, it is basically given by the above condition.

DEFINITION 8.17. *Let $\lambda \in \Omega^1(M)$. Define $d\lambda \in \Omega^2(M)$ by its action on vector fields:*

$$d\lambda(X, Y) = X(\lambda(Y)) - Y(\lambda(X)) - \lambda([X, Y]).$$

This certainly defines a smooth tensor field of rank 2: as X, Y, λ are all smooth, the resulting function above is smooth, and it is apparent that $d\lambda$ is $C^\infty(M)$-bilinear. We can also check easily that it is antisymmetric. Since exact forms are closed, we have $d(df) = 0$ for any smooth f.

Now, let's see what $d\lambda$ looks like in local coordinates. Fix a chart $(U, (x^j))$; then on U $\lambda = \sum_{j=1}^n \lambda_j\, dx^j$. To compute the 2-form $d\lambda$, it suffices to observe its action on any pair of basis vectors $(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^\ell})$ with $1 \le k < \ell \le n$. Notice that the bracket of these vector fields is 0, and so we can quickly compute

$$d\lambda\left(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^\ell}\right) = \frac{\partial}{\partial x^k}\left(\lambda\left(\frac{\partial}{\partial x^\ell}\right)\right) - \frac{\partial}{\partial x^\ell}\left(\lambda\left(\frac{\partial}{\partial x^k}\right)\right) = \frac{\partial \lambda_\ell}{\partial x^k} - \frac{\partial \lambda_k}{\partial x^\ell}.$$

Now, we know we can write this in the form $\sum_{i<j} \omega_{ij} dx^i \wedge dx^j$. To determine the coefficients, we compute that, in general,

$$\sum_{i<j} \omega_{ij}\, dx^i \wedge dx^j \left(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^\ell}\right) = \omega_{k\ell}.$$

Thus, equating coefficients, we have

$$d\left(\sum_{j=1}^n \lambda_j\, dx^j\right) = \sum_{i<j} \left(\frac{\partial \lambda_j}{\partial x^i} - \frac{\partial \lambda_i}{\partial x^j}\right) dx^i \wedge dx^j. \tag{8.1}$$

EXAMPLE 8.18. On (an open subset of) $\mathbb{R}^3$, if we identify a vector field $X = \sum_{j=1}^3 X^j \frac{\partial}{\partial x^j}$ with the 1-form $\flat(X) = \sum_{j=1}^3 X^j\, dx^j$, and then take the exterior derivative, that gives

$$d\flat(X) = \left(\frac{\partial X^2}{\partial x^1} - \frac{\partial X^1}{\partial x^2}\right) dx^1 \wedge dx^2 + \left(\frac{\partial X^3}{\partial x^1} - \frac{\partial X^1}{\partial x^3}\right) dx^1 \wedge dx^3 + \left(\frac{\partial X^3}{\partial x^2} - \frac{\partial X^2}{\partial x^3}\right) dx^2 \wedge dx^3.$$

Now, if we identify this 2-form on $\mathbb{R}^3$ with a vector field via the isomorphism $\beta = (\cdot)\lrcorner(dx^1 \wedge dx^2 \wedge dx^3)\colon \mathscr{X}(\mathbb{R}^3) \to \Omega^2(\mathbb{R}^3)$, this gives

$$\beta(d\flat(X)) = \left(\frac{\partial X^3}{\partial x^2} - \frac{\partial X^2}{\partial x^3}\right) \frac{\partial}{\partial x^1} - \left(\frac{\partial X^3}{\partial x^1} - \frac{\partial X^1}{\partial x^3}\right) \frac{\partial}{\partial x^2} + \left(\frac{\partial X^2}{\partial x^1} - \frac{\partial X^1}{\partial x^2}\right) \frac{\partial}{\partial x^3} = \nabla \times X.$$

That is to say, on $\mathbb{R}^3$, the curl operator $\nabla\times$ is actually just the exterior derivative $d\colon \Omega^1(\mathbb{R}^3) \to \Omega^2(\mathbb{R}^3)$ composed with the (coordinate-dependent) isomorphism $\flat$ and β identifying $\mathscr{X}(\mathbb{R}^3)$ alternately with $\Omega^1(\mathbb{R}^3)$ and $\Omega^2(\mathbb{R}^3)$.

Let us make a further observation about (8.1): it amounts to a simple rule for computing exterior derivatives (in local coordinates). Indeed,

$$d\lambda_j \wedge dx^j = \left(\sum_{i=1}^n \frac{\partial \lambda_j}{\partial x^i} dx^i\right) \wedge dx^j = \sum_{i \neq j} \frac{\partial \lambda_j}{\partial x^i}\, dx^i \wedge dx^j,$$

and so, summing up, we have

$$\begin{aligned}
\sum_{j=1}^n d\lambda_j \wedge dx^j &= \sum_{j=1}^n \sum_{i \neq j} \frac{\partial \lambda_j}{\partial x^i}\, dx^i \wedge dx^j \\
&= \sum_{j=1}^n \left(\sum_{i<j} + \sum_{i>j}\right) \frac{\partial \lambda_j}{\partial x^i}\, dx^i \wedge dx^j \\
&= \sum_{j=1}^n \sum_{i<j} \frac{\partial \lambda_j}{\partial x^i}\, dx^i \wedge dx^j - \sum_{j=1}^n \sum_{j<i} \frac{\partial \lambda_j}{\partial x^i}\, dx^j \wedge dx^i.
\end{aligned}$$

Reversing the variable names $i \leftrightarrow j$ in the second sum shows that

$$\sum_{j=1}^n d\lambda_j \wedge dx^j = \sum_{i<j} \left(\frac{\partial \lambda_j}{\partial x^i} - \frac{\partial \lambda_i}{\partial x^j}\right) dx^i \wedge dx^j = d\left(\sum_{j=1}^n \lambda_j\, dx^j\right).$$

That is: to compute d of a 1-form, simply take d of each of its coefficients and wedge with the proceeding dx terms. It is not a priori clear that this operation in local coordinates is invariant; the fact that it came from the invariant expression of Definition 8.17 means that it indeed is invariant.

We now wish to generalize this to k-forms with $k > 1$. Motivated by the above, we can define the exterior derivative on k-forms on $\mathbb{R}^n$ as follows.

DEFINITION 8.19. *Let $U \subseteq \mathbb{R}^n$ be open, and let $\omega \in \Omega^k(U)$. So we have $\omega = \sum_{I\uparrow} \omega_I\, dx^I$ for $\omega_I \in C^\infty(U)$. Define a $(k+1)$-form $d\omega \in \Omega^{k+1}(U)$ as follows:*

$$d\omega = \sum_{I\uparrow} d\omega_I \wedge dx^I.$$

It is immediate to verify that this defines a(n $\mathbb{R}$-)linear operator $d\colon \Omega^k(U) \to \Omega^{k+1}(U)$. Also, there is no need to express ω in the standard way in terms of increasing multi-indices. We have the following.

LEMMA 8.20. *If I is any length-k multi-index (not necessarily increasing, and possibly containing repeated indices) and f is smooth, then $d(f\,dx^I) = df \wedge dx^I$.*

PROOF. If I has repeated indices then $dx^I = 0$ and so both sides are 0. Otherwise, let σ be the unique permutation of I so that $\sigma \cdot I$ is increasing. Then we have $dx^I = \operatorname{sgn}(\sigma) dx^{\sigma \cdot I}$. So by Definition 8.19 and linearity,

$$d(f\,dx^I) = d(\operatorname{sgn}(\sigma) f\,dx^{\sigma\cdot I}) = \operatorname{sgn}(\sigma) df \wedge dx^{\sigma \cdot I} = df \wedge dx^I.$$

□

It now follows by simple calculation that $d^2 = 0$ in general.

LEMMA 8.21. *If $\omega \in \Omega^k(U)$ for some $U \subseteq \mathbb{R}^n$ open, then $d(d\omega) = 0$.*

PROOF. First, since d is linear, $d^2 \colon \Omega^k(U) \to \Omega^{k+2}(U)$ is linear, and so to show it is 0 it suffices to show $d^2(\omega_I \wedge dx^I) = 0$ for any fixed $I \uparrow$. From the definition, we have

$$d(\omega_I\,dx^I) = d\omega_I \wedge dx^I = \sum_{j=1}^n \frac{\partial \omega_I}{\partial x^j}\,dx^j \wedge dx^I.$$

Now, applying d again, using linearity and Lemma 8.20, since $dx^j \wedge dx^I = dx^J$ for some multi-index K of length $k+1$, we get

$$d(d\omega_I\,dx^I) = \sum_{1 \le i,j \le n} \frac{\partial^2 \omega_I}{\partial x^i \partial x^j}\,dx^i \wedge dx^j \wedge dx^I.$$

Note that the $i = j$ terms are 0 since $dx^i \wedge dx^i = 0$. Therefore, we can break this sum up into the terms with $i < j$ and the terms with $i > j$:

$$\begin{aligned} d\left(d(\omega_I\,dx^I)\right) &= \sum_{i<j} \frac{\partial^2 \omega_I}{\partial x^i \partial x^j}\,dx^i \wedge dx^j \wedge dx^I + \sum_{i>j} \frac{\partial^2 \omega_I}{\partial x^i \partial x^j} dx^i \wedge dx^j \wedge dx^I \\ &= \sum_{i<j} \frac{\partial^2 \omega_I}{\partial x^i \partial x^j}\,dx^i \wedge dx^j \wedge dx^I - \sum_{j<i} \frac{\partial^2 \omega_I}{\partial x^i \partial x^j} dx^j \wedge dx^i \wedge dx^I. \end{aligned}$$

Finally, exchanging the labels $i \leftrightarrow j$ in the second term, and using the fact that mixed partial derivatives of ω_I commute, shows that this is 0. □

EXAMPLE 8.22. Let $\omega = \beta(X) = X^1\,dx^2 \wedge dx^3 - X^2\,dx^1 \wedge dx^3 + X^3\,dx^1 \wedge dx^2$ be a 2-form on $\mathbb{R}^3$. Using Definition 8.19, we compute

$$d\omega = dX^1 \wedge dx^2 \wedge dx^3 - dX^2 \wedge dx^1 \wedge dx^3 + dX^3 \wedge dx^1 \wedge dx^2.$$

Now, $dX^j = \frac{\partial X^j}{\partial x^1}\,dx^1 + \frac{\partial X^j}{\partial x^2}\,dx^2 + \frac{\partial X^j}{\partial x^3}\,dx^3$. Plugging this into each of the three places and expanding, we note that only one of the three terms survives in each case: for example,

$$\left(\frac{\partial X^1}{\partial x^1}\,dx^1 + \frac{\partial X^1}{\partial x^2}\,dx^2 + \frac{\partial X^1}{\partial x^3}\,dx^3\right) \wedge dx^2 \wedge dx^3 = \frac{\partial X^1}{\partial x^1}\,dx^1 \wedge dx^2 \wedge dx^3$$

since each of the other two terms contains a wedge of two equal 1-forms and is 0. Repeating this with the other two, we arrive at

$$\begin{aligned} d\omega &= \frac{\partial X^1}{\partial x^1}\, dx^1 \wedge dx^2 \wedge dx^3 - \frac{\partial X^2}{\partial x^2}\, dx^2 \wedge dx^1 \wedge dx^3 + \frac{\partial X^3}{\partial x^3}\, dx^3 \wedge dx^1 \wedge dx^2 \\ &= \left(\frac{\partial X^1}{\partial x^1} + \frac{\partial X^2}{\partial x^2} + \frac{\partial X^3}{\partial x^3} \right) dx^1 \wedge dx^2 \wedge dx^3 = \nabla \cdot X\, dx^1 \wedge dx^2 \wedge dx^3. \end{aligned}$$

That is: identifying $\Omega^3(\mathbb{R}^3)$ with $C^\infty(\mathbb{R}^3)$ via the (coordinate-dependent) isomorphism $*(f) = f\, dx^1 \wedge dx^2 \wedge dx^3$, we have $* \circ d \circ \beta \colon \mathscr{X}(\mathbb{R}^3) \to \mathscr{X}(\mathbb{R}^3)$ is the divergence operator – divergence is just the exterior derivative $d \colon \Omega^2(\mathbb{R}^3) \to \Omega^3(\mathbb{R}^3)$ composed with the identifications between these spaces of differential forms with vector fields.

Combining this with the (motivating) calculation that $df = \flat(\nabla f)$, we then have the following big commutative diagram:

$$\begin{tikzcd} C^\infty(\mathbb{R}^3) \arrow[r, "\nabla"] \arrow[d, "\mathrm{Id}"] & \mathscr{X}(\mathbb{R}^3) \arrow[r, "\nabla\times"] \arrow[d, "\flat"] & \mathscr{X}(\mathbb{R}^3) \arrow[r, "\nabla\cdot"] \arrow[d, "\beta"] & C^\infty(\mathbb{R}^3) \arrow[d, "*"] \\ \Omega^0(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^1(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^2(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^3(\mathbb{R}^3) \end{tikzcd}$$

In particular, the classical vector calculus identities that $\nabla \times \nabla f = 0$ and $\nabla \cdot \nabla \times X = 0$ are just dressed up versions of the fact that $d^2 = 0$.

Now, we know that d on functions satisfies a product rule: from Proposition 6.12(b), $d(fg) = f\, dg + g\, df$. We expect the more general exterior derivative on higher degree forms should also satisfy a product rule. Since the algebra $\Omega^*(\mathbb{R}^n)$ is anti-commutative, it turns out d is an *anti*-derivation.

PROPOSITION 8.23. *Let $U \subseteq \mathbb{R}^3$, and let $\omega \in \Omega^k(U)$ and $\eta \in \Omega^\ell(U)$. Then*

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

PROOF. by linearity of d, it sufficies to consider terms of the form $\omega = f\, dx^I$ and $\eta = g\, dx^J$ for some multi-indices $|I| = k$ and $|J| = \ell$, where $f, g \in C^\infty(U)$. Then $\omega \wedge \eta = fg\, dx^I \wedge dx^J$. Then we have

$$\begin{aligned} d(\omega \wedge \eta) = d(fg) \wedge dx^I \wedge dx^J &= (f\, dg + g\, df) \wedge dx^I \wedge dx^J \\ &= df \wedge dx^I \wedge (g\, dx^J) + dg \wedge (f\, dx^I) \wedge dx^J \\ &= df \wedge dx^I \wedge \eta + dg \wedge \omega \wedge dx^J. \end{aligned}$$

In the first term, $df \wedge dx^I = d\omega$. For the second term, we need to commute df past dx^I. Since dg is a 1-form, we can expand it as a linear combination of dx^i terms, and each of these satisfies $dx^i \wedge dx^I = (-1)^k dx^I \wedge dx^i$ since $|I| = k$. Thus, by linearity, we recombine and see that $dg \wedge \omega \wedge dx^J = (-1)^k \omega \wedge dg \wedge dx^J = (-1)^\omega \wedge d\eta$, concluding the proof. □

We now know everything we need to about computing exterior derivatives of differential forms on patches of $\mathbb{R}^n$. We would like to parlay this into a definition of an exterior derivative $d \colon \Omega^k(M) \to \Omega^{k+1}(M)$ on any manifold. We have invariant definitions of this for $k = 0, 1$. In general, we'd like to say the following: for $\omega \in \Omega^k(M)$, define $d\omega$ locally: at any point $p \in M$, choose a chart at p, and let $d\omega_p$ be the local $(k+1)$-form given on U by Definition 8.19 in local coordinates on $\hat{U}$. This is a standard way to define d; it requires a tedious calculation to show that

it is well-defined (i.e. that you get the same expression in every coordinate chart). We will take a different approach here: one that emphasizes the computational properties of Lemmas 8.21 and 8.23 which, it turns out, together with the known action of d on 0-forms, uniquely defines its action in general.

PROPOSITION 8.24. *Let M be a smooth n-manifold. For $0 \leq k \leq n$, there are unique operators $d\colon \Omega^k(M) \to \Omega^{k+1}(M)$ (called* **exterior derivatives***), satisfying the following properties:*

(a) *d is $\mathbb{R}$-linear.*
(b) *If $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$, then*
$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$
(c) *$d^2 = 0$.*
(d) *For $f \in \Omega^0(M) = C^\infty(M)$, df is the usual differential $df(X) = X(f)$.*

Moreover, in any local coordinate chart, d is consistent with Definition 8.19.

PROOF. We begin by proving uniqueness. Suppose there is an operator d satisfying (a)–(d) above. First we will show it is locally defined: that is, suppose $\omega_1, \omega_2 \in \Omega^k(M)$ agree on some open set U. We need to see that $d\omega_1 = d\omega_2$ on U as well. To see this, fix $p \in U$ and set $\eta = \omega_1 - \omega_2$. Let $\psi \in C^\infty(M)$ be a bump function supported in U, and equal to 1 on some smaller neighborhood of p. Then $\psi\eta = 0$, and so by (a) and (b),
$$0 = d(0) = d(\psi\eta) = d\psi \wedge \eta + \psi\, d\eta.$$
Evaluating at p, and using $\psi(p) = 1$ and $d\psi_p = 0$, this shows that $d\eta_p = 0$. Thus $d\omega_1(p) = d\omega_2(p)$, and since this holds at every $p \in U$, this shows that d acts locally as claimed.

Now, let $\omega \in \Omega^k(M)$ and fix a coordinate chart $(U, (x^j))$ in M. Then we may write $\omega|_U = \sum_{I\uparrow} \omega_I\, dx^I$. For any $p \in U$, fix a smaller neighborhood of p, and select a bump function that is 1 on that neighborhood and supported in U; then $\psi\omega_I$ and ψx^i have smooth extensions to smooth functions $\widetilde{\omega}_I$ and $\widetilde{x}^i$ on M that are equal to ω_I and x^i on a neighborhood of p. By the locality argument above, this means that $dx^i = d\widetilde{x}^i$ on this neighborhood, and so therefore $dx^I = d\widetilde{x}^I$ on that neighborhood. Thus, setting $\widetilde{\omega} = \sum_{I\uparrow} \widetilde{\omega}_I\, d\widetilde{x}^I$ globally, we have $\omega|_U = \widetilde{\omega}|_U$. Now, from (a) and (b), together with locality, we then have $d\omega|_U = d\widetilde{\omega}|_U$, and
$$d\widetilde{\omega} = d\left(\sum_{I\uparrow} \widetilde{\omega}_I\, d\widetilde{x}^I\right) = \sum_{I\uparrow} d(\widetilde{\omega}_I\, d\widetilde{x}^I).$$
By (b), $d(\widetilde{\omega}_I\, d\widetilde{x}^I) = d\widetilde{\omega}_I \wedge d\widetilde{x}^I - \widetilde{\omega}_I \wedge d(d\widetilde{x}^I)$. Now for this last term, iterating (b) we get
$$d(d\widetilde{x}^I) = \sum_{j=1}^{k} d\widetilde{x}^{i_1} \wedge \cdots d(d\widetilde{x}^{i_j}) \wedge \cdots \wedge dx^{i_k} = 0$$
because, by (c), $d(d\widetilde{x}^{i_j}) = 0$ for any j. Thus, we actually must have
$$d\widetilde{\omega} = \sum_{I\uparrow} d\widetilde{\omega}_I \wedge d\widetilde{x}^I,$$
and on U this reduces exactly to the expression for $d\omega$ in Definition 8.19. (It is at this point that we use (d) to see that $d\omega_I$ has the usual meaning here.) This proves uniqueness (there is only one formula that $d\omega$ can have if it satisfies (a)–(d)), and that, should d exist, it must coincide with Definition 8.19 in local coordinates.

It remains to prove that d exists. One approach is to show that the local Definition 8.19 is invariant under coordinate change. A more appealing alternative is to give an invariant expression for an operator d which satisfies (a)–(d), generalizing Definition 8.17. That is our present approach. For $\omega \in \Omega^k(M)$, and $X_1, \dots, X_{k+1} \in \mathscr{X}(M)$, define $d\omega \in \Omega^{k+1}(M)$ as follows:

$$\begin{aligned} &d\omega(X_1, \dots, X_{k+1}) \\ &\quad = \sum_{i=1}^{k+1} (-1)^{i-1} X_i(\omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1})) \\ &\qquad + \sum_{1 \le i < j \le k+1} (-1)^{i+j} \omega([X_i, X_j], X_1, \dots, \widehat{X_i}, \dots, \widehat{X_j}, \dots, X_{k+1}). \end{aligned}$$

(As usual, the $\widehat{\ }$ denotes this term being omitted.) The proof now consists in showing that $d\omega$ is indeed a $(k+1)$-form (i.e. that it is an antisymmetric $C^\infty(M)$-multilinear map), and that it satisfies properties (a)–(d). Alternatively (and this is easier), one can show that, in any local coordinates, this definition yields the expression of Definition 8.19. Either way, the remainder of the proof is a long sequence of tedious but elementary calculations, that are left to the bored reader. □

This, the exterior derivative is a well-defined invariant operator, which can be calculated in local coordinates in the manner we did above. Let us now conclude our discussion here by showing that it is natural (i.e. commutes with pullbacks), and show further that the Lie derivative can be elegantly expressed in terms of the exterior derivative.

LEMMA 8.25. *Let $F\colon M \to N$ be a smooth map between manifolds. Then for each k, the pullback $F^*\colon \Omega^k(N) \to \Omega^k(M)$ commutes with the exterior derivative d: for all $\omega \in \Omega^k(N)$,*

$$F^*(d\omega) = d(F^*\omega).$$

PROOF. It suffices to show that the equality holds at any point $p \in M$. Let $(V, (y^j))$ be a chart at $F(p)$. Then we can expand ω locally as $\omega|_V = \sum_{I\uparrow} \omega_I \, dy^I$, and so $d\omega|_V = \sum_{I\uparrow} d\omega_I \wedge dy^I$. Now pulling back, we then have (by Lemma 8.11)

$$F^*(d\omega) = \sum_{I\uparrow} F^*(d\omega_I) \wedge F^*(dy^I) = \sum_{I\uparrow} d(\omega_I \circ F) \wedge dF^I$$

where $dF^I = dF^{i_1} \wedge \cdots \wedge dF^{i_k}$, given in terms of the components of F^i of F in the coordinates (y^i). On the other hand,

$$F^*(\omega) = \sum_{I\uparrow} F^*(\omega_I \, dy^I) = \sum_{I\uparrow} (\omega_I \circ F) \, dF^I$$

and now using Proposition 8.24(a,b) we have

$$dF^*(\omega) = \sum_{I\uparrow} \left(d(\omega_I \circ F) \wedge dF^I + \sum_{j=1}^{k} (\omega_I \circ F) dF^{i_1} \wedge \cdots \wedge d(dF^{i_j}) \wedge \cdots \wedge dF^{i_k} \right)$$

and all of the last terms in the internal sum are 0 since $d^2 = 0$ by Proposition 8.24(c). So the two expressions match locally near any point, and this proves the lemma. □

This brings us to Lie derivatives of differential forms. Referring back to Proposition 8.8, we can take the Lie derivative $\mathscr{L}_X \omega$ for any $\omega \in \Omega^k(M) \subseteq \mathscr{T}^k(M)$; the result is a new tensor field in $\mathscr{T}^k(M)$. What's more, from the definition (a limit difference quotient of the pullback along the flow of X), since pullbacks preserve $\Omega^k(M)$, it follows that $\mathscr{L}_X \omega \in \Omega^k(M)$ as well.

Proposition 8.8(b) assters that $\mathscr{L}_X(\omega \otimes \eta) = (\mathscr{L}_X\omega) \otimes \eta + \omega \otimes (\mathscr{L}_X\eta)$. By linearity of both sides, antisymmetrizing shows that, in fact,

$$\mathscr{L}_X(\omega \wedge \eta) = (\mathscr{L}_X\omega) \wedge \eta + \omega \wedge (\mathscr{L}_X\eta). \tag{8.2}$$

That is: $\mathscr{L}_X$ is a genuine $\wedge$-derivation, as opposed to d which is an anti-derivation. Note that $\mathscr{L}_X$ preserves the degree, while d increases it by 1. Despite these differences, there is a remarkable connection between the two kinds of derivatives that is a very useful computational tool. It is known as **Cartan's Magic Formula**.

THEOREM 8.26 (Cartan's Magic Formula). *Let M be a smooth manifold, $X \in \mathscr{X}(M)$, and $\omega \in \Omega^k(M)$ for some k. Then*

$$\mathscr{L}_X\omega = X \lrcorner d\omega + d(X \lrcorner \omega).$$

In other words, using the other notation for contraction $i_X = X \lrcorner$, we have $\mathscr{L}_X = i_X \circ d + d \circ i_X$.

Before we prove Theorem 8.26, we first prove as a lemma a fact that would follow very quickly from the theorem (together with $d^2 = 0$): d commutes with Lie derivatives.

LEMMA 8.27. *Let $X \in \mathscr{X}(M)$ and $\omega \in \Omega^k(M)$. Then $\mathscr{L}_X(d\omega) = d\mathscr{L}_X\omega$.*

PROOF. We do this in local coordinates $(U, (x^j))$. By linearity of both sides, it suffices to prove the equality for $\omega = f dx^I$ for some smooth function f and some increasing k-tuple $I = (i_1, \ldots, i_k)$. Using Proposition 8.8 together with (8.2) yields

$$\mathscr{L}_X\omega = X(f)\, dx^I + \sum_{j=1}^{k} f dx^{i_1} \wedge \cdots \wedge \mathscr{L}_X(dx^{i_j}) \wedge \cdots \wedge dx^{i_k}.$$

Now, in Corollary 6.21, we already proved that $\mathscr{L}_X$ and d commute on 1-forms, and so $\mathscr{L}_X(dx^{i_j}) = d\mathscr{L}_X(x^{i_j}) = d(X(x^{i_j})) = dX^{i_j}$. So this gives

$$\mathscr{L}_X\omega = X(f)\, dx^I + \sum_{j=1}^{k} f dx^{i_1} \wedge \cdots \wedge dX^{i_j} \wedge \cdots \wedge dx^{i_k}.$$

Calculating d of this, we note that $d(dx^{i_1} \wedge \cdots \wedge dX^{i_j} \wedge \cdots \wedge dx^{i_k}) = 0$ for the usual reasons, and so we simply have

$$d\mathscr{L}_X\omega = d(X(f)) \wedge dx^I + \sum_{j=1}^{k} df \wedge dx^{i_1} \wedge \cdots \wedge dX^{i_j} \wedge \cdots \wedge dx^{i_k}.$$

Reversing the above calculation then shows that this is equal to

$$d\mathscr{L}_X\omega = d(X(f)) \wedge dx^I + df \wedge \mathscr{L}_X(dx^I).$$

On the other hand, $d\omega = df \wedge dx^I$, and again using (8.2),

$$\mathscr{L}_X(d\omega) = \mathscr{L}_X(df) \wedge dx^I + df \wedge \mathscr{L}_X(dx^I).$$

Finally, using Corollary 6.21 again, $\mathscr{L}_X(df) = d(\mathscr{L}_X(f)) = d(X(f))$, and this concludes the computation showing that the two sides are equal as claimed. $\square$

PROOF OF THEOREM 8.26. We begin by verifying the formula in the case $k = 0$, i.e. $\omega = f \in C^\infty(M)$. Here we have $\mathscr{L}_X f = X(f)$. On the other hand $X \lrcorner df = df(X) = X(f)$ by definition, while $X \lrcorner f = 0$ by definition (since there are no slots to contract X into f; alternatively, we have $i_X \colon \Omega^k(M) \to \Omega^{k-1}(M)$, and so in the case $k = 0$ this gives output in $\Omega^{-1}(M) = 0$).

Now, for higher k, we note the following: suppose D, D' are two $\wedge$-derivations on $\Omega^*(M)$ that both commute with d, and satisfy $Df = D'f$ on $\Omega^0(M)$. Then they are equal. Indeed, working locally, by linearity it is enough to check equality on Ω^k on terms of the form $f\,dx^I$, and then the $\wedge$-derivation property yields

$$D(f\,dx^I) = Df\,dx^I + \sum_{j=1}^{k} dx^{i_1} \wedge \cdots \wedge D(dx^{i_j}) \wedge \cdots \wedge dx^{i_k}.$$

Now, since D commutes with d, we get $D(dx^{i_j}) = d(Dx^{i_j})$. By assumption, $Df = D'f$ and $Dx^{i_j} = D'x^{i_j}$. Now reversing the above calculation shows that $D(f\,dx^I) = D'(f\,dx^I)$, as claimed.

Now, let $D = i_X \circ d + d \circ i_X$. This is the anticommutator of two antiderivations (by Proposition 8.24(b) in the first term, and Lemma 7.22(b)). In fact, the anticommutator of two antiderivations is always a derivation: if $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$,

$$\begin{aligned} i_X d(\omega \wedge \eta) &= i_X(d\omega \wedge \eta + (-1)^k \omega \wedge d\eta) \\ &= (i_X d\omega) \wedge \eta + (-1)^{k+1} d\omega \wedge i_X\eta + (-1)^k i_X\omega \wedge d\eta + (-1)^{2k} \omega \wedge i_X d\eta, \\ d i_X(\omega \wedge \eta) &= d(i_X\omega \wedge \eta + (-1)^k \omega \wedge i_X\eta) \\ &= (d i_X\omega) \wedge \eta + (-1)^{k+1} i_X\omega \wedge d\eta + (-1)^k d\omega \wedge i_X\eta + (-1)^{2k} \omega \wedge d i_X\eta. \end{aligned}$$

Adding up, the two middle terms cancel out in each case, and since $(-1)^{2k} = 1$, this leaves

$$\begin{aligned} (i_X d + d i_X)(\omega \wedge \eta) &= (i_X d\omega) \wedge \eta + (d i_x \omega) \wedge \eta + \omega \wedge i_X d\eta + \omega \wedge d i_X \eta \\ &= ((i_X d + d i_X)\omega) \wedge \eta + \omega \wedge ((i_X d + d i_X)\eta). \end{aligned}$$

Thus, $D = i_X d + d i_X$ is a $\wedge$-derivation. It also commutes with d, since $d^2 = 0$:

$$\begin{aligned} D \circ d &= i_X \circ d^2 + d \circ i_X \circ d = d \circ i_X \circ d \\ d \circ D &= d \circ i_X \circ d + d^2 \circ i_X = d \circ i_X \circ d. \end{aligned}$$

By Lemma 8.27 and (8.2), $\mathscr{L}_X$ is also a $\wedge$-derivation that commutes with d. Since we verified that $D = \mathscr{L}_X$ on $\Omega^0(M)$, it now follows that the two are equal on $\Omega^*(M)$, concluding the proof. $\square$

CHAPTER 9

Submanifolds

Consider, for the moment, the nicest kind of map between Euclidean spaces: a linear map $T\colon \mathbb{R}^n \to \mathbb{R}^m$. Associated to T are two subspaces: $\mathrm{image}(T)$ and $\ker(T)$. To compute the dimensions of these subspaces, there is exactly one feature of T we need to know: its **rank**. By definition, $\mathrm{rank}(T) = \dim(\mathrm{image}(T))$. We also have, by the fundamental theorem of linear algebra, that $\dim(\ker(T)) = n - r$. So the rank identifies (up to isomorphism) the two subspaces.

Let $F\colon M \to N$ be a smooth map between manifolds. Our goal here is to understand the sets $F(M) \subseteq N$ and $F^{-1}\{q\}$ for each $q \in N$ (the comparable objects to the image and kernel of a linear map). Since F is smooth, it is supposed to be well modeled by its differential dF which, at each point p, is a linear map $dF_p\colon T_pM \to T_{F(p)}N$. However, this linear map can certainly change characteristics from point to point.

EXAMPLE 9.1. Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be the map $F(x,y) = x^2 + y^2$. Then $F(\mathbb{R}^2) = [0,\infty)$, which is not a manifold (in the subspace topology). Also, the level sets $F^{-1}(q)$ are empty for $q < 0$; for $q > 0$ they are circles (1-dimensional manifolds), and for $q = 0$ the level set is the single point $(0,0)$ (a 0-dimensional manifold). If we compute the total derivative, $DF(x,y) = 2[x,y]$; this is a linear map with rank 1 at every point other than $(x,y) = (0,0)$, but has rank 0 at the origin.

This example highlights that, if we want to have good consistent behavior for the image and level sets of F, we must insist that $\mathrm{rank}(dF_p)$ does not vary with p.

1. Maps of Constant Rank

Let $F\colon M \to N$ be a smooth map between manifolds. At each point $p \in M$, $dF_p\colon T_pM \to T_{F(p)}N$ is a linear map. Its **rank** is the dimension of the image $dF_p(T_p(M)) \subseteq T_{F(p)}N$. We say that F has **constant rank** if this integer r is constant over all $p \in M$; we then say $\mathrm{rank}(F) = r$. By elementary linear algebra, we have $0 \le r \le \min\{\dim M, \dim N\}$. If $\mathrm{rank}(dF_p)$ is equal to this upper bound, we say F is **full rank** at p; if this holds true at all points p, we say F is **full rank**.

In the case of a linear map, if the rank is equal to the dimension of the codomain, then the map is surjective; if the rank is equal to the dimension of the domain, then the map is injective. These two cases, for constant rank smooth maps, are the most important.

DEFINITION 9.2. *Let $F\colon M \to N$ be a smooth map with constant rank. If $\mathrm{rank}(F) = \dim M$, we call F an* **immersion***; equivalently, this means dF_p is one-to-one at each $p \in M$. If $\mathrm{Rank}(F) = \dim N$, we call F a* **submersion***; equivalently, dF_p is surjective at each $p \in M$.*

In fact, these two conditions make sense locally (i.e. immersion or submersion on a neighborhood), and in that context they are open conditions.

LEMMA 9.3. *Let $F\colon M \to N$ be a smooth map, and let $p \in M$. If dF_p is injective, then there is a neighborhood U of p such that $F|_U$ is an immersion. If dF_p is surjective, then there is a neighborhood U of p such that $F|_U$ is a submersion.*

PROOF. In a chart at p, the differential dF_p is represented by the Jacobian matrix $D\hat{F}(p)$. Let r be the rank of this map. If $D\hat{F}(p)$ is injective, this means $r = m = \dim M$, and so there are m linearly independent rows. By reordering coordinates, we may assume that the upper $m \times m$ submatrix is invertible. Invertibility is an open condition, and since the component functions of $D\hat{F}$ in the chart are continuous, it follows that this submatrix is invertible on a neighborhood U of p. Thus, $\operatorname{rank}(dF_{p'})$ is constantly equal to m for all $p' \in U$, showing F is an immersion there. The argument for local submersion is similar. □

EXAMPLE 9.4.
(a) If $M_1, \dots, M_k$ are smooth manifolds, then the projection maps $\pi_j \colon M_1 \times \cdots \times M_k \to M_j$ are submersions. Notably, the map $\pi(x_1, \dots, x_k, x_{k+1}, \dots, x_{k+\ell}) = (x_1, \dots, x_k)$ from $\mathbb{R}^{k+\ell}$ to $\mathbb{R}^k$ is a submersion.
(b) If $a < b$ in $\mathbb{R}$ and $\alpha \colon (a, b) \to M$ is a smooth curve, then α is an immersion iff $\dot{\alpha}(t) \neq 0$ for all $t \in (a, b)$.
(c) The bundle projection $\pi \colon TM \to M$ is always a submersion. After all, in the standard coordinates for TM (given any chart in M), π is given by item (a) above.

If a map is both a submersion and an immersion (at least locally), it need not be a diffeomorphism; but it is *locally*. This is basically the content of the inverse function theorem.

THEOREM 9.5 (Inverse Function Theorem for Manifolds). *Let $F \colon M \to N$ be a smooth map, and let $p \in M$. If dF_p is invertible, then there are connected neighborhoods U_0 of p and V_0 of $F(p)$ such that $F|_{U_0} \colon U_0 \to V_0$ is a diffeomorphism.*

PROOF. Since dF_p is invertible, T_pM and $T_{F(p)}N$ must have the same dimension, and so M and N have the same dimension. Choose charts (U, φ) at p and (V, ψ) at $F(p)$. In these coordinates, setting $\hat{F} = \psi \circ F \circ \varphi^{-1}$ and $\hat{p} = \varphi(p)$ as usual, we have $D\hat{F}(\hat{p}) = d\psi_{F(p)} \circ dF_p \circ d(\varphi^{-1})_{\hat{p}}$ is invertible. By the ordinary inverse function theorem, there are connected open neighborhoods $\hat{U}_0 \subseteq \varphi(U)$ and $\hat{V}_0 \subseteq \psi(V)$ such that $\hat{F}$ is a diffeomorphism from $\hat{U}_0$ onto $\hat{V}_0$. Taking $U_0 = \varphi^{-1}(\hat{U}_0)$ and $V_0 = \psi^{-1}(\hat{V}_0)$ completes the proof. □

A map that is both a submersion and an immersion (i.e. a map whose differential is invertiable at each point) is called a **local diffeomorphism**; the inverse function theorem explains the terminology. To be clear, the definitions are equivalent, as the following sort of converse to the inverse function theorem shows.

LEMMA 9.6. *Suppose $F \colon M \to N$ is a smooth map such that, at each $p \in M$, there is a neighborhood U of p where $F|_U$ is a diffeomorphism onto its image. Then F is both a submersion and an immersion.*

This is easily verified, and left to the reader.

Here is a list of basic properties of local diffeomorphisms, all easy to verify.

PROPOSITION 9.7. *Here are some basic properties of local diffeomorphisms.*
(a) *Compositions of local diffeomorphisms are local diffeomorphisms.*
(b) *A finite Cartesian product of local diffeomorphisms is a local diffeomorphism.*
(c) *Local diffeomorphisms are open maps.*
(d) *A restriction of a local diffeomorphism to an open set is a local diffeomorphism.*
(e) *Diffeomorphisms are local diffeomorphisims.*
(f) *Conversely, a bijective local diffeomorphism is a diffeomorphism.*

Some of the properties of submersions and immersions carry over to constant rank maps in general. The most important (local) theorem for such maps is the following.

THEOREM 9.8 (Rank Theorem). *Let M^m and N^n be smooth manifolds, and let $F\colon M \to N$ be a smooth map of constant rank r. For each $p \in M$, there are charts (U, φ) at p and (V, ψ) at $F(p)$ with $F(U) \subseteq V$, such that $\hat{F} = \psi \circ F \circ \varphi^{-1}$ has the coordinate representation*

$$\hat{F}(x^1, \dots, x^r, x^{r+1}, \dots, x^m) = (x^1, \dots, x^r, 0, \dots, 0).$$

As the theorem is local, the proof is really a statement about smooth maps between open submanifolds of Euclidean spaces. The proof is a matter of detailed analysis of the structure of the derivative matrix $D\hat{F}(\hat{p})$, and is actually very similar to the proof of the implicit function theorem, Theorem 0.20. The niggly details are left to the interested reader, and can be found as [**3**, Theorem 4.12]. Note, in particular, two important special cases:

- If F is an immersion, then $r = m \leq n$, and $\hat{F}(x^1, \dots, x^m) = (x^1, \dots, x^m, 0, \dots, 0)$.
- If F is a submersion, then $r = n \leq m$, and $\hat{F}(x^1, \dots, x^n, x^{n+1}, \dots, x^m) = (x^1, \dots, x^n)$.

The most important characteristic of these representations is that they are *linear* functions. This is an invariant characterization of what it means to be constant rank.

COROLLARY 9.9. *Let $F\colon M \to N$ be a smooth map, and let M be connected. TFAE:*

(a) *F has constant rank.*
(b) *For each $p \in M$, there are charts at p and $F(p)$ so that, in these local coordinates, $\hat{F}$ is linear.*

PROOF. The rank theorem shows that (a) $\Longrightarrow$ (b). Conversely, suppose (b) holds true. Since any linear map has constant rank, and the rank of the total derivative of the coordinate representation at a point is the same as the rank of the differential, it follows that F has constant rank throughout each chart. By connectedness, it follows that the rank is constant everywhere. $\square$

We conclude this section with the following global version of the rank theorem.

THEOREM 9.10 (Global Rank Theorem). *Let $F\colon M \to N$ be a smooth map of constant rank.*

(a) *If F is injective, then it is an immersion.*
(b) *If F is surjective, then it is a submersion.*
(c) *If F is bijective, then it is a diffeomorphism.*

PROOF. Let $r = \mathrm{rank} F$, and let $m = \dim M$ and $n = \dim N$. For (a), suppose F is not an immersion, meaning $r < m$. Fix some point $p \in M$. By the rank theorem, we can choose charts (U, φ) at p and (V, ψ) at $F(p)$ so that, in these coordinates, $\hat{F}(x^1, \dots, x^m) = (x^1, \dots, x^r, 0, \dots, 0)$. In particular, this means that $\hat{F}(0, \dots, 0, t, 0, \dots, 0) = 0 = \hat{F}(0, \dots, 0)$ for all sufficiently small $t \in \mathbb{R}$ (where there are r 0s before the t). This means that $\hat{F}$ is not injective, and so neither is F.

For (b), suppose that F is not a submersion, meaning that $r < n$. By the rank theorem, at each $p \in M$ we may choose charts (U, φ) at p and (V, ψ) at $F(p)$ with $F(U) \subseteq V$ such that, in these coordinates, $\hat{F}(x^1, \dots, F^m) = (x^1, \dots, x^r, 0, \dots, 0)$. By shrinking the coordinate neighborhood U if necessary, we may assume that $\overline{U}$ is compact and $F(\overline{U}) \subseteq V$; thus $F(\overline{U})$ is a compact subset of $\{y \in V\colon y^{r+1} = \cdots = y^n = 0\}$. This is a closed subset of N that contains no open subset; hence it is nowhere dense in N. Now, choose a countable cover M by such charts (U_i, φ_i) and (V_i, ψ_i); then $F(\overline{U_i})$ is closed and nowhere dense in N for each i, and hence $F(M) \subseteq \bigcup_i F(\overline{U_i})$ is

a countable union of nowhere dense sets. By the Baire category theorem, it follows that $F(M)$ has empty interior in N. In particular, $F(M) \neq N$, and so F is not surjective.

Finally, if F is bijective, then by (a) and (b), F is both a submersion and an immersion, so it is a local diffeomorphism. So F is a bijective local diffeomorphism, and by Proposition 9.7(f), it is therefore a diffeomorphism. □

2. Immersions and Embeddings

An injective immersion is a good candidate for a manifold-version of an injective linear map, and the framework for defining submanifolds. This notion is lacking in subtle ways, however, as the following two examples illustrate.

EXAMPLE 9.11. Let $\beta\colon (-\pi,\pi) \to \mathbb{R}^2$ be the curve

$$\beta(t) = (\sin 2t, \sin t).$$

This is a parametrization of a lemniscate (figure-eight); it is the solution set to $x^2 = 4y^2(1-y^2)$. It is injective: if $\sin 2t_1 = \sin 2t_2$, we have $2\sin t_1 \cos t_1 = 2\sin t_2 \cos t_2$; additionally having $\sin t_1 = \sin t_2$ gives either $\sin t_1 = \sin t_2 = 0$ (which only happens at $t_1 = t_2 = 0$ in this domain) or $\cos t_1 = \cos t_2$. Since $\cos$ is $1-1$ on $(-\pi, 0]$ and $[0,\pi)$, and is even, the second condition means $t_2 = \pm t_1$. If $t_2 = -t_1$, then $\sin t_1 = \sin t_2 = \sin(-t_1) = -\sin t_1$ and so again this gives $\sin t_1 = \sin t_2 = 0$, meaning $t_1 = t_2 = 0$. Thus, β is injective. It is also an immersion: $\beta'(t) = (2\cos 2t, \cos t)$. The second coordinate only vanishes at $t = \pm\frac{\pi}{2}$ in the domain, and at both those points the first coordinate is equal to $-2 \neq 0$; thus $\mathrm{rank}\beta = 1$ constantly, which is the dimension of the domain manifold; thus β is an immersion.

However, the image lemniscate $\beta(-\pi,\pi)$ is not a manifold at all: the origin $(0,0)$ has no neighborhood that is locally Euclidean. Moreover, if we give its image the subspace topology, β is not a homeomorphism: the image is compact, while the domain is not. That is: β is an injective immersion, but it is not a topological embedding.

EXAMPLE 9.12. Let $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{C}$ denote the torus, and define a curve $\alpha\colon \mathbb{R} \to \mathbb{T}^2$ by

$$\alpha(t) = (e^{2\pi i t}, e^{2\pi a t}),$$

where a is some irrational number. This map is injective: for any real number b, $e^{2\pi i b t_1} = e^{2\pi i b t_2}$ iff $b(t_1 - t_2)$ is an integer, and this cannot happen simultaneously with $b = 1$ and $b = a \notin \mathbb{Q}$ unless $t_1 - t_2 = 0$. It is also an immersion: neither component of $\alpha'(t)$ vanishes for any $t \in \mathbb{R}$.

But the image $\alpha(\mathbb{R})$ is even worse than Example 9.11: it is a (fun, well-known) fact that $\alpha(\mathbb{R})$ is *dense* in $\mathbb{T}^2$. (The proof involves continued fractions; for an outline, see [**3**, Example 4.20 & Lemma 4.21].) So α is about as far from a topological embedding as any inejctive map could be! Nevertheless, such immersions will play a role in the theory of Lie groups.

To avoid such complications wherever we can, we are most interested in injective immersions that are topological embeddings.

DEFINITION 9.13. *Let M, N be smooth manifolds. A map $F\colon M \to N$ is called a* **smooth embedding** *if F is a topological embedding (i.e. a homeomorphism onto its image in the subspace topology), and also an immersion.*

We will try to consistently use the full phrase *smooth embedding*, since *embedding* often refers to a topological embedding. Not, even if a topological embedding happens to be smooth, it need not be a smooth embedding.

EXAMPLE 9.14. Let $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ be the map $\gamma(t) = (t^3, 0)$. This is a topological embedding: it is a one-to-one map, whose image is $\mathbb{R} \times \{0\}$, and the map is a homeomorphism onto this image (with continuous inverse $\gamma^{-1}(s, 0) = s^{1/3}$). It is also a smooth map. But it is not an immersion, since $\gamma'(0) = 0$. (In particular, this means that γ is not a diffeomorphism onto its image.)

EXAMPLE 9.15. Here are some examples of smooth embeddings.

(a) If M is a smooth manifold and $U \subseteq M$ is open, then the inclusion $U \hookrightarrow M$ is a smooth embedding.
(b) If $M_1, \dots, M_k$ are smooth manifolds, $p_i \in M_i$ are specified points, then for each j the map $\iota_j\colon M_j \hookrightarrow M_1 \times \cdots \times M_k$ given by
$$\iota_j(p) = (p_1, \dots, p_{j-1}, p, p_{j+1}, \dots, p_k)$$
is a smooth embedding. In particular: taking $k = 2$, $M_1 = \mathbb{R}^n$ and $M_2 = \mathbb{R}^m$, and taking $p_2 = 0$, the usual embedding $x \mapsto (x, 0)$ of $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+m}$ is a smooth embedding.
(c) The map $T\colon \mathbb{T}^2 \to \mathbb{R}^3$ given by
$$T(e^{iu}, e^{iv}) = ((2 + \cos u)\cos v, (2 + \cos u)\sin v, \sin u)$$
is a smooth embedding (see Homework 1).

There are a few easy conditions that automatically guarantee that an injective immersion is, in fact, and embedding.

PROPOSITION 9.16. *Let M, N be smooth manifolds, and let $F\colon M \to N$ be an injective immersion. If any of the following hold, then F is a smooth embedding.*

(1) *F is an open map, or a closed map.*
(2) *F is a proper map (the preimage of a compact set is compact).*
(3) *M is compact.*
(4) $\dim M = \dim N$.

PROOF. First, note (c) $\Longrightarrow$ (b) $\Longrightarrow$ (a): if M is compact then the preimage of any closed set in N under (the continuous map) F is compact; if F is a proper map between locally compact Hausdorff spaces, it is closed (cf. [**3**, Theorem A.57]). Hence, it suffices to prove that (a) implies F is an embedding to show that (b) or (c) implies F is an embedding. But this is another general topological fact: a continuous injective open or closed map is always an embedding (cf. [**3**, Theorem A.38]). So it remains only to prove that (d) implies F is an embedding.

As F is an immersion, dF_p is injective at each p; but since the domain and codomain of dF_p have the same dimension, it follows that dF_p is a bijection, meaning that F is an injective local diffeomorphism. By Proposition 9.7(c), F is an open map. A continuous injective open map is an embedding, concluding the proof. □

EXAMPLE 9.17. Let $\iota\colon \mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1}$ be the inclusion map. As we computed in Example 2.11, this map is smooth. We can also readily compute (using the same charts in Example 2.11) that it is an immersion. Hence, ι is an injective immersion. Since $\mathbb{S}^n$ is compact, it follows from Propostion 9.16(3) that ι is a smooth embedding.

Examples 9.11 and 9.12 show two ways that injective immersions can fail to be smooth embeddings; however, as can be seen readily in both examples, in each case, the maps are *locally* smooth embeddings: every point in the domain has a neighborhood on which the map is a smooth embedding. (In Example 9.12, the neighborhood can be taken to be *any* finite-length interval containing the point.) This turns out to characterize the difference.

THEOREM 9.18. *Let $F\colon M \to N$ be a smooth map. Then F is an immersion iff each point in M has a neighborhood U such that $F|_U$ is a smooth embedding.*

PROOF. First, suppose very point p has such a neighborhood. Then dF_p is full-rank for each p, so F is an immersion. Conversely, suppose F is an immersion, and let $p \in M$. By the Rank Theorem, there is a coordinate chart $(U_1, (x^j)_{j=1}^m)$ where, in local coordinates, F has the form $\hat{F}(x^1, \dots, x^m) = (x^1, \dots, x^m, 0, \dots, 0)$. Hence, $F|_{U_1}$ is injective. Choose an open subset $U \subset U_1$ such that $\overline{U}$ is compact and $\overline{U} \subset U_1$. Then $F|_{\overline{U}}$ is an injective continuous map on a compact set, so it is a topological embedding (cf. Proposition 9.16(3)). The same remains true, therefore, on U, concluding the proof. □

3. Submanifolds

An **embedded submanifold** $S \subseteq M$ of a manifold M is a subset S that is a topological manifold in the subspace topology, and which possesses a smooth structure for which the inclusion map $S \hookrightarrow M$ is a smooth embedding. Embedded submanifolds are also sometimes called **regular submanifolds**. If S is an embedded submanifold of M, we call M the **ambient manifold**. The **codimension** of S is $\dim(M) - \dim(S)$.

So, by definition, an embedded submanifold is the image of a smooth embedding (namely its inclusion into its ambient manifold). The converse is also true.

PROPOSITION 9.19. *Let $F\colon N \to M$ be a smooth embedding of manifolds, and let $S = F(N)$. Then S is a topological manifold in the subspace topology, and there is a unique smooth structure on S with respect to which S is an embedded submanifold, and F is a diffeomorphism onto S.*

PROOF. Since F is a smooth embedding, it is a topological embedding, meaning that F is a homeomorphism of N onto S (embued with the subspace topology). Thus S is a topological manifold in the subspace topology. We define smooth charts on S to be of the form $(F(U), \varphi \circ F^{-1})$ for any smooth charts (U, φ) on N; the S-transition maps are then the same as the N-transition maps, and so these form an atlas. This smooth structure makes F into a diffeomorphism (it is given by the identity map in local coordinates), and it is easy to see it is the unique smooth structure with this property. Finally, the inclusion map $S \hookrightarrow M$ is given by

$$S \xrightarrow{F^{-1}} N \xrightarrow{F} M$$

so it is a composition of a smooth embedding and a diffeomorphism, and is thus a smooth embedding. Thus, S is an embedded submanifold. □

This, an embedded submanifold is precisely the image of some smooth embedding. Another term for a smooth embedding is a (global) **parametrization**: we parametrize a submanifold S by some other intrinsic manifold N. (Some authors reserve the term *parametrization* for when the parameter domain N is an open subset of some $\mathbb{R}^k$.)

The simplest embedded submanifolds are the ones we've been working with all along: open subsets. These turn out to be the only codimension 0 embedded submanifolds.

PROPOSITION 9.20. *A subset $S \subseteq M$ is a codimension 0 embedded submanifold iff S is open in M.*

PROOF. Suppose $U \subseteq M$ is open. In local coordinates, the inclusion map is given by the identity, and so it is an injective immersion; and it is immediate that it is a topological embedding. Thus, U is an embedded submanifold. Since it has the same dimension as M, its codimension is 0.

Conversely, suppose $S \subseteq M$ is a codimension 0 embedded submanifold. Then the inclusion map ι is an immersion, and since $\dim(S) = \dim(M)$, $d\iota_p$ is injective at each p, it follows that it is also surjective at each p. Thus, ι is a local diffeomorphism, and hence it is an open map by Proposition 9.7(c). Thus $\iota(S)$ is open in M. □

Some other standard examples of embedded submanifolds follow.

EXAMPLE 9.21. Let M, N be smooth manifolds. For any $q_0 \in N$, $S = M \times \{q_0\} \subseteq M \times N$ is an embedded submanifold. Indeed, S is the image of $F\colon M \to M \times N; p \mapsto (p, q_0)$, which is easily verified to be a smooth embedding.

EXAMPLE 9.22. Let M, N be smooth manifolds and $U \subseteq M$ open. Let $f\colon U \to N$ be a smooth map. Then the graph $\Gamma(f) \subset M \times N$

$$\Gamma(f) = \{(p, f(p)) \in M \times N \colon p \in U\}$$

is an embedded submanifold if dimension $\dim(M)$ (and so codimension $\dim(N)$). Indeed, $\Gamma(f) = \gamma_f(U)$ where $\gamma_f\colon U \to M \times N$ is the map $\gamma_f(p) = (p, f(p))$. This is an evidently smooth map. If $\pi_M\colon M \times N \to M$ is the usual projection, then $\pi_M \circ \gamma_f = \mathrm{Id}_U$, and so by the chain rule

$$d(\pi_M)_{(p,f(p))} \circ d(\gamma_f)_p = \mathrm{Id}_{T_pU}.$$

Since this composition is one-to-one, it follows that $d(\gamma_f)_p$ is one-to-one, and so γ_f is an immersion. It is also a topological embedding: its inverse is the continuous map $\pi_M|_{\Gamma(f)}$. Thus, $\Gamma(f)$ is an embedded submanifold diffeomorphic to U.

In particular, any affine subspace of $\mathbb{R}^n$ is an embedded submanifold (as any such space is, by definition, the graph of an affine function). The simpliest examples are the usual embeddings $\mathbb{R}^k \hookrightarrow \mathbb{R}^{k+n}$ given by $(x^1, \dots, x^k) \mapsto (x^1, \dots, x^k, 0, \dots, 0)$ (giving a special case of Example 9.21. In fact, this is a local model for *any* codimension n embedded submanifold.

Let M be a smooth n-manifold, and let $(U, \varphi = (x^j)_{j=1}^n)$ be a local chart in M. A k**-slice** of U is a subset $S \subseteq U$ is a subset with the property that

$$\varphi(S) = \{(x^1, \dots, x^k, x^{k+1}, \dots, x^n) \in \varphi(U) \colon x^{k+1} = c^{k+1}, \dots, x^n = c^n\}$$

for some constants $c^{k+1}, \dots, c^n \in \mathbb{R}$. More generally, a subset $S \subseteq M$ is said to satisfy the **local** k**-slice condition** if, for each $p \in S$, there is a chart (U, φ) at p so that $U \cap S$ is a k-slice.

THEOREM 9.23 (Slice Theorem). *Let M be a smooth manifold, and let $k \leq \dim(M)$. If $S \subseteq M$ is a k-dimensional embedded submanifold, then it satisfies the local k-slice condition. Conversely, if $S \subseteq M$ is a subset that satisfies the local k-slice condition, then it is a k-dimensional topological manifold in the subspace topology, and possesses a unique smooth structure making it an embedded submanifold.*

PROOF. If U is an embedded submanifold, then the inclusion map $\iota\colon U \hookrightarrow M$ is an immersion, and so by the rank theorem, for every $p \in S$ there are charts (U, φ) at p and (V, ψ) at $\iota(p)$ so that $\hat{\iota}(x^1, \dots, x^k) = (x^1, \dots, x^k, 0, \dots, 0)$ in these local coordinates. Shrinking the neighborhoods as necessary, we can arrange for $V \cap S = U$, which shows that S satisfies the local k-slice condition.

For the converse, we can build an atlas for S by using the local slice charts (presumed to exist) composed with the projections from $\mathbb{R}^n \to \mathbb{R}^k$ where $n = \dim(M)$. The details are left to the reader (and can be read in the proof of [**3**, Theorem 5.8]). □

By far the most useful kinds of embedded submanifolds are those that are identified as *level sets*. If $F\colon M \to N$ is any function, and $q_0 \in N$, the q_0-level set of F is the set $F^{-1}(q_0) \subseteq M$. In general, such sets can be arbitrarily nasty: using partitions of unity, you showed on an early homework set that *any* closed subset of M is a level set of some smooth function $f\colon M \to \mathbb{R}$. So we need some signficant restrictions on F to ensure that level sets are, in fact, embedded submanifolds. One simple condition is having constant rank.

PROPOSITION 9.24. *Let $F\colon M \to N$ be a smooth map of constant rank r. Then for any point $q_0 \in F(M)$, the level set $F^{-1}(q_0)$ is an embedded submanifold of codimension r.*

PROOF. Let $S = F^{-1}(q_0)$, and fix any $p \in S$. By the rank theorem, there are charts (U, φ) at p and (V, ψ) at $F(p) = q_0$ so that $\hat{F}(x^1, \dots, x^r, x^{r+1}, \dots, x^m) = (x^1, \dots, x^r, 0, \dots, 0)$ (where $m = \dim(M)$). So in these coordinates $\widehat{q_0} = (c^1, \dots, c^r, 0, \dots, 0)$ for some constants $c^1, \dots, c^r$, and hence, $S \cap U$ is the slice

$$S \cap U = \{(x^1, \dots, x^m) \in U\colon x^1 = c^1, \dots, x^r = c^r\}.$$

Thus, S satisfies the local $(m-r)$-slice condition, and thus by Theorem 9.23, S is an embedded submanifold of dimension $m - r$, thus codimension $m - (m - r) = r$. □

COROLLARY 9.25. *Let $F\colon M \to N$ be a submersion. Then each level set of F is an embedded submanifold whose codimension in M is equal to the dimension of N.*

In the special case of full rank, it is not actually necessary for the function F to be a submersion everywhere. It is enough that it is a submersion on a neighborhood of the "level". Such levels are called *regular values*.

DEFINITION 9.26. *Let $F\colon M \to N$ be a smooth map. A point $p \in M$ is called a* **regular point** *of F if dF_p is surjective; otherwise p is a* **critical point**. *(So, for example, if $\dim(M) < \dim(N)$, then all points in M are critical for F.) A point $q \in F(N)$ is called a* **regular value** *of F if every point $p \in F^{-1}(q)$ is a regular point; it is called a* **critical value** *otherwise. If q is a regular value, then the level set $F^{-1}(q)$ is called a* **regular level set**.

COROLLARY 9.27. *Every regular level set of a smooth map $F\colon M \to N$ is an embedded submanifold with codimension equal to $\dim(N)$.*

PROOF. A regular level set has the form $F^{-1}(q)$ where $q \in F(M)$ is a regular value of F. By Lemma 9.3, the set U of points $p \in M$ where dF_p is surjective is open, and by assumption $F^{-1}(q) \subseteq U$. Thus $F|_U\colon U \to N$ is a submersion, and the result follows by Corollary 9.25. □

EXAMPLE 9.28. We have already seen directly that the sphere $\mathbb{S}^n$ is a manifold. In fact, it is a codimension-1 embedded submanifold of $\mathbb{R}^{n+1}$: by definition, $\mathbb{S}^n = f^{-1}(1)$ where $f\colon \mathbb{R}^{n+1} \to \mathbb{R}$ is the map $f(x) = |x|^2$. The differential of this map is $df_x(v) = 2\langle x, v\rangle$, which is a surjective map everywhere other than $x = 0$. Since $x = 0$ is not in $f^{-1}(1)$, it follows that 1 is a regular value of f, so $\mathbb{S}^n = f^{-1}(1)$ is a regular level set, and thus is an embedded submanifold whose codimention is equal to the dimension of the codomain $\mathbb{R}$.

Not every embedded submanifold is a regular level set (just as not every embedded submanifold is the graph of a smooth function), but this is true *locally* (as it is for graphs, à la the local slice theorem).

PROPOSITION 9.29. *Let $S \subseteq M$ be a subset of an m-dimensional manifold. Then S is an embedded k-dimensional submanifold of M if and only if every point of S has a neighborhood U in M such that $U \cap S$ is a level set of a submersion $U \to \mathbb{R}^{m-k}$.*

PROOF. Suppose S is an embedded k-dimensional submanifold. Fix $p \in S$ and let $(U, \varphi = (x^j))$ be a chart at p in M that satisfies the local k-slice condition. Then we can define $F\colon U \to \mathbb{R}^{m-k}$ by $F(x^1, \dots, x^k, x^{k+1}, \dots, x^m) = (x^{k+1}, \dots, x^m)$. This is evidently a submersion, and by definition of the k-slice condition, $U \cap S$ is a level set of F.

Conversely, suppose every point has a neighborhood U such that there is a submersion $F\colon U \to \mathbb{R}^{m-k}$ with $S \cap U$ being a level set of F. Since F is a subjersion, it follows from Corollary 9.25 that $S \cap U$ is an embedded submanifold with codimension equal to the dimension of the codomain; thus the dimension is $m - (m-k) = k$. It then follows from the slice theorem (Theorem 9.23) that $S \cap U$ satisfies the local k-slice condition. This is true at each point, so S satisfies the local k-slice condition, and so again by the slice theorem, S is an embedded k-dimensional submanifold. □

4. Embeddings into Euclidean Space

It turns out that every smooth n-dimensional manifold has a smooth embedding into $\mathbb{R}^{2n+1}$. We will not prove this completely here (as this fact is only for general interest and not important for us at all). But we will prove some special cases. The first result is that all compact manifolds embed in some Euclidean space.

THEOREM 9.30. *Let M be a compact smooth manifold. Then there is a smooth embedding $F\colon M \hookrightarrow \mathbb{R}^k$ for some $k \in \mathbb{N}$.*

PROOF. For each $p \in M^m$, choose a chart (U_p, φ_p) at p so that $\phi_p(U_p)$ is an open ball in $\mathbb{R}^m$. For each p, fix an open subset $U'_p \subset U_p$ containing p such that $\overline{U'_p}$ is compact. The $\{U'_p\}_{p \in M}$ from and open cover of M, and since M is compact, there is a finite subcover $\{U'_{p_1}, \dots, U'_{p_n}\}$, each of which has compact closure, and since $\varphi_{p_j}\colon U_{p_j} \to \mathbb{R}^m$ is a diffeomorphism onto its image, $\varphi_{p_j}(\overline{U'_{p_j}})$ is compact in $\mathbb{R}^m$. Fix bump functions $\rho_1, \dots, \rho_n\colon M \to \mathbb{R}$ so that $\rho_j|_{\overline{U'_j}} = 1$ and $\operatorname{supp}(\rho_j) \subseteq U_j$. Then the function $U_{p_j} \ni p \mapsto \rho_j(p)\varphi_{p_j}(p) \in \mathbb{R}^m$ has a smooth extension f_j to M by setting it equal to 0 outside U_j, as usual. Then define $F\colon M \to \mathbb{R}^{mn+n}$ by

$$F(p) = (f_1(p), \dots, f_n(p), \rho_1(p), \dots, \rho_n(p)).$$

We will show that F is an injective immersion; since M is compact, it then follows by Proposition 9.16(3) that F is a smooth embedding, proving the theorem (with $k = mn + n$).

First, we show that F is injective. Suppose $p, q \in M$ with $F(p) = F(q)$. There is some p_j with $p \in U'_{p_j}$, and so $\rho_j(p) = 1$. Since $F(p) = F(q)$ and the last n coordinates of F are $\rho_1, \dots, \rho_n$, it follows that $\rho_j(q) = 1$ as well, which shows that $q \in \operatorname{supp}(\rho_j) \subseteq U_{p_j}$. Therefore p, q are both in the domain of φ_{p_j}, and moreover

$$\varphi_{p_j}(p) = \rho_j(p)\varphi_{p_j}(p) = f_j(p) = f_j(q) = \rho_j(q)\varphi_{p_j}(q) = \varphi_{p_j}(q)$$

where we used the fact that the first coordinates of F are given by $f_1, \dots, f_n$ and so $f_j(p) = f_j(q)$ for all j by the assumption that $F(p) = F(q)$. Since φ_{p_j} is a diffeomorphism from U_{p_j} onto its image, it is one to one, and so it now follows that $p = q$. Thus, F is injective.

Now, let $p \in M$ and again fix j so that $p \in U'_{p_j}$. Since $\rho_j \equiv 1$ on U'_{p_j}, we have $d(f_j)_p = d(\rho_j\varphi_j)_p = \varphi_j(p)d(\rho_j)_p + \rho_j(p)d(\varphi_j)_p = d(\varphi_j)_p$, which is injective since φ_j is a local diffeomorphism. Thus, dF_p is injective (if any component of a linear map is injective, the whole map is injective). So F is an injective immersion. This concludes the proof. □

This theorem can be dramatically improved in two ways: first, the result holds true for non-compact manifolds as well. To prove this requires us to deal with "regular" subsets that are not manifolds, but rather manifolds *with boundary*. We have avoided such discussion so far, and so we will not prove the result presently. The general idea is to break up M into a countable collection of compact pieces, each of which is a manifold with boundary, and apply the above embedding in a clever fashion (summing over bump functions). The hard part is showing that any manifold can be appropriately decomposed into such "regular domains". The key tool there is *Sard's theorem*, which says (in a strong sense) that regular values of smooth functions are generic (the set of singular values has measure 0). This is a local theorem, and really has nothing to do with differential geometry: it is a theorem about smooth maps between open subsets of Euclidean space.

The other improvement, sometimes called "Whitney's trick", is to reduce the potentially enormous k in Theorem 9.30: it turns out it can always be chosen no bigger than $2n+1$. This is a very geometric result, relying on a clever "folding" trick (showing that any proper embedding and be folder to fit inside an arbitrarily thin tubular neighborhood of a one-dimensional subspace of $\mathbb{R}^k$, and then using some affine geometry to show how this allows an embedding into $\mathbb{R}^{k-1}$ whenever $k > 2n$). As this result does not pertain to us in any way, we will not discuss it further here. It is worth noting that $2n+1$ is still not generally the least possible dimension; in fact, using more sophisticated tools from algebraic topology, Whitney later showed every n-manifold embeds in $\mathbb{R}^{2n}$. This is optimal for $n = 1, 2$ in general, but for $n = 3$, the optimal general imbedding dimension is 5. The general optimum is unknown for all manifolds, but for compact n-manifolds it is $2n + 1 - a(n)$, where $a(n)$ is the number of 1s in the binary expansion of n (and this is sharp).

5. Restrictions and Extensions

We begin this section by briefly discussion restrictions and extensions of smooth maps.

PROPOSITION 9.31. *Let $F\colon M \to N$ be a smooth map between manifolds, and let $S \subseteq M$ be an embedded submanifold. Then $F|_S$ is smooth.*

PROOF. By definition, the inclusion map $\iota\colon S \hookrightarrow M$ is an immersion, hence is smooth. Thus $F|_S = F \circ \iota$ is smooth. □

Note: this result does not even require and embedding: it holds true whenever the inclusion is an injective smooth map. For example, an *immersed submanifold* (the same as an embedded submanifold but without the requirement that the inclusion is a topological embedding) has this restriction property.

Complementaty to restricting a map, the natural question is whether a smooth map on a submanifold $S \subseteq M$ has a smooth extension to M. In general, the answer is no, as we saw in Remark 2.30: the identity map $\mathbb{S}^1 \to \mathbb{S}^1$, which is certainly smooth, has no smooth (or even continuous) extension to $\mathbb{R}^2 \to \mathbb{S}^1$. On the other hand, in the preceding Proposition 2.29, we saw that it is always possible to extend a smooth map on a closed subset to a neighborhood; but that proposition used a stronger sense of smooth. Recall that, if $A \subseteq M$ is closed, we called a map $F\colon A \to N$ smooth if each point of A has a neighborhood U in M such that F is the restriction to A of a smooth map $U \to N$. If A is a submanifold and $F\colon A \to N$ is smooth in this sense, then it is smooth in the intrinsic sense of smoothness between the manifolds A and N (via Proposition 9.31), but that is a priori stronger than what we ask for here. As with Proposition 9.31, in the case of a Euclidean codomain, the two notions are equivalent.

PROPOSITION 9.32. *Let $S \subseteq M$ be an embedded submanifold, and let $f\colon S \to \mathbb{R}^k$ be a smooth function. Then there is some open neighborhood of S in M and a smooth function $\widetilde{f}\colon U \to \mathbb{R}^k$ so that $f = \widetilde{f}|_S$. Moreover, if S is closed in M, then U may be taken to be all of M.*

The proof is very similar to the proof of Proposition 2.29 (utilizing a partition of unity), and is left as an exercise (on the homework).

Now we consider tangent spaces. Let $S \subseteq M$ be an embedded submanifold. Then S is a smooth manifold in its own right, and so for any $p \in S$ we have the tangent space T_pS (typically identified as Der_pS). But p is also in M, and so we have a tangent space T_pM as well. We would like to think of T_pS as a subspace of T_pM, but it is not. As we did in the case that S is an open submanifold (Proposition 3.11), we can identify T_pS with a natural isomorphic copy inside T_pM using the differential: since $\iota\colon S \hookrightarrow M$ is an immersion, $d\iota_p\colon T_pS \to T_pM$ is an injective map. Thus, we typically identifiy T_pS with $d\iota_p(T_pS) \subseteq T_pM$. In terms of derivations, given $X_p \in T_pS$, we identify it with $\widetilde{X}_p = d\iota_p(X_p) \in T_pM$ whose action on $f \in C^\infty(M)$ is

$$\tilde{X}_p(f) = d\iota_p(X_p)(f) = X_p(f \circ \iota) = X_p(f|_S).$$

Note that this identification only requires local structure, so it makes sense even if S is only immersed. In the embedded case, there is a very useful way to identify this subspace $T_pS \cong d\iota_p(T_pS)$ inside T_pM.

PROPOSITION 9.33. *Let $S \subseteq M$ be an embedded submanifold, and let $p \in S$. Then, as a subspace of T_pM,*

$$T_pS = \{X_p \in T_pM\colon X_p(f) = 0 \textit{ whenever } f \in C^\infty(M) \textit{ and } f|_S = 0\}.$$

PROOF. First, let $X_p \in T_pS \subseteq T_pM$. To be more precise, $X_p = d\iota_p(v)$ for some $v \in T_pS$, where $\iota\colon S \hookrightarrow M$ is the inclusion. Now let $f \in C^\infty(M)$ vanishes on S, then $f \circ \iota = 0$, and so $X_p(f) = d\iota_p(v)(f) = v(f \circ \iota) = v(0) = 0$. This shows the forward containment above.

Conversely, suppose $X_p \in T_pM$ satisfies $X_p(f) = 0$ whenever $f|_S = 0$. Fix a slice chart $(U, (x^j))$ at p, so that $U \cap S$ is the subset $x^{k+1} = \cdots = x^n = 0$ (where $k = \dim(S)$ and $n = \dim(M)$). Then $(x^1, \ldots, x^k)$ are coordinate for $S \cap U$, and the inclusion map $\iota\colon S \cap U \hookrightarrow M$ has the coordinate representation $\iota(x^1, \ldots, x^k) = (x^1, \ldots, x^k, 0, \ldots, 0)$. Thus T_pS is the subspace of T_pM spanned by $\frac{\partial}{\partial x^1}\big|_p, \ldots, \frac{\partial}{\partial x^k}\big|_p$. Given the coordinate representation of X_p,

$$X_p = \sum_{j=1}^n X_p^j \left.\frac{\partial}{\partial x^j}\right|_p$$

we see that $X_p \in T_pS$ if and only if $X_p^j = 0$ for $j = k+1, \ldots, n$.

Fix a bump function ψ supported in U that is 1 on a neighborhood of p. For any $j > k$, let $f_j(x) = \psi(x)x^j$, which extends to a smooth function on M by setting it equal to 0 outside U. Then, since $x^j = 0$ in $S \cap U$, $f_j \equiv 0$ on S, and so by our assumption, $X_p(f_j) = 0$. Thus

$$0 = X_p(f_j) = \sum_{i=1}^n X_p^i \frac{\partial f_j}{\partial x^j}(p) = X_p^j$$

since $\psi = 1$ on a neighborhood of p and $\frac{\partial x^j}{\partial x^i} = \delta_i^j$. So this shows that $X_p^j = 0$ for $j > k$, meaning that $X_p \in T_pS$, as desired. $\square$

Similarly, if the embedded submanifold is given (locally) as a regular level set of some function F, then the tangent space is the kernel of dF.

PROPOSITION 9.34. *Let $S \subseteq M$ be an embedded submanifold, and suppose $U \subseteq M$ is an open set so that $S \cap U = F^{-1}(q)$ for some smooth map $F\colon U \to N$ where $q \in N$ is a regular value of F. Then $T_pS = \ker(dF_p)$ for each $p \in S \cap U$.*

PROOF. Let $\iota\colon S \hookrightarrow M$ be the inclusion map as usual. Since $F \circ \iota$ is the constant map (with constant value q) on $S \cap U$, we have $0 = d(F \circ \iota)_p = dF_p \circ d\iota_p$ for any $p \in S \cap U$. Thus, if $X_p \in T_pS \subseteq T_pM$ (by which we mean $X_p \in T_pM$ has the form $X_p = d\iota_p(v)$ for some $v \in T_pS$), then $dF_p(X_p) = d(F \circ \iota)_p(v) = 0$, and so $X_p \in \ker(dF_p)$ as claimed. On the other hand, since q is a regular value, $dF_p\colon T_pM \to T_{F(p)}N$ is surjective, and so the fundamental theorem of linear algebra yields

$$\dim \ker(dF_p) = \dim(T_pM) - \dim(T_{F(p)}N) = \dim(T_pS) = \dim(\mathrm{image}(d\iota_p)).$$

Since the above argument shows $\mathrm{image}(d\iota_p) \subseteq \ker(dF_p)$, it thus follows that $\mathrm{image}(d\iota_p) = \ker(dF_p)$, which is precisely the statement of the proposition. □

6. Vector Fields and Tensor Fields

If $S \subseteq M$ is an embedded submanifold and $X \in \mathscr{X}(M)$, then we say X is **tangent to** S if $X_p \in T_pS \subseteq T_pM$ for each $p \in S$. Proposition 9.33 immediately yields the following characterization of fields on M tangent to S.

PROPOSITION 9.35. *Let $S \subseteq M$ be an embedded submanifold and let $X \in \mathscr{X}(M)$. Then X is tangent to S if and only if $(Xf)|_S = 0$ for every $f \in C^\infty(M)$ such that $f|_S = 0$.*

Now, suppose $S \subseteq M$ is an embedded (or even immersed) submanifold, and let $Y \in \mathscr{X}(M)$. Let $\iota\colon S \hookrightarrow M$ be the inclusion. If there is a vector field $X \in \mathscr{X}(S)$ that is (X, Y) are ι-related, then Y is tengent to S: ι-relation means that for each $p \in S$ $Y_p = d\iota_p(X_p)$, meaning that Y_p is in the image of $d\iota_p$ which is precisely T_pS. In fact, the converse of this is true in a strong sense as well.

PROPOSITION 9.36. *Let $S \subseteq M$ be an embedded submanifold and let $\iota\colon S \hookrightarrow M$ be the inclusion map. If $Y \in \mathscr{X}(M)$ is tangent to S, then there is a* unique *vector field $X \in \mathscr{X}(S)$ that is ι-related to Y.*

We refer to this unique vector field X as $X = Y|_S$.

PROOF. By assumption of tangency to S, $Y_p \in \mathrm{image}(d\iota_p)$ for each p, so there is some vector $X_p \in T_pS$ such that $Y_p = d\iota_p(X_p)$. Since ι is an immersion, $d\iota_p$ is injective, and so X_p is unique. Thus, we have a unique (rough) vector field X on S with $Y_p = d\iota_p(X_p)$ for each p; it thus suffices only to show that X is smooth.

Fix $p \in S$. Let $(U, (x^j))$ be a slice chart for S (in M) at p: so $S \cap U$ is the subset $x^{k+1} = \cdots = x^n = 0$ (where $k = \dim(S)$ and $n = \dim(M)$), and $(x^1, \ldots, x^k)$ are coordinates for $S \cap U$. If $Y = Y^1 \frac{\partial}{\partial x^1} + \cdots + Y^n \frac{\partial}{\partial x^n}$ in these coordinates, then (exactly as in the proof of Proposition 9.33) $X = Y^1 \frac{\partial}{\partial x^1} + \cdots + Y^k \frac{\partial}{\partial x^k}$. This is smooth on $S \cap U$. So every $p \in S$ has a neighborhood in S on which X is smooth, and so X is smooth. □

Finally, note that if $A \in \mathscr{T}^k(M)$ is a tensor field on M and $S \subseteq M$ is an embedded submanifold, then we can "restrict" A to S simply by pulling back along the inclusion map $\iota\colon S \hookrightarrow M$: $A|_S \equiv \iota^* A$. If we are thinking of T_pS already as a subset of T_pM for each p (identified using $d\iota_p$, then this amounts to the simple-minded restriction of A_p to the subset $(T_pS)^k$ of $(T_pM)^k$. Thus, a k-form on M restricts to a k-form on S. Note: if $\dim(S) < k \leq \dim(M)$ then a non-zero k-form on M restricts to the 0 k-form on S. In particular, the restriction map (which is linear) is *not* injective.

CHAPTER 10

Integration of Differential Forms

1. Orientation

Let V be a finite dimensional vector space. We can think of an orientation of V as being encoded by an *ordered basis*.

EXAMPLE 10.1. In $\mathbb{R}^3$, (e_1, e_2, e_3) gives the "standard" orientation, also called "right-handed": if you curl the fingers of your right hand around first in the e_1 direction then e_3, your thumb points in the e_3 direction (as opposed to $-e_3$). On the other hand (literally), (e_2, e_1, e_3) is different: this time you need to use your left hand to curl from e_2 to e_1 for your thumb to point along e_3.

The ordered basis (e_2, e_3, e_1) has the same property as (e_1, e_2, e_3): it is right-handed.

We really want to identify the orientations (e_1, e_2, e_3) and (e_2, e_3, e_1): what we want to keep track of is "handedness". After all, the transformation $(x^1, x^2, x^3) \mapsto (x^2, x^3, x_1)$ is a rotation, and we want orientation to be invariant under rotations.

DEFINITION 10.2. *Two ordered bases $(b_1, \ldots, b_n)$ and $(b'_1, \ldots, b'_n)$ for a finite-dimensional vector space V are called* **consistently oriented** *if the change of basis matrix B*

$$b'_i = \sum_{j=1}^{n} B_i^j b_j$$

has positive determinant. This is an equivalence relation (as is easy to check). An **orientation** *for V is an equivalence class of consistently oriented ordered bases.*

Note: every change of basis matrix is nonsingular, and so has either positive or negative determinant. Thus, the set of orientations has size 2. In the above Example, (e_1, e_2, e_3) and (e_2, e_3, e_1) are consistently oriented so yield the same orientation, while (e_2, e_1, e_3) give the opposite orientation. (In the case that the two bases are related by a permutation of their order, the change of basis matrix is the corresponding permutation matrix, whose determinant is the sgn of the permutation).

A vector space together with a selected orientation is called an **oriented vector space**. A basis for an oriented vector space is called **positively oriented** if its equivalence class is the given orientation; otherwise it is **negatively oriented**.

Now, given a positively oriented basis $b_1, \ldots, b_n$ for V, consider the n-form $\omega = b_1^* \wedge \cdots \wedge b_n^*$. If $(b'_1, \ldots, b'_n)$ is any other basis, with change of basis matrix B, then we can compute

$$\omega(b'_1, \ldots, b'_n) = \omega(Bb_1, \ldots, BB_n) = \det(B)\omega(b_1, \ldots, b_n) = \det(B).$$

Thus, $(b'_1, \ldots, b'_n)$ is positively oriented if and only if $\omega(b'_1, \ldots, b'_n) > 0$. This gives us an alternative way to define orientations in terms of top-forms.

PROPOSITION 10.3. *Let V be an n-dimensional vector space ($n \geq 1$). Each non-zero n-form $\omega \in \Lambda^n(V^*)$ determines an orientation for V as described above: a basis $b_1, \ldots, b_n$ is positively oriented iff $\omega(b_1, \ldots, b_n) > 0$. Two nonzero n-forms determine the same orientation iff they are positive scalar multiples of each other.*

PROOF. The preceding calculation shows how this works: given two ordered bases $(b_1, \dots, b_n)$ and $(b'_1, \dots, b'_n)$ with change of basis matrix B, we have

$$\omega(b'_1, \dots, b'_n) = \omega(Bb_1, \dots, Bb_n) = \det(B)\omega(b_1, \dots, b_n).$$

So the two bases are consistently oriented ($\det(B) > 0$) iff ω assigns them the same sign, which proves the first part. For the second part, we know $\Lambda^n(V^*)$ is 1-dimensional, so any two forms are scalar multiples of each other. Since they are both non-zero, the scalar multiple is either $+$ or $-$, and this precisely encodes positive or negative orientation. $\square$

Thus, an orientation could equally be described as an equivalence class of top-forms. This is the method we will use to generalize to manifolds.

DEFINITION 10.4. *Let M be a smooth manifold. A* **pointwise orientation** *for M is a choice of orientation for T_pM for each $p \in M$. A* **smooth orientation for** *M is a smooth n-form $\omega \in \Omega^n(M)$ which vanishes nowhere on M. A smooth manifold is called* **orientable** *if it possesses a smooth orientation ω. We call the pair (M, ω) an* **oriented manfiold***; often dropping the ω and just calling M an oriented manifold.*

As in Proposition 10.3, a smooth orientation form ω determines a pointwise orientation. This gives us a convenient way to talk about the orientation being smooth as it varies from point to point. It is possible to define this independently, locally: any coordinate chart gives a pointwise orientation within the chart, and one can then look for consistent orientation along the transition maps, giving a smoothness criterion. The use of forms is much easier and more computationally effective.

Not every manifold is orientable: the standard counterexamples being the Möbius strip, and Klein bottle. Another good class of examples of non-orientable manifolds are projective spaces. We will be focusing on orientable manifolds here; to prove a manifold is orientable, it is necessary and sufficient to specify a smooth non-vanishing top-form on the manifold.

Following the above definitions, a non-vanishing top form η on an oriented manifold M (with orientation form ω) is called **positively oriented** iff $\eta = f\omega$ for some strictly-positive function f. The form η is **negatively oriented** if $-\eta$ is positively oriented.

EXAMPLE 10.5. Let $M_1, \dots, M_k$ be orientable smooth manifolds. There is an orientation ω on $M_1 \times \cdots \times M_k$, called the **product orientation**, defined as follows: if ω_j is an orientation form on M_k, and $\pi_j \colon M_1 \times \cdots \times M_k \to M_j$ is the projection, then $\omega = \pi_1^* \wedge \cdots \wedge \pi_k^* \omega_k$. It is straightforward to check that this is a non-vanishing top-form, and so defines a smooth orientation. What's more, it is well-defined: if η_j are positively oriented forms on M_j, then $\pi_1^*\eta_1 \wedge \cdots \wedge \pi_k^*\eta_k$ is positively oriented with respect to ω, as the reader should readily check.

EXAMPLE 10.6. Let $U \subseteq M$ be a codimension-0 (i.e. open) submanifold of an oriented manifold M. If ω is an orientation form for M, and $\iota \colon U \hookrightarrow M$ is the inclusion map, then $\iota^*\omega$ is an orientation form for U.

EXAMPLE 10.7. Any parallelizable manifold is orientable. Indeed, let $X_1, \dots, X_n$ be vector fields on M^n with the property that $\{X_1|_p, \dots, X_n|_p\}$ is a basis for T_pM for each $p \in M$. This defines a pointwise orientation, simply by choosing the orientation given by the ordered basis $(X_1|_p, \dots, X_n|_p)$ in each tangent space T_pM. To show it is a smooth orientation, we could go a number of routes. For example, we could fix a Riemannian metric g on M (a symmetric 2-tensor field so that g_p is an inner product on T_pM for each p), and then define 1-forms $\eta_j(X) = g(X_j, X)$; then $\omega = \eta_1 \wedge \cdots \wedge \eta_n$ is an orientation form for the given pointwise orientation, as can be easily

checked. Note that the specific form ω depends on the choice of Riemannian metric; but we don't care about the specific form, only that one exists to induce this orientation.

Just as on a vector space, there are only two orientations on a *connected* manifold.

PROPOSITION 10.8. *Let M be a connected, orientable manifold. Then there are precisely two orientations: any non-vanishing top-form is either positively or negatively oriented. Moreover, if two forms are consistently oriented at any one point, then they are consistently oriented at all points.*

The proof is left as a homework exercise.

Now, let M, N be oriented manifolds, and let $F\colon M \to N$ be a local diffeomorphism. We call F **orientation-preserving** if, given any $p \in M$ and any positively oriented ordered basis $(X_1, \dots, X_n)$ for M, the ordered basis $(dF_p(X_1), \dots, dF_p(X_n))$ is positively oriented for $T_{F(p)}N$; if the image basis is always negatively oriented, we call F **orientation-reversing**. Note that a composition of orientation-preserving maps is orientation-preserving; a composition of orientation-reversing maps is also orientation-preserving; and a composition of an orientation-preserving map with an orientation-reversing map is orientation reversing (all easy to check).

LEMMA 10.9. *Let M, N be oriented manifolds, and let $F\colon M \to N$ be a local diffeomorphism. Then F is orientation-preserving if and only if, for each positively oriented form ω on N, $F^*\omega$ is positively oriented on M; F is orientation-reversing if and only if, for each positively oriented form ω on N, $F^*\omega$ is negatively oriented.*

The proof is simple definition chasing. This lemma leads to a nice way to *define* an orientation on a manifold using a local diffeomorphism.

PROPOSITION 10.10. *Let $F\colon M \to N$ be a local diffeomorphism. If N is oriented, then there is a unique orientation on M, called the* **pullback orientation induced by** *F, such that F is orientation-preserving.*

PROOF. For each $p \in M$, there are two orientations on T_pM, and one of them makes $dF_p\colon T_pM \to T_{F(p)}N$ orientation-preserving. So there is a unique pointwise orientation on M that does the trick. To see that it is actually a (smooth) orientation, note that $F^*\omega$ is an orientation form for it, by definition. $\square$

Example 10.7 shows that parallelizable manifolds are orientable, so we now know that the spheres $\mathbb{S}^1$, $\mathbb{S}^3$, and $\mathbb{S}^7$ are orientable. In fact, all spheres are orientable: on $\mathbb{R}^{n+1}$, let $\eta = \sum_{j=1}^{n+1} x_j(-1)^j\, dx_1 \wedge \cdots \wedge \widehat{dx^j} \wedge \cdots \wedge dx^{n+1}$, and $\omega = \iota^*\eta$ be the pullback along the inclusion $\iota\colon \mathbb{S}^n \hookrightarrow \mathbb{R}^{n+1}$ to a top-form on $\mathbb{S}^n$. Then ω is an orientation form for $\mathbb{S}^n$. A quick way to see this is to note that $\eta = E \lrcorner (dx^1 \wedge \cdots \wedge dx^n)$ where E is the Euler vector field $E = \sum_{j=1}^{n+1} x^j \frac{\partial}{\partial x^j}$. This is a *normal field* to $\mathbb{S}^n$: since the sphere is the regular level set of $F(x) = |x|^2$ (at height 1), by Proposition 9.34 the tangent space $T_x\mathbb{S}^n$ is the kernel of dF_x whose action is $dF_x(v) = 2\langle E_x, v\rangle$; thus the kernel is precisely the orthogonal complement to E_x. Thus, for any vectors $v_1, \dots, v_n$ at $x \in \mathbb{S}^{n+1}$,

$$\omega_x(v_1, \dots, v_n) = \eta_x(d\iota_x(v_1), \dots, d\iota_x(v_n)) = dx^1 \wedge \cdots \wedge dx^{n+1}(E_x, d\iota_x(v_1), \dots, d\iota_x(v_n)).$$

If we choose $v_1, \dots, v_n$ to be linearly independent, so are their images under the immersion $d\iota_x$; and they are all orthogonal to E_x, so $E_x, v_1, \dots, v_{n+1}$ are linearly independent, so the form $dx^1 \wedge \cdots \wedge dx^{n+1}$ (which is the determinant) does not vanish on them.

This kind of procedure works for any codimension-1 submanifold that possesses a nonvanishing "normal" field – in fact, all that is necessary is that the field be nowhere tangent. The same precise idea works for regular level sets of of smooth $\mathbb{R}$-valued functions in all oriented manifolds.

PROPOSITION 10.11. *Let M be an oriented manifold, and let $S \subseteq M$ be a codimension-1 embedded submanifold that is a regular level set of a smooth function $f\colon M \to \mathbb{R}$. Then S is orientable.*

PROOF. We will construct a nowhere tangent vector field to S as follows (generalizing the sphere case done above): choose a Riemnannian metrix g on M, and define a vector field ∇f to be the dual vector field to df: $df(X) = g(\nabla f, X)$. By the Riesz-Fisher theorem, this defines ∇f uniquely as a rough vector field. To see it is smooth, we compute in local coordinates: in a chart $(U, (x^j))$, g has components g_{ij}, and by symmetry $g_{ij} = g_{ji}$. The condition that g_p is an inner product means precisely that the matrix $[g_{ij}]_{i,j}$ is positive definite; in particular, it is invertible. The standard (terrible) notation for its inverse's components is to write the components *up*: $[g^{-1}]_{ij} = g^{ij}$. Now, the identity $df = g(\nabla f, \cdot\,)$ in local coordinates reads $(df)_j = \sum_i g_{ij}(\nabla f)^i$, which means that $(\nabla f)^i = \sum_j g^{ij}(df)_j$. Since f is smooth, the coefficient functions $(df)_j$ are smooth; since g is a smooth tensor field, its coefficient functions g_{ij} are smooth, and therefore by the inverse function theorem so are the raised coefficients g^{ij}. Thus, ∇f is a smooth vector field on M.

Now, suppose $X \in \mathscr{X}(M)$ is tangent to S. Since S is a regular f-level set, it follows from Proposition 9.34 that X_p is in the kernel of df_p for each $p \in S$. This means precisely that $df(X) = 0$, and so since $df = g(\nabla f, \cdot\,)$, it follows that X is g-orthogonal to ∇f.

Thus, follows exactly the same outline as above for the sphere, we see that if ω is an orientation form for M, then $\nabla f \lrcorner\, \omega$ is an orientation form for S. $\square$

2. Integration of Differential Forms

Let $\omega \in \Omega^n(\mathbb{R}^n)$ be a compactly supported top-form. Using our global Euclidean coordinates, this means that $\omega = f\, dx^1 \wedge \cdots \wedge dx^n$ for some smooth compactly-supported function f. We will then simply use the Lebesgue integral to integrate the function f: letting U be any measurable subset that contains the support of f, we define

$$\int_U \omega \equiv \int_U f\, dx^1 \cdots dx^n.$$

This looks like a trick of notation: we simply "erase the wedges". As we will see, this is very well-behaved: the only purpose of the wedges was to help us remember which orientation on $\mathbb{R}^n$ we used.

To begin, let us note that the definition does not depend on which set U: if U, U' are two measurable sets that contain the support of f then $f = 0$ on $U \triangle U'$ (symmetric difference) and this implies that $\int_U f\, dx^1 \cdots dx^n = \int_{U'} f\, dx^1 \cdots dx^n$. More importantly, this definition – which is prima facie coordinate dependent – is invariant under diffeomorphisms, up to the orientation.

PROPOSITION 10.12. *Let U, V be open subsets in $\mathbb{R}^n$, and let $G\colon U \to V$ be an orientation-preserving or orientation-reversing diffeomorphism. For any n-form compactly supported in V,*

$$\int_U G^*\omega = \begin{cases} \displaystyle\int_V \omega & \text{if } G \text{ is orientation-preserving}, \\ \displaystyle -\int_V \omega & \text{if } G \text{ is orientation-reversing}. \end{cases}$$

PROOF. Let $(y^1, \dots, y^n)$ be standard coordinates on V and $(x^1, \dots, x^n)$ standard coordinates on U. Any n-form compactly supported in V may be expressed as $\omega = f\, dy^1 \wedge \cdots \wedge dy^n$ for some smooth, compactly-supported function f on V. By the change of variables theorem from calculus,

$$\int_V \omega = \int_V f\, dy^1 \cdots dy^n = \int_U (f \circ G)|\det[DG]|\, dx^1 \cdots dx^n.$$

Suppose that G is orientation-preserving. This means that DG maps the standard basis (in x^j-coordinates) to a positively-orientated bases (relative to the y^j-basis), which is precisely to say that $\det[DG] > 0$. Thus

$$\int_V \omega = \int_U (f \circ G)(\det[DG])\, dx^1 \cdots dx^n \equiv \int_U \eta, \quad \text{where} \quad \eta = (f \circ G)(\det[DG])\, dx^1 \wedge \cdots \wedge dx^n.$$

Now using Proposition 8.13 (the change of variables formula for pulling back top-forms), we have $\eta = G^*(f\, dy^1 \wedge \cdots \wedge dy^n) = G^*(\omega)$ thus proving the first half of the proposition. If G is orientation-reversing, the only difference is that $|\det[DG]| = -\det[DG]$, introducing the global minus-sign. □

This leads the way to defining the integral of a top-form on an oriented manifold. Let M be an oriented n-manifold, and let ω be an n-form that is compactly supported within a single chart (U, φ) of M. Note that $\varphi\colon U \to \hat{U} \subseteq \mathbb{R}^n$ is a diffeomorphism, and we fix the standard orientation on $\mathbb{R}^n$, so φ is either orientation-preserveing or orientation-reversing. We then define

$$\int_M \omega \equiv \pm \int_{\varphi(U)} (\varphi^{-1})^*\omega \tag{10.1}$$

with the $+$ in the case φ is orientation-preserving, and $-$ in the case it is orientation-reversing.

Aside from the deficiency that this only covers forms supported in a single chart, it is also prima facie dependent on the chart (U, φ) in question. In fact, this is not so.

LEMMA 10.13. *If (U, φ) and (V, ψ) are two charts on an n-manifold M, and ω is an n-form compactly supported in $U \cap V$, then*

$$\int_{\varphi(U)} (\varphi^{-1})^*\omega = \pm \int_{\psi(V)} (\psi^{-1})^*\omega$$

with the $+$ in the case that φ and ψ are both orientation-preserving or both orientation-reversing, and $-$ otherwise.

PROOF. Supposing φ and ψ are either both orientation-preserving of both orientation-reversing, $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ is an orientation-preserving diffeomorphism. Thus, Proposition 10.12 shows that

$$\int_{\psi(V)} (\psi^{-1})^*\omega = \int_{\psi(U \cap V)} (\varphi^{-1})^*\omega = \int_{\varphi(U \cap V)} (\psi \circ \varphi^{-1})^*(\psi^{-1})^*\omega = \int_{\varphi(U \cap V)} (\varphi^{-1})^*\omega,$$

and this equals the desired quantity (as in the first equality above) since $(\varphi^{-1})^*\omega$ is equal to 0 on $\varphi(U) \setminus \varphi(U \cap V)$. This proves the proposition in this case; the case with the $-$ sign comes from the fact that $\psi \circ \varphi^{-1}$ is orientation-reversing in this case. $\square$

Thus, 10.1 is well-defined: if a different chart is chosen, we have the same value for $\int_M \omega$. It now behooves us to extend the definition to the case that ω need not be supported in a single chart. To that end, we use a partition of unity. Let $\omega \in \Omega^n(M)$ be compactly-supported, and fix a finite open cover $\{(U_j, \varphi_j)\}$ of $\mathrm{supp}\omega$ by charts that are all orientation-preserving (with respect to the standard orientation on their target $\mathbb{R}^n$). Let $\{f_j\}$ be a partition of unity subordinate to $\{U_j\}$. We define

$$\int_M \omega \equiv \sum_j \int_M f_j \omega \tag{10.2}$$

subject to our definition (10.1) of the integral of a form supported in a single chart (since $\psi_j\omega$ is compactly supported in U_j). This definition prima facie depends on the choice of charts *and* of the partition of unity. Just as the integral in a single chart is actually chart dependent, this definition is also well-defined: it does not depend on the partition of unity.

PROPOSITION 10.14. *Let $\{(U_j, \varphi_j)\}$ and $\{(V_j, \psi_j)\}$ be finite covers of $\mathrm{supp}\omega$ by positively-oriented charts, and let $\{f_j\}$ and $\{g_j\}$ be subordinate partitions of unity. Then*

$$\sum_j \int_M f_j \omega = \sum_j \int_M g_j \omega.$$

PROOF. We simply use the defining properties of partitions of unity, to compute that, for each i,

$$\int_M f_i \omega = \int_M \left(\sum_j g_j \right) f_i \omega = \sum_j \int_M g_j f_i \omega.$$

Summing over i, we then have

$$\sum_i \int_M f_i \omega = \sum_{i,j} \int_M g_j f_i \omega.$$

Inside the sum, each integral is well-defined regardless of which chart containing $\mathrm{supp}(g_j f_i \omega)$ we use (either U_i or V_j), by Lemma 10.13. Hence, we may write this instead as

$$\sum_{i,j} \int_M g_j f_i \omega = \sum_j \int_M \left(\sum_i f_i \right) g_j \omega = \sum_j \int_M g_i \omega$$

as desired. $\square$

Hence, we have a well-defined notion of the integral of a (compactly-supported) n-form on any oriented n-manifold. One common source of top-forms is from form on ambient manifolds: if $\iota\colon S \hookrightarrow M$ is an oriented embedded submanifold of dimension $n \leq \dim(M)$, and $\omega \in \Omega^n(M)$, then $\iota^*\omega \in \Omega^n(S)$. We generally shorten notation and simply write

$$\int_S \omega = \int_S \iota^*\omega.$$

Let us note that it is possible to make sense of integral of (some!) non-compactly-supported forms, but this requires a much richer integration theory mirroring Lebesgue integration. We will not need this level of generality, so we will content ourselves with the compactly-supported case.

PROPOSITION 10.15 (Properties of Integrals of Differential Forms). *Let M, N be oriented n-manifolds ($n \geq 1$), and let $\omega, \eta \in \Omega^n(M)$ be compactly-supported.*

(a) LINEARITY*: If $a, b \in \mathbb{R}$, then*

$$\int_M a\omega + b\eta = a\int_M \omega + b\int_M \eta.$$

(b) ORIENTATION REVERSAL*: Let $-M$ denote the manifold M with the oppposite orientation. Then*

$$\int_{-M} \omega = -\int_M \omega.$$

(c) POSITIVITY*: If ω is a positively-oriented orientation form, then $\displaystyle\int_M \omega > 0$.*

(d) DIFFEOMORPHISM INVARIANCE*: If $F\colon M \to N$ is an orientation-preserving or orientation-reversing diffeomorphism, then*

$$\int_M \omega = \begin{cases} \displaystyle\int_N F^*\omega & \text{if } F \text{ is orientation-preserving,} \\ \displaystyle -\int_N F^*\omega & \text{if } F \text{ is orientation-reversing.} \end{cases}$$

PROOF. Parts (a) and (b) are trivial definition chasing. Part (c) is immediate in local coordinates, and so follows from the definition (via partition of unity, consisting of non-negative bump functions). Similarly, (d) can be reduced to local coordintes since (using the partition of unity definition) ω is a finite sum of orientation forms each supported in a single chart, and then the result follows from Proposition 10.12. □

Although the definition of the integral, using a partition of unity, is useful for proving all of the above properties, it is completely useless for actual computation: one can basically never explicitly write down a partition of unity, and even if one can, it will involve functions like $e^{-1/x}$, which is not the kind of function you want hanging around when you're trying to do an integral. Instead, to actually compute the integral of a form, we always use (local) parametrizations.

PROPOSITION 10.16. *Let M be an oriented n-manifold, and let $\omega \in \Omega^n(M)$ be compactly-supported. Let $D_1, \ldots, D_k$ be open subsets of $\mathbb{R}^n$ such that ∂D_j has measure 0 for each j, with smooth maps $F_j\colon \overline{D_j} \to M$ satisfying*

(i) *$F_j|_{D_j}\colon D_j \to F(D_j)$ is an orientation-preserving diffeomorphism.*
(ii) *$F(D_i) \cap F(D_j) = \varnothing$ when $i \neq j$.*
(iii) $\operatorname{supp}(\omega) \subseteq \overline{F(D_1)} \cup \cdots \cup \overline{F(D_k)}$.

Then

$$\int_M \omega = \sum_{j=1}^k \int_{D_j} F_j^*\omega.$$

PROOF. By the linearity of both sides of the desired equality, it sufficies to prove the proposition in the case that ω is compactly-supported inside a single chart (by using a partition of unity to express ω in general as a sum of such forms). By choosing our covering charts carefully to begin with, it suffices to assume that $\operatorname{supp}(\omega)$ is contained in a chart domain U for which $\overline{U}$ is compact, $\varphi(U)$ is an open ball in $\mathbb{R}^n$, and φ extends to a homeomorphism $\overline{U} \to \overline{\varphi(U)}$.

We now resrict our parametrization domains to this chart as follow: $A_j \subseteq D_j$, $B_j \subseteq F(D_j)$, and $C_j \subseteq \varphi(U)$ are the sets

$$A_j = F_j^{-1}(U \cap F(D_j)), \quad B_j = U \cap F_j(D_j) = F_j(A_j), \quad C_j = \varphi(B_j) = \varphi \circ F_j(A_j).$$

The support of $(\varphi^{-1})^*\omega$ is contained in $\overline{C_1} \cup \cdots \cup \overline{C_k}$, and any two of these intersect along their boundaries. Note that $\partial C_j = \varphi(\partial B_j)$, and B_j is the intersection of an open ball with $F_j(D_j)$. Since ∂D_j has measure 0, it follows that ∂C_j also heas measure 0, and so is ignored by the integral; thus

$$\int_M \omega = \int_{\varphi(U)} (\varphi^{-1})^*\omega = \sum_{j=1}^k \int_{C_j} (\varphi^{-1})^*\omega.$$

We now compute using the change of variables (Proposition 10.12),

$$\int_{C_j} (\varphi^{-1})^*\omega = \int_{A_j} (\varphi \circ F_j)^*(\varphi^{-1})^*\omega = \int_{A_j} F_j^*\omega = \int_{D_j} F_j^*\omega,$$

and summing over j proves the desired result. □

EXAMPLE 10.17. We saw above that $\omega = x\,dy\wedge dz + y\,dz\wedge dx + z\,dx\wedge dy$ is an orientation form on $\mathbb{S}^2$. Let's compute the integral of ω over the sphere. To do this, we use spherical coordinates:

$$F(\phi, \theta) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$$

which is a diffeomorphism from $(0, \pi) \times (0, 2\pi)$ into the open set in $\mathbb{S}^2$ given by the removal of the arc from the north to the south pole in the $x \geq 0$ half of the $y = 0$ plane. Since $F([0, \pi] \times [0, 2\pi])$ covers the sphere, and F is orientation-preserving (by definition if we use ω to orient $\mathbb{S}^2$), it follows from the preceding proposition that

$$\int_{\mathbb{S}^2} \omega = \int_0^\pi \int_0^{2\pi} F^*\omega.$$

This is now a standard double integral. It is left to the bored reader to compute that $F^*\omega = \cos\phi\, d\phi \wedge d\theta$. Thus

$$\int_{\mathbb{S}^2} \omega = \int_0^\pi \int_0^{2\pi} \cos\phi\, d\phi \wedge d\theta = 4\pi.$$

This example gives the usual volume form on $\mathbb{S}^2$. In general, an orientation form can be though of as a choice of volume form. We will use this idea in the study of Lie groups.

Part 2

Lie Groups.

CHAPTER 11

Lie Groups, Subgroups, and Homomorphisms

DEFINITION 11.1. *A* **Lie group** *is a smooth manifold G which is also a group, and has the property that the group operations*

$$\mathrm{m}(g,h) = gh, \qquad \mathrm{inv}(g) = g^{-1}$$

are smooth maps $\mathrm{m}\colon G \times G \to G$ *and* $\mathrm{inv}\colon G \to G$.

The notation above is typical: Lie groups are G, H, K (with K usually denoting a compact group), elements are $g, h \in G$ (although we may also use $a, b, c, \ldots$ or p, q or x, y, z). The identity element of the group may be denoted 1_G, but is more commonly called e (from the German *Einselement*, "unit element").

The multiplication map gives rise to two all-important families of diffeomorphisms of G: the **left-translation** and **right-translation** maps $L_g, R_g\colon G \to G$ for $g \in G$:

$$L_g(h) = gh, \qquad R_g(h) = hg.$$

These are smooth maps: letting $\iota_g\colon G \to G \times G$ be the map $\iota_g(h) = (g,h)$ which is clearly smooth, note that $L_g = \mathrm{m} \circ \iota_g$ is smooth as well. (A similar argument shows R_g is smooth.) Since L_g, R_g are bijections whose inverses are $L_{g^{-1}}, R_{g^{-1}}$ which are also smooth, L_g, R_g are diffeomorphisms for all $g \in G$.

As a first application of the efficacy of these maps, we note that the two conditions in Definition 11.1 are redundant: smoothness of multiplication alone is sufficient.

LEMMA 11.2. *If G is a smooth manifold that is a group such that the multiplication map* $\mathrm{m}\colon G \times G \to G$ *is smooth, then G is a Lie group.*

PROOF. Let $F\colon G \times G \to G \times G$ be the map $F(g,h) = (g, gh)$. By assumption, F is smooth. Note that it is also a bijection: the inverse is $F^{-1}(g,h) = (g, g^{-1}h)$. We will show that F is a local diffeomorphism: it then follows (since it is a bijection) that F is a diffeomorphism, and hence its inverse is smooth. This means that the map $(g,h) \mapsto g^{-1}h$ (the second component of F^{-1}) is smooth, and therefore the map $g \mapsto (g,e) \mapsto g^{-1}e = g^{-1}$ is smooth, as claimed.

Thus, we need to compute the differential of F. Since $F(g,h) = (g, \mathrm{m}(g,h))$, we have

$$dF_{g,h}(X,Y) = (X, d\mathrm{m}_{g,h}(X,Y)), \qquad X \in T_gG, Y \in T_hG.$$

So, let $\alpha, \beta\colon (-\epsilon, \epsilon) \to G$ be smooth curves with $\alpha(0) = g$ and $\dot\alpha(0) = X$, while $\beta(0) = h$ and $\dot\beta(0) = Y$. Then, by definition,

$$dm_{g,h}(X,Y) = \frac{d}{dt}\bigg|_0 \mathrm{m}(\alpha(t), \beta(t)).$$

For any test function $f \in C^\infty(G)$,

$$\frac{d}{dt}\bigg|_0 \mathrm{m}(\alpha(t), \beta(t))(f) \equiv \frac{d}{dt}\bigg|_0 (f \circ \mathrm{m})(\alpha(t), \beta(t)).$$

We now apply the chain rule, giving

$$\frac{d}{dt}\bigg|_0 (f \circ \mathrm{m})(\alpha(t), \beta(t)) = \frac{d}{dt}\bigg|_0 (f \circ \mathrm{m})(\alpha(t), \beta(0)) + \frac{d}{dt}\bigg|_0 (f \circ \mathrm{m})(\alpha(0), \beta(t))$$

which shows by definition that

$$\begin{aligned}\frac{d}{dt}\bigg|_0 \mathrm{m}(\alpha(t), \beta(t)) &= \frac{d}{dt}\bigg|_0 \mathrm{m}(\alpha(t), h) + \frac{d}{dt}\bigg|_0 \mathrm{m}(g, \beta(t)) \\ &= \frac{d}{dt}\bigg|_0 R_h(\alpha(t)) + \frac{d}{dt}\bigg|_0 L_h(\beta(t)) \\ &= dR_h|_g(X) + dL_g|_h(Y).\end{aligned}$$

All together, then, we have

$$dF_{g,h}(X, Y) = (X, dR_h|_g(X) + dL_g|_h(Y)).$$

We know that $R_h, L_g \colon G \to G$ are diffeomorphisms, so the differentials $dR_h|_g \colon T_gG \to T_{hg}G$ and $dL_g|_g \colon T_hG \to T_{gh}G$ are isomorphisms.

So, suppose $(X, Y) \in \ker(dF_{g,h})$. Then $X = 0$ and $dR_h|_g(X) + dL_g|_h(Y) = 0$, thus $dL_g|_h(Y) = 0$ and since $dL_g|_h$ is an isomorphism, $Y = 0$. This shows $dF_{g,h}$ is injective. Since the domain $T_{g,h}G \times G$ and codomain $T_{g,gh}G \times G$ of F has the same dimension $2\dim(G)$, it follows that $dF_{g,h}$ is a linear isomorphism, so F is a local diffeomorphism, concluding the proof. □

1. Examples

The first example of a Lie group is the group of invertible matrices.

EXAMPLE 11.3. Let $\mathrm{GL}_n(\mathbb{F})$ denote the **general linear group** of invertible $n \times n$ matrices over the field $\mathbb{F}$. With $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, we know this is an open subset of the vector space of $n \times n$ matrices over $\mathbb{R}$ (dimension n^2) or $\mathbb{C}$ (dimension $2n^2$), and hence is a smooth manifold of dimension n^2 or $2n^2$. Matrix multiplication is a smooth map (indeed, its coordinate functions are all polynomials), and hence $\mathrm{GL}_n(\mathbb{F})$ and $\mathrm{GL}_n(\mathbb{C})$ are both Lie groups.

One easy way to generate Lie groups is as subgroups of a given Lie group, in the following sense.

LEMMA 11.4. *Let G be a Lie group, and let $H \subseteq G$ be a subgroup that is also an embedded submanifold. Then H is a Lie group.*

PROOF. By definition, the inclusion map $\iota_H \colon H \hookrightarrow G$ is a smooth embedding, and so H is a smooth manifold in its own right. Since H is a subgroup, its multiplication map m_H is the restriction of the multiplication map m_G on G: $\mathrm{m}_H = \mathrm{m}_G \circ (\iota_H \times \iota_H)$. This is a composition of smooth maps, hence is smooth. □

A subgroup satisfying the conditions of Lemma 11.4 is called a(n embedded) **Lie subgroup**. (Some authors also allow immersed subgroups to be called Lie subgroups.) Now let us delve into a long list of examples of Lie groups, mostly given as Lie subgroups of $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{GL}_n(\mathbb{C})$.

EXAMPLE 11.5. (1) We denote $\mathrm{GL}_1(\mathbb{R}) = \mathbb{R}^*$ and $\mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^*$ the multiplicative groups of real and complex numbers.

(2) The subgroup $\mathrm{GL}_n^+(\mathbb{R})$ of $\mathrm{GL}_n(\mathbb{R})$ of those invertible matrices whose determinant is > 0. Since $\det(AB) = \det(A)\det(B)$, and $\det(I_n) = 1 > 0$, this is indeed a subgroup. Moreover, by continuity of $\det$, it is an open subset, so it is an embedded submanifold of codimension 0. Thus, it is a Lie subgroup.

(3) The **special linear groups** $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{C})$ are defined to be the subgroups of GL_n with determinant 1. Since $\det(AB) = \det(A)\det(B)$, SL_n is a subgroup. We need to check that it is a Lie subgroup. Since it is defined as the level set of a smooth function $\det$, is sufficies to show that it is a regular level set. To see this, we use the following formula: for any $A \in \mathrm{GL}_n$ and $X \in \mathrm{M}_n$,

$$\left.\frac{d}{dt}\right|_0 \det(A + tX) = \det(A)\mathrm{Tr}(A^{-1}X).$$

To prove this, first note that

$$A + tX = A(I + tA^{-1}X) = Ae^{tA^{-1}X} + o(t).$$

Thus, by smoothness of $\det$,

$$\det(A + tX) = \det(Ae^{tA^{-1}X}) + o(t) = \det(A)\det(e^{tA^{-1}X}) + o(t).$$

Now, if Y is a diagonalizable matrix $Y = P^{-1}DP$, then we have

$$\begin{aligned}\det(e^Y) = \det(e^{P^{-1}DP}) &= \det(P^{-1}e^DP)\\ &= \det(e^D) = e^{D_{11}} \cdots e^{D_{nn}} = e^{\mathrm{Tr}(D)} = e^{\mathrm{Tr}(P^{-1}DP)} = e^{\mathrm{Tr}\,(Y)}.\end{aligned}$$

Since the set of diagonalizable matrices is dense, and $\det$ and Tr are continuous, it follows that the identity $\det(e^Y) = e^{\mathrm{Tr}(Y)}$ holds for all matrices Y. Thus, w ehave

$$\det(A + tY) = \det(A)e^{t\mathrm{Tr}(tA^{-1}X)}.$$

Taking the derivative now gives

$$D(\det)_A(X) = \left.\frac{d}{dt}\right|_0 \det(A + tX) = \det(A)\mathrm{Tr}(A^{-1}X).$$

For any $A \in \mathrm{GL}_n$, so that $\det(A) = 1$, we therefore have $D(\det)_A(X) = \mathrm{Tr}(A^{-1}X)$. This linear map is clearly surjective: for any real or complex number c, $D(\det)_A(cA/n) = \mathrm{Tr}(A^{-1}Ac/n) = c/n \cdot \mathrm{Tr}(I) = c$. Thus, SL_n is a regular level set of the smooth function $\det$, and so it is an embedded submanifold of GL_n of codimension 1 in the real case . Thus, is a Lie group of dimension $n^2 - 1$ in the real case and $2n^2 - 2$ in the complex case.

(4) The **orthogonal group** $\mathrm{O}_n = \mathrm{O}_n(\mathbb{R})$ and **unitary group** $\mathrm{U}_n = \mathrm{U}_n(\mathbb{C})$ are the groups of $\mathbb{R}$-linear / $\mathbb{C}$-linear isometries of $\mathbb{R}^n$ / $\mathbb{C}^n$. In particular, this means that

$$\mathrm{O}_n = \{Q \in \mathbb{M}_n(\mathbb{R}) \colon Q^\top Q = I\}, \mathrm{U}_n = \{U \in \mathbb{M}_n(\mathbb{C}) \colon U^*U = I\}.$$

They are therefore level sets of the smooth functions $A \mapsto A^\top A$ or $A \mapsto A^*A$. Viewing the range not as the full space $\mathbb{M}_n$ by rather the space $\mathbb{M}_n^{\mathrm{s.a.}}$ of *self-adjoint* matrices ($X = X^\top$ in the real case, $X = X^*$ in the complex case), we showed in Example 1.5 that O_n is a regular level set; a nearly identical proof shows the same for U_n. Since the group operation is still given by matrix multiplication which is smooth, O_n and U_n are also Lie groups. Since the codomain $\mathbb{M}_n^{\mathrm{s.a.}}(\mathbb{R})$ of the defining function of O_n has dimension $\frac{n(n+1)}{2}$, that's the codimension of O_n; thus $\dim(\mathrm{O}_n) = \frac{n(n-1)}{2}$. On the other hand, the space $\mathbb{M}_n^{\mathrm{s.a.}}(\mathbb{C})$

has dimension n^2: the upper-triangular entries are all independent complex coordinates, giving $2 \cdot \frac{n(n-1)}{2}$ coordinates, and the diagonal has n independent real coordinates. Thus, $\dim(\mathrm{U}_n) = 2n^2 - n^2 = n^2$.

(5) Combining the last two examples: suppose $Q \in \mathrm{O}_n$. Then $Q^\top Q = I$, and so $1 = \det(Q^\top Q) = \det(Q)^2$, meaning $\det(Q) = \pm 1$. In fact, O_n has two connected components, given by $\det^{-1}(\pm 1)$. These components are open subsets. By the homomorphism property of $\det$, the component of the identity (where $\det Q = 1$) is a subgroup. It is called the **special orthogonal group** SO_n. As an open subgroup, it has the same dimension as O_n.

The complex case is different: the same argument shows that, for $U \in \mathrm{U}_n$, $|\det U| = 1$, so in fact $\det\colon \mathrm{U}_n \to \mathbb{S}^1$. Now one can use the same argument we did in the case of SL_n to see that $\det$ is full rank onto the circle, and thus the subgroup $\mathrm{SU}_n = \{U \in \mathrm{U}_n\colon \det U = 1\}$ is an embedded submanifold of codimension 1. Thus SU_n is a Lie group of dimension $n^2 - 1$.

(6) We can generalize the construction of O_n by taking a different inner product. Any inner product on $\mathbb{R}^n$ has the form $\langle x, y\rangle = x^\top A y$ for some positive definite matrix A; then we might define O_n^A to be the set of matrices Q with $Q^\top A Q = A$ (which is the same as requiring Q to be a linear isometry of the inner product given by A). All the same arguments show this is a Lie group; in fact, it is the image of O_n under the map $Q \mapsto A^{-1/2} Q A^{1/2}$, which gives (as we'll discuss below) a Lie group isomorphism.

But we could also take a matrix A that is not positive definite. If A is degenerate, the resulting object will not be a manifold. But we could take a non-degenerate *anti*symmetric A. The canonical example is to work in even dimension and take $J \in \mathbb{M}_{2n}$:

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

The **symplectic group** $\mathrm{Sp}(n, \mathbb{R})$ is the set of matrices $A \in \mathbb{M}_{2n}$ such that $A^\top J A = J$. We could also do this over $\mathbb{C}$, noting that in this case we still have to take $\top$ and not $*$, to give $\mathrm{Sp}(n, \mathbb{C})$. Arguments just like those for O_n show that these are Lie groups. The same computations as in the previous example show that, if $A \in \mathrm{Sp}(n)$, then $\det A = \pm 1$. It is less obvious, but not too hard to show, that $\det A = 1$ always. So there is no "special symplectic group" – it's already special enough!

Note: some authors use the notation $\mathrm{Sp}(n)$ to refer to the **compact symplectic group**, which is (in our notation) $\mathrm{Sp}(n, \mathbb{C}) \cap \mathrm{U}_{2n}(\mathbb{C})$.

(7) The **Heisenberg group** $\mathrm{H}_n(\mathbb{R})$ is the subgroup of $\mathrm{GL}_n(\mathbb{R})$ consisting of matrices of the form $I + T$ where T is a strictly upper-triangular matrix. It is easy to see this is a manifold: its $\frac{n(n-1)}{2}$ upper-triangular entries are global coordinates making it isomorphic as a manifold to $\mathbb{R}^{n(n-1)/2}$. It is therefore a Lie group since (as usual) matrix multiplication is smooth. Sometimes "Heisenberg group" refers to just $\mathrm{H}_3(\mathbb{R})$, which is deeply connected to quantum mechanics (as we'll explore a little later).

(8) The **Euclidean group** $\mathrm{E}(n, \mathbb{R})$ is the set of bijections of $\mathbb{R}^n$ that preserve the Euclidean distance. So $\mathrm{O}_n \subset \mathrm{E}(n, \mathbb{R})$, but they are not equal: for example, the translation map $T_x(y) = y + x$ is in $\mathrm{E}(n, \mathbb{R})$. It is a theorem that this covers everything: every element of $\mathrm{E}(n, \mathbb{R})$ can be written uniquely in the form $T_x Q$ for some translation T_x and orthogonal

map Q. By writing out how the product of two such tranformation acts, we see that we can view $\mathrm{E}(n,\mathbb{R})$ as the subgroup of O_{n+1} of elements of the form

$$\begin{bmatrix} Q & x \\ 0 & 1 \end{bmatrix}.$$

It is then straightforward to build charts for $\mathrm{E}(n,\mathbb{R})$ out of any atlas for O_n, showing that $\mathrm{E}(n,\mathbb{R})$ is a manifold. It is globally isomorphic to $\mathrm{O}_n \times \mathbb{R}^n$, and so has dimension $\frac{n(n-1)}{2} + n = \frac{n(n+1)}{2}$.

We can also build new examples as products of old ones:

LEMMA 11.6. *If $G_1, \dots, G_k$ are Lie groups, so is their product $G_1 \times \cdots \times G_k$ (given the manifold structure and group structure of the product).*

This is easy to check. For example, since $\mathbb{S}^1 \cong \mathrm{SO}_2$ is a Lie group, so is the torus $\mathbb{T}^n = (\mathbb{S}^1)^n$. Note: the final example above $\mathrm{E}(n,\mathbb{R})$ is *not* $\mathrm{O}_n \times \mathbb{R}^n$ as a Lie group. As a manifold, this is true, but the group structure is different. In fact it is a *semidirect product*, as we'll discuss soon. First we need to discuss the morphisms in the category of Lie groups.

2. Lie Group Homomorphisms

DEFINITION 11.7. *Let G, H be Lie groups. A* **Lie group homomorphism** *is a smooth map $F\colon G \to H$ that is also a group homomorphism. a* **Lie group isomorphism** *is a Lie group homomorphism that is also a diffeomorphism; hence, its inverse is also a Lie group isomorphism. If there exists a Lie group isomorphism $G \to H$, we say G and H are* **isomorphic** *as Lie groups. A Lie group isomorphism from G to itself is called a* **Lie group automorphism**.

The map $F\colon \mathrm{O}_n \times \mathbb{R}^n \to \mathrm{E}(n,\mathbb{R})$ given by $(Q,x) \mapsto T_xQ$ is a diffeomorphism (as the matrix representation above shows), but it is not a group homomorphism; so, although these two groups are diffeomorphic as manifolds, they are not isomorphic as Lie groups. (To be clear: we have not *shown* that they are not isomorphic, we have only shown that *this* diffeomorphism is *not* an isomorphism. In fact, they are not isomorphic, as we'll prove later.)

EXAMPLE 11.8. Here are some examples of Lie group homomorphisms.

(1) If $H \subseteq G$ is an embedded Lie subgroup, then the inclusion $H \hookrightarrow G$ is a Lie group homomorphism. Examples 11.5(2-6) are all embedded Lie subgroups of GL_n, and so their inclusions are all Lie group homomorphisms.
(2) The exponential map $\exp\colon \mathbb{R} \to \mathbb{R}^*$ or $\exp\colon \mathbb{C} \to \mathbb{C}^*$ is a homomorphism: it is a homomorphism of groups (with the additive group structure on $\mathbb{R}, \mathbb{C}$ and the multiplicative ones on $\mathbb{R}^*, \mathbb{C}^*$), and it is smooth, so it is a Lie group homomorphism. It is also injective. It is not surjective, however: its image is equal to $\mathbb{R}_+ = (0,\infty)$. In fact, $\exp\colon \mathbb{R} \to \mathbb{R}_+$ is a Lie group isomorphism, since it has an inverse $\log\colon \mathbb{R}_+ \to \mathbb{R}$ which is also smooth. In the complex case, the exponential map is surjective but not injective: its kernel is $2\pi i\mathbb{Z}$.
(3) The map $\varepsilon\colon \mathbb{R} \to \mathbb{S}^1$ given by $\varepsilon(t) = \exp(2\pi i t)$ is a surjective Lie group homomorphism, whose kernel is $\mathbb{Z}$. Similarly, $\varepsilon^n\colon \mathbb{R}^n \to \mathbb{T}^n$ is a Lie group homomorphism whose kernel is the integer lattice $\mathbb{Z}^n$.
(4) The determinant $\det\colon \mathrm{GL}_n(\mathbb{F}) \to \mathbb{F}$ with $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ (additive groups) is a Lie group homomorphism – it is smooth because it is a polynomial in the entries. It is never an isomorphism if $n > 1$, of course.

(5) Let G be a Lie group, and let $g \in G$. The **inner automorphism** of G is the map $C_g \colon G \to G$ given by $C_g(h) = ghg^{-1}$ (conjugation by g). Because multiplication and inversion are smooth, C_g is smooth; inner automorphisms are group isomorphisms, so this is a Lie group automorphism. As a reminder: a subgroup H is called **normal** if $C_g(H) = H$ for all $g \in G$.

Lie group homomorphisms are the best kind of smooth maps: they have constant rank.

THEOREM 11.9. *Every Lie group homomorphism has constant rank.*

PROOF. Let G, H be Lie groups, and let $F \colon G \to H$ be a Lie group homomorphism. Let $g_0 \in G$, and denote the identity of G as e_G (and the identity of H as e_H). Since F is a homomorphism, we have, for all $g \in G$,

$$F(L_{g_0}(g)) = F(g_0 g) = F(g_0)F(g) = L_{F(g_0)}(F(g)).$$

That is: $F \circ L_{g_0} = L_{F(g_0)} \circ F$. Taking differentials of both sides at the identity, the chain rule then tells us

$$dF_{g_0} \circ d(L_{g_0})_{e_G} = d(L_{F(g_0)})_{e_H} \circ dF_{e_G}.$$

Since L_{g_0} and $L_{F(g_0)}$ are diffeomorphisms, their differentials at any points are isomorphisms. It follows therefore that dF_{g_0} has the same rank as dF_{e_G}. As this holds true for any g_0, we see that dF_{g_0} has constant rank. □

A corollary is that the relation between "Lie group homomorphism" and "Lie group isomorphism" is the same as the relation between "group homomorphism" and "group isomorphism": it is just a question of being a bijections.

COROLLARY 11.10. *A Lie group homomorphism is a Lie group isomorphism iff it is a bijection.*

PROOF. The 'only if' direction is in the definition of isomorphism. For the 'if' direction, we apply the Global Rank Theorem 9.10: since the Lie group homomorphism is constant rank, if it is a bijection, it is therefore a diffeomorphism, hence a Lie group isomorphism. □

3. Lie Subgroups

As we saw in Lemma 11.4, if $H \subseteq G$ is a subgroup that is also an embedded submanifold, then it is a Lie group in its own right. The simplest such example is an *open* subgroup. However, the group structure poses serious restrictions on what kind of open submanifolds are subgroups.

LEMMA 11.11. *Let $H \subseteq G$ be an open subgroup of a Lie group. Then H is an embedded Lie subroup. In addition, H is closed, and so it is a union of connected components of G.*

PROOF. Any open submanifold is embedded, and so the first statement follows from Lemma 11.4. Now, for any $g \in G$, the left coset $gH = \{gh \colon h \in H\} = L_g(H)$ is open, since it is the image of the open set H under the diffeomorphism L_g. The cosets of any group are all disjoint, and so $G \setminus H$ is the union of all cosets excluding H itself. Thus $G \setminus H$ is a union of open sets, so is open, and thus H is closed. Therefore H is clopen, and hence it is a union of connected components. □

In fact, the components of a Lie group are all diffeomorphic. To prove this, we first study the component containing e_G.

PROPOSITION 11.12. *Let G be a Lie group, and let $W \subseteq G$ be a neighborhood of G.*

(a) *The subgroup of G generated by W is open.*
(b) *If W is connected, the subgroup generated by W is a connected, the subgroup it generates is connected.*
(c) *If G is connected, then W generates G.*

Note: the subgroup generated by W is the smallest subgroup containing W. Alternatively, it can be described as the set of all finite words in elements in $W \cup W^{-1}$, as can be readily verified.

PROOF. Let $W \subseteq G$ be a neighborhood of the identity, and let H be the subgroup it generates. For $k \in \mathbb{N}$, let H_k denote the set of all elements of G that are equal to a product of k or fewer elements from $W \cup W^{-1}$; so $H = \bigcup_k H_k$. Now, $W^{-1} = \mathrm{inv}(W)$ is the image of the open set W under a diffeomorphism, so it is open. Hence $H_1 = W \cup W^{-1}$ is open. For each $k > 1$, note that

$$H_k = H_1 H_{k-1} = \bigcup_{g \in H_1} L_g(H_{k-1}).$$

Proceeding by induction, once we've shown H_{k-1} is open, its image under the diffeomorphism L_g is open, and so H_k is a union of open sets, hence is open. Thus, H_k is open for each k, and therefore its union H is open, proving (a).

Now, suppose W is connected. Then so is W^{-1} (since it is the image of a connected set under a diffeomorphism). Then $H_1 = W \cup W^{-1}$ is the union of two connected sets that intersect (at e_G), hence H_1 is connected. Now we proceed by induction again: suppose we've shown H_{k-1} is connected. Then the Cartesion product $H_1 \times H_{k-1}$ is connected, and so $H_k = H_1 H_{k-1} = \mathrm{m}(H_1 \times H_{k-1})$ is connected since it is the image of a connected set under a continuous map. Thus, H_k is connected for all k. Finally, this shows $H = \bigcup_k H_k$ is connected, sicne all H_k intersect at e_G, proving (b).

Item (c) now follows from Lemma 11.11: since H is an open subgroup, it is also closed, and hence if G is connected and $e_G \in H \neq \varnothing$, $H = G$. □

The connected component of G that contains e_G is called the **identity component**, and is often denoted G_0. Item (b) above thus says that any connected neighborhood of G generates the identity component of G. In fact, we can say more.

COROLLARY 11.13. *Let G be a Lie group and let G_0 be its identity component. Then G_0 is a normal subgroup of G, and is the only connected open subgroup. Every connected component of G is diffeomorphic to G_0.*

The proof is left as a homework exercise.

A good way to generate Lie groups is via the image and kernel of a Lie group homomorphism.

PROPOSITION 11.14. *Let G, H be Lie groups, and let $F\colon G \to H$ be a Lie group homomorphism. Then the kernel $\ker F = F^{-1}(e_H) \subseteq G$ is an embedded Lie subgroup whose codimension is the rank of F.*

PROOF. This follows immediately from Proposition 9.24 and Theorem 11.9: since F is a Lie group homomorphism it has constant rank r, and thus the e_H-level set $\ker F$ is an embedded submanifold of codimension r. The kernel of a group homomorphism is a group, and so $\ker F$ is an embedded Lie group. □

PROPOSITION 11.15. *Let $F\colon G \to H$ be an* injective *Lie group homomorphism. Then the image $F(G)$ has a unique smooth structure such that $F(G)$ is a (not necessarily embedded) Lie subgroup of H, and $F\colon G \to F(G)$ is a Lie group isomorphism.*

PROOF. Since F has constant rank and is injective, by the global rank Theorem 9.10 it is an injective immersion. Hence, by Theorem 9.18, F is a local smooth embedding, and this defines the topology and smooth structure on $F(G)$ since F is injective. It is clear this is the unique smooth structure for which F is a diffeomorphism onto its image. Since F is an immersion, this makes $F(G)$ into an immersed submanifold. The image of a group under a homomorphism is a group, and the group multiplication is inherited from H so is smooth, cf. Proposition 9.31. Hence $F(G)$ is a Lie group. Since F is a group homomorphism and a bijection $G \to F(G)$, it follows from Corollary 11.10 that it is a Lie group isomorphism. □

EXAMPLE 11.16. (1) With Proposition 11.14 in hand, we didn't need to work so hard to show SL_n is a Lie group. Note that $\mathrm{SL}_n(\mathbb{F}) = \ker(\det\colon \mathrm{GL}_n(\mathbb{F}) \to \mathbb{F}^*)$ with $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and since $\det$ is a Lie group homomorphism, it follows that $\mathrm{SL}_n(\mathbb{F})$ is an embedded Lie subgroup. Similarly, $\mathrm{SO}_n(\mathbb{F})$ and $\mathrm{SU}_n(\mathbb{C})$ are kernels of the restrictions of $\det$ to $\mathrm{O}_n(\mathbb{F}) \to \mathbb{F}^*$ and $\mathrm{U}_n(\mathbb{C}) \to \mathbb{C}^*$.

(2) Consider the map $\beta\colon \mathrm{GL}_n(\mathbb{C}) \to \mathrm{GL}_{2n}(\mathbb{R})$ defined by replacing each complex entry $a+ib$ with its representation as a 2×2 real matrx $\left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]$. It is straightforward to verify that β is an injective Lie group homomorphism, and thus we can identify $\mathrm{GL}_n(\mathbb{C})$ as a Lie sugroup of $\mathrm{GL}_{2n}(\mathbb{R})$. Moreover, it is also easy to check that the image is, in this case, an embedded submanifold (the 2×2 blocks gives us slice charts). Note: β arises nauturally from the identification of $(x^1 + iy^1, \dots, x^n + iy^n) \in \mathbb{C}^n$ with $(x^1, y_1, \dots, x^n, y^n) \in \mathbb{R}^{2n}$.

(3) Define $\alpha\colon \mathbb{R} \to \mathbb{T}^2$ by $\alpha(t) = (e^{2\pi i t}, e^{2\pi i a t})$ for some irrational a. Then α is a Lie group homomorphism and is injective (cf. Example 9.12), and so its image is an immersed Lie sugroup. But the image is dense, and so it is definitely not an embedded Lie subgroup. The *closure* of the image is all of $\mathbb{T}^2$, of course, which is an embedded Lie subgroup. In fact, we could have considered a boosting of this example: $\gamma\colon \mathbb{R} \to \mathbb{T}^2$ given by $\gamma(t) = (\alpha(t), 1)$; then the image is an immersed Lie sugroup whose closure is $\mathbb{T}^2 \subset \mathbb{T}^3$, an embedded Lie subgroup. This is the generic situation: if $S \subseteq G$ is any Lie subgroup, then its closure is an embedded Lie subgroup.

4. Smooth Group Actions

Let G be a group and let M be a set. A **left action of** G **on** M is a map $\theta\colon G \times M \to M$, usually written as $\theta(g, p) = \theta_g(p) = g \cdot p$, which satisfies the following two group laws:

$$\begin{aligned} g_1 \cdot (g_2 \cdot p) &= (g_1 g_2) \cdot p, && \text{for all } g_1, g_2 \in G \text{ and } p \in M, \text{ and} \\ e_G \cdot p &= p && \text{for all } p \in M. \end{aligned}$$

Written in terms of the notation θ_g, these say $\theta_{g_1} \circ \theta_{g_2} = \theta_{g_1 g_2}$ and $\theta_{e_G} = \mathrm{Id}_M$. We may also talk about **right actions**, where $\theta_{g_2} \circ \theta_{g_1} = \theta_{g_1 g_2}$, better understood in the notation $(p \cdot g_1) \cdot g_2 = p \cdot (g_1 g_2)$. Note, if θ is a left action, then $\phi(g, p) = \theta(g^{-1}, p)$ is a right action, and vice versa. Note also that θ_g is a bijection for each g, since by the group law it has inverse $\theta_{g^{-1}}$.

Now, let G be a Lie group and let M be a smooth manifold; we call an action $\theta\colon G \times M \to M$ a **smooth action** if the map θ is smooth. In this case, θ_g is a diffeomorphism for each $g \in G$.

Here is some standard notation for group actions. (We work here with left actions, but of course the comparable terminology applies to right actions.)

- For $p \in M$, the **orbit of** p is the set $G \cdot p = \{g \cdot p\colon g \in G\}$.

- For $p \in M$, the **stabilizer of** p, also called the **isotropy group of** p is the set $G_p = \{g \in G\colon g \cdot p = p\}$. That is, it is the set of group elements that fix p. Note that G_p is a subgroup.
- An action is said to be **transitive** if there is only one orbit, containing all elements of M. In other words, it is transitive if, for each pair $p, q \in M$, there is some $g \in G$ with $g \cdot p = q$.
- An action is said to be **free** if all stabilizers are trivial: $G_p = \{e_G\}$ for all p. In other words, only the group unit fixes any element.

Here are some basic and important examples of smooth actions of Lie groups.

EXAMPLE 11.17. (a) If G is any Lie group and M is any smooth manifold, the **trivial action** of G on M is defined by $g \cdot p = p$ for all $g \in G, p \in M$. It is smooth (θ is just the projection map $G \times M \to M$); each orbit is a single point, and the isotropy group of each point is all of G.

(b) Let $G \subseteq \mathrm{GL}(n, \mathbb{R})$ be any Lie subgroup of a general linear group. It acts on $\mathbb{R}^n$ (viewed as column vectors) by matrix multiplication: $A \cdot x = Ax \in \mathbb{R}^n$. It is a smooth action: its components are polynomials in the entries. Note that $A \cdot 0 = 0$ for all A, so the stabilizer of 0 is all of G, and the orbit of 0 is just $\{0\}$. If $G = \mathrm{GL}(n, \mathbb{R})$, then there is only one other orbit: for any two non-zero vectors x, y, there is some invertible matrix A with $Ax = y$. For subgroups, it can be more interesting. For example, if $G = \mathrm{O}(n, \mathbb{R})$, then there exists $Q \in G$ with $Qx = y$ if and only if $|x| = |y|$, and so the orbits of Q are spheres centered at 0. In general, stabilizers are intersections of G with affine spaces: the condition $Ax = x$ is the same as $(A - I)x = 0$, meaning that the matrix $A - I$ has rows in $x^\perp$.

(c) Every Lie group G acts smoothly on itself by left translation $L_g\colon G \to G$. Given $g_1, g_2 \in G$, there is a unique element $g \in G$ with $g \cdot g_1 = g_2$ (namely $g = g_2 g_1^{-1}$); this shows that the action is transitive and free. We may also consider the restriction of this to a Lie subgroup $L\colon H \times G \to G$ with $H \subseteq G$. The action is still free: $h \cdot g = g$ implies $h = e_H$; but it is not transitive unless $H = G$. (Indeed: if H acts transitively, then for any $g \in G$ there is some $h \in H$ for which $h \cdot e_G = g$; but $h \cdot e_G = h$, so $g \in H$.)

(d) Every Lie group acts smoothly on itself by conjugation $g \cdot h = ghg^{-1}$. (Written this way yields a left action.) The orbits of this action are called the *conjugacy classes* of G. The stabilizer of h is the commutator of h: $G_h = \{g \in G\colon gh = hg\}$.

(e) A sub-example of (d): let $G = \mathrm{U}(n)$ act on itself by conjugation. Since unitary matrices are normal, by the spectral theorem, every unitary matrix V can be unitarily diagonalized: $V = UDU^{-1}$ for some diagonal matrix D and $U in \mathrm{U}(n)$. Since $D = U^{-1}VU$ is a product of unitary matrices, it too is unitary, meaning $D^*D = I$. But since D is diagonal, this implies that all the diagonal entries d_{jj} satisfy $|d_{jj}|^2 = 1$. In other words, the diagonal entries are in the unit circle (i.e. the eigenvalues of a unitary matrix are in the unit circle). Hence, the conjugacy classes of $U(n)$ correspond to diagonal matrices with entries on the unit circle (a set which can be identified with the n-dimensional torus $\mathbb{T}^n$). Note, however, that these conjugacy classes are not all distinct: if $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ are two vectors in $\mathbb{T}^n$, and if they are permutations of each other, $\mu_j = \lambda_{\sigma(j)}$ for some $\sigma \in S_n$, then if $D_{\boldsymbol{\mu}}$ and $D_{\boldsymbol{\lambda}}$ are the associated diagonal matrices, we have $D_{\boldsymbol{\mu}} = S_\sigma D_{\boldsymbol{\lambda}} S_{\sigma^{-1}}$, where S_σ is the permutation matrix associates to σ (meaning that it has exactly one non-zero entry in each column, a 1, and in the jth column the 1 is in position $\sigma(j)$). Since $S_\sigma \in \mathrm{U}(n)$, this shows that $D_{\boldsymbol{\mu}}$ and $D_{\boldsymbol{\lambda}}$ are in the same conjugacy class. On the other hand, if two normal matrices are conjugate to each other by a unitary

matrix then they have the same eigenvalues, and so the conjugacy classes of $U(n)$ are in one-to-one correspondence with points in $\mathbb{T}^n$ whose arguments increase in $[0, 2\pi)$ in their components.

Group actions allow is to impart some nice properties of Lie groups to the manifolds they act on. This can be described through a property called **equivariance**.

DEFINITION 11.18. *Let M, N be smooth manifolds, and let $F\colon M \to N$ be a smooth map. Suppose that M, N both possess smooth (left) actions by some Lie group G. We call F* **equivariant** *under the actions of G if*

$$F(g \cdot p) = g \cdot F(p), \qquad \textit{for all } g \in G, p \in M.$$

This is often expressed as a commutative diagram:

$$\begin{tikzcd} M \arrow[r, "F"] \arrow[d, "g\cdot"'] & N \arrow[d, "g\cdot"] \\ M \arrow[r, "F"'] & N \end{tikzcd}$$

We equivalently say that F **intertwines** *the actions of G on M and N.*

EXAMPLE 11.19. Let $v = (v^1, \dots, v^n) \in \mathbb{R}^n$ be a fixed nonzero vector. Define actions of $\mathbb{R}$ on $\mathbb{R}^n$ and $\mathbb{T}^n$ as follows: for $t \in \mathbb{R}$,

$$\begin{aligned} t \cdot (x^1, \dots, x^n) &= (x^1 + tv^1, \dots, x^n + tv^n), \qquad (x^1, \dots, x^n) \in \mathbb{R}^n \\ t \cdot (z^1, \dots, z^n) &= (e^{2\pi i t v^1} z^1, \dots, e^{2\pi i t v^n} z^n), \qquad (z^1, \dots, z^n) \in \mathbb{T}^n. \end{aligned}$$

Let $\varepsilon^n\colon \mathbb{R}^n \to \mathbb{T}^n$ be the usual covering map $\varepsilon^n(x^1, \dots, x^n) = (e^{2\pi i x^1}, \dots, e^{2\pi i x^n})$. Then it is easy to verify that ε^n is equivariant under these actions.

EXAMPLE 11.20. Let $F\colon \mathrm{GL}(n, \mathbb{R}) \to \mathbb{M}_n(\mathbb{R})$ be the defining map of the orthogonal group: $F(A) = A^\top A$. Define (right) actions of $\mathrm{GL}(n, \mathbb{R})$ on the domain and codomain of F as follows: for the domain, $\mathrm{GL}(n, \mathbb{R})$ acts transitively by right multiplication on itself; for the codomain, define the right action $X \cdot B = B^\top X B$ for $B \in \mathrm{GL}(n, \mathbb{R})$ and $X \in \mathbb{M}_n(\mathbb{R})$. Then we have

$$F(A \cdot B) = F(AB) = (AB)^\top (AB) = B^\top A^\top AB = B^\top F(A) B = F(A) \cdot B$$

and so F is equivariant for these two actions.

We know (cf. Theorem 11.9) that all Lie group homomorphisms have constant rank. This property extends to the much wider class of equivariant maps (under transitive actions).

THEOREM 11.21 (Equivariant Rank Theorem). *Let $F\colon M \to N$ be a smooth map between manifolds. Let G be a Lie group that acts smoothly on both M and N, and suppose the action on M is transitive. If F is equivariant with respect to these actions, then F has constant rank.*

PROOF. Denote the action on M by θ and the action on N by ϕ. Let $p, q \in M$. By the transitivity assumption, there is some $g \in G$ with $\theta_g(p) = q$. The equivariance of F is the statement that $F \circ \theta_g = \phi_g \circ F$ for all g. We now apply the chain rule at the point p: $dF_q \circ (d\theta_g)_p = (d\phi_g)_{F(p)} \circ dF_p$. Since θ_g and ϕ_g are diffeomorphisms, the differentials $(d\theta_g)_p$ and $(d\phi_g)_{F(p)}$ are linear isomorphisms, and it follows that dF_p and dF_q have the same rank. $\square$

Of course, the same applies to right actions.

EXAMPLE 11.22. Returning to Example 11.20, the two smooth actions of $\mathrm{GL}(n,\mathbb{R})$ there were intertwined by $F(A) = A^\top A$, and since the domain action was transitive, it follows by the Equivariant Rank Theorem that F has constant rank. Hence, by Proposition 9.24, any level set $F^{-1}(X)$ of F is an embedded submanifold of $\mathrm{GL}(n,\mathbb{R})$. In particular, this gives a very short proof that $\mathrm{O}(n,\mathbb{R}) = F^{-1}(I)$ is an embedded submanifold (as we computed more directly above).

Generalizing the previous example, the Equivariant Rank Theorem is a powerful tool for finding embedded submanifolds.

PROPOSITION 11.23. *Let θ be a smooth action of a Lie group G on a manifold M. For each $p \in M$, the* **orbit map** $\theta^{(p)}\colon G \to M$ *is defined by $\theta^{(p)}(g) = \theta_g(p) = g \cdot p$. Note that $\theta^{(p)}(G) = G \cdot p$, while $(\theta^{(p)})^{-1}(p) = G_p$. For each $p \in M$, the orbit map has constant rank, and so the isotropy group G_p is an embedded Lie subgroup of G. If $G_p = \{e\}$, then $\theta^{(p)}$ is an injective smooth immersion, and so the orbit $G \cdot p$ is an immersed submanifold of M.*

REMARK 11.24. In fact, *every* orbit of a smooth action is an immersed submanifold, as we will see later. For now, we only have the tools to prove this when the associated isotropy group is trivial.

PROOF. The orbit map is smooth since the action θ is smooth. Now, G acts on itself by left multiplication, and by the definition of group action we have

$$\theta^{(p)}(g \cdot g_0) = (g \cdot g_0) \cdot p = g \cdot (g_0 \cdot p) = g \cdot \theta^{(p)}(g_0)$$

showing that $\theta^{(p)}$ is equivariant with tespect to the actions of G on G and M. Moreover, the left multiplication action of G on itself is transitive. Thus, by the Equivariant Rank Theorem, $\theta^{(p)}$ has constant rank. It now follows by Proposition 9.24 that the level set $G_p = (\theta^{(p)})^{-1}(p)$ is an embedded submanifold of G; since it is also a subgroup, it is therefore a Lie subgroup.

Now, suppose $G_p = \{e\}$. Then $\theta^{(p)}$ is injective: if $\theta^{(p)}(g) = \theta^{(p)}(g')$ then $g \cdot p = g' \cdot p$ which implies, by the group law, that $g^{-1}g' \cdot p = p$, and so $g^{-1}g' \in G_p = \{e\}$, so $g = g'$. By the Equivariant Rank Theorem and the Global Rank Theorem 9.10, it now follows that $\theta^{(p)}$ is an injective immersion, and so its image $G \cdot p$ is, by definition, an immersed submanifold of M. $\square$

5. Semidirect Products

Smooth actions give us tools to construct many new Lie groups from old ones. Let H, N be Lie groups, and suppose $\theta\colon H \times N \to N$ is a smooth action. Then we know that, for each $h \in H$, $\theta_h\colon N \to N$ is a diffeomorphism. If θ_h is also a group homomorphism, we say that H **acts by automorphisms on** N.

DEFINITION 11.25. *Let H, N be Lie groups $\theta\colon H \times N \to N$ be a smooth action by automorphisms. The* **semidirect product** *of N, H by θ, denoted $N \rtimes_\theta H$ (or simply $N \rtimes H$ if θ is understood from context) is the smooth manifold $N \times H$ given group operation*

$$(n_1, h_1)(n_2, h_2) \equiv (n_1\theta_{h_1}(n_2), h_1h_1).$$

It is immediate to verify that this operation defines a group, with identity element (e, e), and inverse $(n, h)^{-1} = (\theta_{h^{-1}}(n^{-1}), h^{-1})$. Moreover, since θ is smooth and N, H are Lie groups, it is also smooth. So $N \rtimes H$ is a Lie group.

EXAMPLE 11.26. Recall the Euclidean group $E(n,\mathbb{R})$ of Example 11.5(8). Let $\mathrm{O}(n,\mathbb{R})$ act on $\mathbb{R}^n$ by the natural action $\mathrm{O}(n,\mathbb{R}) \times \mathbb{R}^n \to \mathbb{R}^n : (Q,x) \mapsto Qx$ viewing x as a column vector. We view $\mathbb{R}^n$ as an additive Lie group, and so $\mathrm{O}(n,\mathbb{R})$ acts by automorphisms (i.e. linear isomorphisms). In fact, the group $E(n,\mathbb{R})$ was defined precisely as $\mathbb{R}^n \rtimes \mathrm{O}(n,\mathbb{R})$. Letting $T_y(v) = v+y$, for any $Q \in \mathrm{O}(n,\mathbb{R})$ we have $QT_y(v) = Q(v+y) = Qv + Qy = T_{Qy}Qv$, and so

$$(x,Q)(y,R) \equiv T_xQT_yR = T_xT_{Qy}QR = T_{x+Qy}QR \equiv (x+Qy, QR).$$

Here are some basic properties of semidirect products.

PROPOSITION 11.27. *Let N, H be Lie groups, let $\theta\colon H \times N \to N$ be a smooth action by automorphisms, and let $G = N \rtimes_\theta H$. Let $\widetilde{N} = N \times \{e\}$ and $\widetilde{H} = \{e\} \times H$.*

(a) *The slices $\widetilde{N}, \widetilde{H}$ are Lie subgroups of G isomorphic to N, H respectively.*
(b) *$\widetilde{N}$ is a normal subgroup of G.*
(c) *$\widetilde{N} \cap \widetilde{H} = \{(e,e)\}$, and $\widetilde{N}\widetilde{H} = G$.*

PROOF. By definition of the group operation, $(n_1,e)(n_2,e) = (n_1\theta_e(n_2), ee) = (n_1n_2, e)$ since θ is an action so $\theta_e(n) = n$. Similarly, $(e,h_1)(e,h_2) = (e\theta_{h_2}(e), h_1h_2) = (ee, h_1h_2) = (e, h_1h_2)$ since θ acts by automorphisms so $\theta_{h_2}(e) = e$. The slices are diffeomorphism to N, H as manifolds, and so they are therefore also isomorphic as groups, proving item (a). For item (b), fix an element $(n_0,h_0) \in G$; then for any $(n,e) \in \widetilde{N}$, we have

$$\begin{aligned}(n_0,h_0)(n,e)(n_0,h_0)^{-1} &= (n_0\theta_{h_0}(n), h_0e)(n_0,h_0)^{-1}\\ &= (n_0\theta_{h_0}(n), h_0)(\theta_{h_0^{-1}}(n_0^{-1}), h_0^{-1})\\ &= (n_0\theta_{h_0}(n)\cdot\theta_{h_0}(\theta_{h_0^{-1}}(n_0^{-1})), h_0h_0^{-1})\\ &= (n_0\theta_{h_0}(n)n_0^{-1}, e).\end{aligned}$$

Since θ is an action of H on N, the first term is in $\widetilde{N}$, and this shows $(n_0,h_0)\widetilde{N}(n_0,h_0)^{-1} \subseteq \widetilde{N}$, proving the claim. Finally, for item (c), the first statement is immediate from the definition. For the secondd, fix $(n,h) \in G$; then we have

$$(n,h) = (n\theta_e(e), eh) = (n,e)(e,h)$$

as claimed. □

The decomposition of Proposition 11.27 is its characterizing property, as the following (essentially purely algebraic) proposition demonstrates.

PROPOSITION 11.28. *Let G be a Lie group, and let $N, H \subseteq G$ be embedded Lie subgroups such that N is normal, $N \cap H = \{e\}$, and $NH = G$. Then the product map $\mathrm{m}\colon (n,h) \mapsto nh$ is a Lie group isomorphism $N \rtimes_\theta H \to G$, where $\theta\colon H \times N \to N$ is the conjugation action $\theta_h(n) = hnh^{-1}$.*

PROOF. The action θ is smooth, so $N \rtimes_\theta H$ is a Lie group, as is G. Thus, we need only show that the map m is a bijective homomorphism. It is onto because $NH = G$. To see it is a homomorphism, we note that $\mathrm{m}(e,e) = e$, and we compute

$$\mathrm{m}\left((n_1,h_1)\cdot(n_2,h_2)\right) = \mathrm{m}(n_1\theta_{h_1}(n_2), h_1h_2) = \mathrm{m}(n_1h_1n_2h_1^{-1}, h_1h_2) = n_1h_1n_2h_1^{-1}h_1h_2$$

and this is equal to $n_1h_1n_2h_2 = \mathrm{m}(n_1,h_1)\cdot\mathrm{m}(n_2,h_2)$, as claimed. Finally, let $(n_0,h_0) \in \ker(\mathrm{m})$; then $n_0h_0 = e$. Thus $h_0^{-1} = n_0$, which shows $h_0^{-1} \in N$; but N is a subgroup, and so $h_0 \in N$. Thus

$h_0 \in H \cap N = \{e\}$, so $h_0 = e$. Then $n_0 e = e$ so $n_0 = e$, and so $(n_0, h_0) = (e, e)$. Thus m has trivial kernel. Thus it is a group isomorphism, and hence a Lie group isomorphism. $\square$

CHAPTER 12

Invariant Vector Fields and Measures

1. Left-Invariant Vector Fields

Let G be a Lie group. Unlike a general manifold, G has a special point: e. Now, let $X_e \in T_eG$ be a vector tangent to the identity. Then the group operations allow us to translate X_e around to other points: using the diffeomorphism $L_g \colon G \to G$ for $g \in G$, we have $(dL_g)_e \colon T_eG \to T_gG$ is a linear isomorphism. So left multiplication gives us a natural way to identify all tangent spaces with each other. Tanginbly, we transform X_e to the vector X_g defined by

$$X_g = (dL_g)_e(X_e).$$

So any vector at e can be translated around to any other point. In fact, this defines a vector field, which most authors call $\widetilde{X}$. This is actually a smooth vector field.

PROPOSITION 12.1. *Let G be a Lie group, and let X_e be any vector in T_eG, and define the (rough) vector field $\widetilde{X}$ on G by $\widetilde{X}_g = (dL_g)_e(X_e)$. Then $\widetilde{X} \in \mathscr{X}(G)$.*

PROOF. Let $f \in C^\infty(G)$. Let $\alpha \colon (-\epsilon, \epsilon) \to G$ be a smooth curve so that $\alpha(0) = e$ and $\dot{\alpha}(0) = X_e$. Then by definition, for any $g \in G$,

$$\left(\widetilde{X}(f)\right)(g) = \widetilde{X}_g(f) = (dL_g)_e(X_e)f = X_e(f \circ L_g) = \left.\frac{d}{dt}\right|_{t=0} (f \circ L_g \circ \alpha)(t).$$

That is

$$\left(\widetilde{X}(f)\right)(g) = \left.\frac{d}{dt}\right|_{t=0} f(g\alpha(t)).$$

The function $(g, t) \mapsto g\alpha(t)$ is a smooth map $G \times (-\epsilon, \epsilon) \to G$, and so its partial derivatives are smooth. This concludes the proof. □

Thus, we have a map $T_eG \to \mathscr{X}(G)$, given by $X_e \mapsto \widetilde{X}$. It is elementary to verify that it is an $\mathbb{R}$-linear map. Note that $\widetilde{X}_e = X_e$, and so this map is also clearly injective. It is not, however, surjective: the space $\mathscr{X}(G)$ is infinite-dimensional (if $\dim G \geq 1$), while T_eG is finite-dimensional. It is easy to see why this map cannot be surjective: the vector field $\widetilde{X}$ is *left-invariant*.

DEFINITION 12.2. *Let G be a Lie group. A smooth vector field $X \in \mathscr{X}(G)$ is called* **left-invariant** *if $(L_g)_*(X) = X$ for all $g \in G$. That is: for all $g, g' \in G$,*

$$X_{g'} = (L_g)_*(X)_{g'} = d(L_g)_{(L_g)^{-1}(g')}(X_{(L_g)^{-1}(g')}).$$

As $(L_g)^{-1}(g') = g^{-1}g'$, taking $h = g^{-1}g'$, it is more convenient to write left-invariance as the condition

$$X_{gh} = d(L_g)_h(X_h).$$

In particular, if X is left-invariant, then $X_g = d(L_g)_e(X_e) = \widetilde{X}_g$, and so the vector field $\widetilde{X}$ determined by left-translating X_e around is left-invariant.

Denote by $\mathscr{X}^L(G)$ the set of left-invariant vector fields on G. Given $a, b \in \mathbb{R}$, note that $(L_g)_*(aX + bY) = a(L_g)_*(X) + b(L_g)_*(Y)$, it follows that $\mathscr{X}^L(G)$ is an $\mathbb{R}$-linear subspace of $\mathscr{X}^L(G)$. In fact, this subspace is finite-dimensional: it is the image of the map $T_eG \to \mathscr{X}(G)$ in Proposition 12.1.

COROLLARY 12.3. *The map $\lambda\colon T_eG \to \mathscr{X}^L(G)$ given by $X_e \mapsto \widetilde{X}$ is an $\mathbb{R}$-linear isomorphism.*

PROOF. We noted above that $\widetilde{X}$ is left-invariant, so λ is well-defined. We showed above that λ is $\mathbb{R}$-linear and injective, so it only remains to show that it is surjective. Thus, let $X \in \mathscr{X}^L(G)$. Let $V = \widetilde{X} = \lambda(X_e)$, so $V_g = d(L_g)_e(X_e)$. Since X is left-invariant, we have $X_g = d(L_g)_e(X_e) = V_g = \lambda(X_e)_g$, i.e. $X \in \lambda(T_eG)$. This concludes the proof. □

Hence, $\mathscr{X}^L(G)$ is finite-dimensional: it is isomorphic, as a vector space, to T_eG. One immediate consequence is that Lie groups have global frames.

PROPOSITION 12.4. *Any Lie group possesses a left-invariant global smooth frame; hence, it is parallelizable, and hence orientable.*

PROOF. Let G be a Lie group, and fix any basis $\{E^1, \dots, E^n\}$ of T_eG, and let $\{\widetilde{E}^1_g, \dots, \widetilde{E}^n\}$ be the left-invariant extensions of them (i.e. $\widetilde{E}^j = \lambda(E^j)$). We claim these form a smooth frame. They are smooth vector fields by Proposition 12.1, so we need only show they are linearly independent at each point. Let $g \in G$, and suppose $c_1\widetilde{E}^1_g + \cdots + c_n\widetilde{E}^n_g = 0$ for some $c_1, \dots, c_n \in \mathbb{R}$. Then using the isomorphism λ of Corollary 12.3, we have

$$0 = c_1\lambda(E^1) + \cdots + c_n\lambda(E^n) = \lambda(c_1E^1 + \cdots + c_nE^n)$$

and since λ is injective, it follows that $c_1E^1 + \cdots + c_nE^n = 0$. But $\{E^1, \dots, E^n\}$ are a basis for T_eG, and so $c_1 = \cdots = c_n = 0$. Hence, these smooth vector fields form a global smooth frame of left-invariant vector fields. It follows (by definition) that G is parallelizable. Hence, from Example 10.7, it follows that G is orientable. □

Now, let G be a Lie group and fix any basis $\{E^1, \dots, E^n\}$ of T_eG; let $\{\widetilde{E}^1, \dots, \widetilde{E}^n\}$ be the corresponding left-invariant frame. For each $g \in G$, let $\{\omega_1|_g, \dots, \omega_n|_g\}$ be the dual basis of T^*_gG. This defines n (rough) covariant vector fields $\omega_1, \dots, \omega_n$. Since E^j are left-invariant, we expect ω_j to also behave well under the group action, and they do: they are preserved under *pull back* by L_g. We have

$$(L^*_g\omega_j)_h(\widetilde{E}^i_h) = (\omega_j)_{L_gh}(d(L_g)_h(\widetilde{E}^i_h)).$$

As $\widetilde{E}^i$ is left-invariant, $d(L_g)_h(\widetilde{E}^i_h) = \widetilde{E}^i_{gh}$, and so

$$(L^*_g\omega_j)_h(\widetilde{E}^i_h) = (\omega_j)_{gh}(\widetilde{E}^i_{gh}) = \delta^i_j = (\omega_j)_h(\widetilde{E}^i_h).$$

That is: $L^*_g\omega_j = \omega_j$. This motivates a definition/lemma.

LEMMA 12.5. *Let ω be a (rough) covariant vector field on a Lie group G. Call ω* **left-invariant** *if $L^*_g\omega = \omega$ for each $g \in G$. Any left invariant covariant vector field is a 1-form, $\omega \in \Omega^1(G)$; i.e. it is smooth.*

PROOF. Let Fix a left-invariant frame $\widetilde{E}^1, \dots, \widetilde{E}^n$ on G. Then we may expand any $X \in \mathscr{X}(G)$ as $X = \sum_{j=1}^n f_j \widetilde{E}^j$ for some $f_j \in C^\infty(G)$. Thus, for any covariant vector field ω, we have

$$\omega(X)(g) = \sum_{j=1}^n f_j(g)\omega(\widetilde{E}^j)(g).$$

Now, if ω is left-invariant, since $\widetilde{E}^j$ is left-invariant, we have

$$\omega(\widetilde{E}^j)(g) = \omega_g(\widetilde{E}^j_g) = \omega_g(d(L_g)_e(\widetilde{E}^j_e) = (L_g^*\omega)_e(\widetilde{E}^j_e) = \omega_e(\widetilde{E}^j_e).$$

That is: these functions are constant. S owe have

$$\omega(X) = \sum_{j=1}^n \omega_e(\widetilde{E}^j_e) f_j \in C^\infty(G).$$

Thus ω is smooth. □

Denote the subspace of left-invariant 1-forms as $\Omega^1_L(G)$. So, if $\{E^1, \dots, E^1\}$ is a basis for T_eG and $\{\widetilde{E}^1, \dots, \widetilde{E}^n\}$ is the corresponding left-invariant frame, then the dual frame $\{\omega_1, \dots, \omega_n\}$ is in $\Omega^1_L(G)^n$. This allows us to construct a left-invariant orientation form, which is (up to scale) unique. (Note: we can extend the definition of left-invariance to k-forms by the same formula $L_g^*\omega = \omega$ for all $g \in G$. We denote the space of such forms as $\Omega^k_L(G)$. Lemma 12.5 applies equally well to covariant tensor fields, with virtually the same proof.)

PROPOSITION 12.6. *Let G be a Lie group. Then there is a left-invariant orientation form $\vartheta \in \Omega^n_L(G)$, which is unique up to scale. If G is compact, there is a unique left-invariant orientation form ϑ_G with $\int_G \vartheta_G = 1$.*

PROOF. Let $\{E^1, \dots, E^n\}$ be a basis of T_eG, let $\{\widetilde{E}^1, \dots, \widetilde{E}^n\}$ be the associated left-invariant frame, and let $\{\omega_1, \dots, \omega_n\}$ be the associated left-invariant dual frame as above, and set $\vartheta = \omega_1 \wedge \cdots \wedge \omega_n$. By (induction on) Lemma 8.11, we have

$$L_g^*\vartheta = L_g^*(\omega_1 \wedge \cdots \wedge \omega_n) = L_g^*\omega_1 \wedge \cdots \wedge L_g^*\omega_n = \omega_1 \wedge \cdots \wedge \omega_n = \vartheta$$

so ϑ is left-invariant. By definition $\vartheta(\widetilde{E}^1, \dots, \widetilde{E}^n)(g) = 1$ for all $g \in G$, which shows that ϑ does not vanish at any g, so it is an orientation form.

Now, suppose η is another left-invariant orientation form on G. Since $\dim \Lambda^n(T_e^*G) = 1$, there is a constant $c \in \mathbb{R}^*$ with $\eta_e = c\vartheta_e$. Now using left-invariance, for any $g \in G$ we have

$$\eta_g = L_{g^{-1}}^*\eta_e = cL_{g^{-1}}^*\vartheta_e = c\vartheta_g.$$

Thus, ϑ is indeed unique up to scale.

Finally, fix any left-invariant orientation form ϑ on G. We use it to give G an orientation. If G is compact, then ϑ is compactly-supported, and so $\int_G \vartheta$ is a positive real number (by Proposition 10.15(c)). Hence, we may define

$$\vartheta_G = \frac{1}{\int_G \vartheta}\vartheta.$$

Then $\int_G \vartheta_G = 1$, and since any two such forms are related by a scalar multiple, it follows that ϑ_G is unique. □

The left-invariant orienation forms in Proposition 12.6 are called **Haar forms**. It is more common to use them to define **Haar measure**. At least for $f \in C^\infty(G)$, we can define integration against Haar measure as follows:

$$\int_G f(g)\, dg \equiv \int_G f\, \vartheta_G.$$

(This is uniquely defined if G is compact; otherwise it requires a choice of scale for the Haar form.) The point if this is as follows: thinking of f as a 0-form, for any $h \in G$ we have

$$(L_h^* f)\vartheta_G = (L_h^* f)(L_h^* \vartheta_G) = L_h^*(f\vartheta_G)$$

and so

$$\int_G (L_h^* f)(g)\, dg = \int_G L_h^*(f\vartheta_G) = \int_G f\vartheta_G = \int_G f(g)\, dg$$

where the second equality follows by Proposition 10.15(d), since L_g is a diffeomorphism which is orientation preserving (by left-invariance). Note that $(L_h^* f)(g) = f(hg)$ – i.e. it is the left-translation of the argument of f. Thus, Haar measure is **translation invariant**. For example, on $\mathbb{R}^n$, the Lebesgue measure(s up to scale) are the Haar measures. That every Lie group posseses a Haar measure is an important technical tool we will use later. (Note: this holds in greater generality: on any locally-compact topological group, there is a unique-up-to-scale Haar measure. But this is much harder to prove without the tools of differential geometry.)

2. Smooth Measures, Density Fields, and Haar Measure

In the preceding section, we noted how to construct left- (and right-)translation invariant measures, known has **Haar measures**, on Lie groups, using integration of volume forms. This is really the only application of differential forms (of rank ≥ 2) in basic Lie theory. Since the development of differential forms, orientation, and integration of forms on manifolds is a time-consuming subject, we also include a self-contained section here that constructs Haar measure in a more direct way, via *smooth measures* and *density fields* on manifolds.

2.1. Smooth Measures on Manifolds. For motivation, we begin with the change-of-variables formula for Lebesgue measure in local charts.

DEFINITION 12.7. *Let (U, x) be a chart on a smooth d-dimensional manifold M. Then $x(U) \subset \mathbb{R}^d$ possesses a Lebesgue measure m_x (whose scale is determined by x). Let dx denote the push-forward of this measure to U:*

$$dx = x_*\mu = \mathrm{m}_x \circ x.$$

(This is a measure because x is a diffeomorphism; i.e. this is really $\mathrm{m}_x \circ (x^{-1})^{-1}$.) Note that dx is a Borel measure on U.

LEMMA 12.8. *Let (U, x) and (V, y) be charts on a smooth manifold M. Then on $U \cap V$,*

$$dy = \left|\det[D(y \circ x^{-1})]\right| \circ x\, dx.$$

We therefore use the notation

$$\left|\frac{\partial y}{\partial x}\right| := \left|\det[D(y \circ x^{-1})]\right| \circ x$$

for the Jacobian determinant. Thus $|\partial y/\partial x|$ is the smooth function on $U \cap V$ with the property that

$$dy = \left|\frac{\partial y}{\partial x}\right| dx.$$

PROOF. To avoid confusion, let s denote a point in $x(U \cap V)$ and let $t = y \circ x^{-1}(s)$ denote the corresponding point in $y(U \cap V)$. Then by the measure-theoretic change-of-variables theorem, for a (say) bounded continuous function f on $U \cap V$,

$$\int_{U \cap V} f\, dy = \int_{U \cap V} (f \circ y^{-1}) \circ y\, d(\mathrm{m}_y \circ y) = \int_{y(U \cap V)} f \circ y^{-1}(t)\, dt.$$

We now apply the change-of-variables formula from multivariate calculus, with the diffeomorphism $x \circ y^{-1} \colon U \cap V \to U \cap V$, to convert this to

$$\int_{y(U \cap V)} f \circ y^{-1}(t)\, dt = \int_{x(U \cap V)} f \circ x^{-1}(s) \left|\det[D(y \circ x^{-1})](s)\right|\, ds.$$

Meanwhile, by definition

$$\int_{U \cap V} f\, \left|\det[D(y \circ x^{-1})]\right| \circ x\, dx = \int_{U \cap V} (f \circ x^{-1}) \circ x\, \left|\det[D(y \circ x^{-1})]\right| \circ x\, d(\mathrm{m}_x \circ x)$$

which precisely equals the above expression, again by the measure theoretic change of variables theorem. □

The presence of the Jacobian determinant of the transition function should be no surprise. It motivates what kind of invariance we need to build into a global object which we can only access in local form. Such is life for an "smooth measure" on a manifold.

DEFINITION 12.9. *A Radon measure μ on a smooth manifold M is called* **absolutely continuous** *if, for each chart (U, x) in an atlas for M, $d\mu|_U \ll dx$. If, in addition, $d\mu|_U / dx \in C^\infty(U)$ for each chart, we call μ a* **smooth measure** *on M.*

In order to construct absolutely continuous (and in particular smooth) measures, they key is the invariance property of their densities.

LEMMA 12.10. *Let μ be a smooth measure on a smooth manifold M. In any chart (U, x), denote the Radon–Nikodym derivative as*

$$\frac{d\mu|_U}{dx} =: \rho^x \colon U \to \mathbb{R}_+.$$

If (V, y) is another chart, then on $U \cap V$,

$$\rho^x = \rho^y \left|\frac{\partial y}{\partial x}\right|.$$

PROOF. By definition, we have

$$\rho^x\, dx = d\mu = \rho^y\, dy \quad \text{on } U \cap V.$$

By Lemma 12.8, we therefore have

$$\rho^x\, dx = \rho^y \left|\frac{\partial y}{\partial x}\right|\, dx.$$

Since these two measures are equal on $U \cap V$, it follows that $\rho^x = \rho^y \left|\frac{\partial y}{\partial x}\right|$ a.e. with respect to dx. The images of ρ^x and $\rho^y \left|\frac{\partial y}{\partial x}\right|$ under x are smooth, hence continuous, and they are equal a.e. with respect to Lebesgue measure. It follows that they are equal. □

(The above proof only required the Radon–Nikodym derivatives to be continuous in all charts; for our purposes, continuous measures that are not smooth will not be important.)

In fact, the transition-invariance property of the local densities ρ^x of a smooth measure, cf. Lemma 12.10, is all that is needed to construct such measures.

THEOREM 12.11. *Let M be a smooth manifold. Suppose that, for each chart (U, x) in an atlas for M, there is a smooth function $\rho^x \colon U \to \mathbb{R}_+$, and suppose that for any pair of charts,*

$$\rho^x = \rho^y \left| \frac{\partial y}{\partial x} \right|.$$

Then there is a unique smooth measure μ on M such that $d\mu|_U = \rho^x \, dx$ on each chart (U, x).

To prove this, we use the following elementary measure theory lemma.

LEMMA 12.12. *Let $(M, \mathscr{B})$ be a measurable space, and let $\mathscr{A} \subseteq \mathscr{B}$ be a collection of measurable sets containing a countable subset $\{A_1, A_2, \ldots\} \subseteq \mathscr{A}$ with $M = \bigcup_j A_j$. For each $A \in \mathscr{A}$, let μ_A be a measure on A (with respect to the σ-field $\mathscr{B}_A := \{A \cap B \colon B \in \mathscr{B}\}$). If $\mu_A = \mu_{A'}$ on $A \cap A'$ for all $A, A' \in \mathscr{A}$, then there is a unique measure μ on $(M, \mathscr{B})$ such that $\mu|_A = \mu_A$ for all $A \in \mathscr{A}$.*

PROOF. We redistribute the A_j so they are disjoint: let $A'_1 = A_1$ and $A'_j := A_j \setminus (A_1 \cup \cdots A_{j-1})$ for $j \geq 2$. Then $\bigcup_{j=1}^n A_j = \bigcup_{j=1}^n A'_j$ for each $n \in \mathbb{N}$, and so $\bigcup_j A'_j = M$. By construction $\{A'_j\}$ are disjoint, and $A'_j \subseteq A_j$.

Now, let us first address uniqueness. If μ is such a measure on M, then for any $B \in \mathscr{B}$, since $B = \bigsqcup_j B \cap A'_j$ (disjoint union), we have

$$\mu(B) = \sum_j \mu(B \cap A'_j) = \sum_j \mu_{A_j}(B \cap A'_j) \tag{12.1}$$

where the second equality comes from the defining property of μ and the fact that $A'_j \subseteq A_j$. Note that (12.1) completely defines μ in terms of the given measures μ_j, demonstrating uniqueness.

For existence, we simply *define* μ by (12.1). By the disjointness of the A'_j it is easily seen that this defines a measure; we just need to check that $\mu|_A = \mu_A$ for each $A \in \mathscr{A}$. Fixing A, let $B \in \mathscr{B}_A$. Then

$$\mu(B) = \sum_j \mu_{A_j}(B \cap A'_j).$$

Since $B \subseteq A$ and $A'_j \subseteq A_j$, $B \cap A'_j \subseteq A \cap A_j$. By assumption, it therefore follows that $\mu_A(B \cap A'_j) = \mu_{A_j}(B \cap A'_j)$. Thus

$$\mu(B) = \sum_j \mu_{A_j}(B \cap A'_j) = \sum_k \mu_A(B \cap A'_j) = \mu_A\left(\bigsqcup_j B \cap A'_j\right) = \mu_A(B)$$

as desired. □

PROOF OF THEOREM 12.11. Let $\mathscr{A}$ be the set of all domains of charts in an atlas for M. By the second countability of M, there is a countable subset of $\mathscr{A}$ that covers M (cf. Proposition 1.20). On each chart (U, x), define a measure μ^x on U by $d\mu^x = \rho^x \, dx$. Then for any two charts (U, x) and (V, y),

$$d\mu^x = \rho^x \, dx = \rho^y \left| \frac{\partial y}{\partial x} \right| dx = \rho^y \, dy = d\mu^y$$

where the penultimate equality follows from Lemma 12.8. It now follows from Lemma 12.12 that there is a unique measure μ on M with $d\mu|_U = d\mu^x = \rho^x\,dx$ for each chart (U, x), as claimed. Moreover, since ρ^x is smooth by assumption, it follows by definition that μ is a smooth measure. □

2.2. Smooth Density Fields on Manifolds. We now explore how to construct a global object on a manifold which gives rise to the local functions ρ^x in Theorem 12.11 defining a smooth measure. The idea is to rely, once again, on the change-of-variables theorem from multivariate calculus, in infinitesimal form: we define a "density" on each tangent space, glued together to form a global object. The requisite notion from multilinear algebra is as follows.

DEFINITION 12.13. *Let V be a real d-dimensional vector space. A* **density** *on V is a function $\delta\colon V^d \to \mathbb{R}_+$ with the property that*

$$\delta(Av_1, \dots, Av_d) = |\det A|\delta(v_1, \dots, v_d)$$

for all $v_1, \dots, v_d \in V$ and $A \in \mathrm{End}(V)$.

REMARK 12.14. Without the absolute value around the determinant, this would mirror the transformation property of a d-form on V. In fact, such a density δ is always the absolute value of a d-form; as such, there is a unique density up to scale, as we now show.

LEMMA 12.15. *Let V be a d-dimensional real vector space, and fix any basis $\mathbf{e} = \{e_1, \dots, e_d\}$. Let $\{e_1^*, \dots, e_d^*\}$ denote the dual basis of V^*. Then*

$$\delta^{\mathbf{e}}(v_1, \dots, v_d) := \left|\det\left(\left[e_i^*(v_j)\right]_{i,j=1}^d\right)\right| \tag{12.2}$$

defines a non-zero density on V. Moreover, if δ is any density on V, then $\delta = c \cdot \delta^{\mathbf{e}}$ where c is the constant $c = \delta(e_1, \dots, e_d)$.

PROOF. Expanding $v_j = \sum_k e_k^*(v_j)e_k$, we have

$$e_i^*(Av_j) = \sum_k e_k^*(v_j)e_i^*(Ae_k).$$

Letting $S_{ik} = e_i^*(Ae_k)$ and $T_{kj} = e_k^*(v_j)$, we therefore have $e_i^*(Av_j) = [ST]_{ij}$. Therefore

$$\delta^{\mathbf{e}}(Av_1, \dots, Av_d) = \left|\det\left(\left[e_i^*(Av_j)\right]_{i,j=1}^d\right)\right| = |\det[ST]| = |\det S||\det T|.$$

By definition, $|\det T| = \delta^{\mathbf{e}}(v_1, \dots, v_d)$. Also, S_{ik} are the components of the matrix of A in the basis $\{e_1, \dots, e_d\}$; therefore $\det S = \det A$. This demonstrates that $\delta^{\mathbf{e}}$ is a density on V. Note, in particular, that $\delta^{\mathbf{e}}(e_1, \dots, e_d) = 1$ (since $[e_i^*(e_j)]_{i,j=1}^d$ is the identity matrix); thus $\delta^{\mathbf{e}}(Ae_1, \dots, Ae_d) = |\det A|$ for any $A \in \mathrm{End}(V)$.

Now, let δ be any density on V. Then

$$\delta(Ae_1, \dots, Ae_d) = |\det A|r(e_1, \dots, e_d) = \delta^{\mathbf{e}}(Ae_1, \dots, Ae_d)r(e_1, \dots, e_d).$$

Since $\{e_1, \dots, e_d\}$ form a basis for V, given any vectors $\{v_1, \dots, v_d\}$ in V there is some $A \in \mathrm{End}(V)$ such that $Ae_j = v_j$ for $1 \le j \le d$. Thus, we have shown that, for all such d-tuples of vectors,

$$\delta(v_1, \dots, v_d) = \delta(e_1, \dots, e_d)\delta^{\mathbf{e}}(v_1, \dots, v_d)$$

as desired. □

REMARK 12.16. By scaling only one basis vector, the change of basis matrix is diagonal with one non-1 entry, having determinant given by that non-1 constant. It therefore follows, by the uniqueness up to scaling above, that every density has the form (12.2) for some basis of V. (A further calculation will show that, fixing an inner product on V, the value of the density is the same for any orthonormal basis; so densities on V are in one-to-one correspondence with inner products on V.)

We now extend this definition to a global object, gluing together densities at each point.

DEFINITION 12.17. *A* **smooth density field** *on a smooth d-dimensional manifold M is a function $\delta\colon \bigsqcup_{p\in M}(T_pM)^d \to \mathbb{R}_+$ with the following properties:*

(1) *For each $p \in M$, $\delta_p\colon (T_pM)^d \to \mathbb{R}_+$ is a density on T_pM.*
(2) *If $U \subseteq M$ is open and $X_1, \dots, X_d \in \mathscr{X}(U)$ is a smooth local frame of vector fields on U, then the function $p \mapsto \delta_p(X_1(p), \dots, X_d(p))$ is smooth.*

REMARK 12.18. We could construct δ more abstractly as a section of a certain line bundle over M; then the smoothness condition would be shown to be equivalent to smoothness as a section, and therefore also equivalent to the nominally weaker statement that $p \mapsto \delta_p(X_1(p), \dots, X_d(p))$ is smooth for all $X_1, \dots, X_d \in \mathscr{X}(M)$.

We could now prove the existence of a multitude of smooth density fields by using a standard partition of unity argument; in the case we care about, construction of a smooth density field will be straightforward. The more relevant point for us now is that a smooth density field gives rise to the local density functions of a smooth measure, as follows.

PROPOSITION 12.19. *Let δ be a smooth density field on a smooth manifold M. For any chart (U, x) on M, define the function $\rho^x\colon U \to \mathbb{R}_+$ by*

$$\rho^x(p) = \delta_p\left(\left.\frac{\partial}{\partial x^1}\right|_p, \dots, \left.\frac{\partial}{\partial x^d}\right|_p\right).$$

Then for any charts (U, x) and (V, y), we have $\rho^x = \rho^y \left|\frac{\partial y}{\partial x}\right|$ on $U \cap V$. Hence, by Theorem 12.11, δ gives rise to a unique smooth measure μ_δ on M.

PROOF. This follows from the change of variables for coordinate vector fields, (3.5), which states that for $p \in U \cap V$

$$\left.\frac{\partial}{\partial x^j}\right|_p = \sum_{k=1}^d \frac{\partial y^k}{\partial x^j}(x(p)) \left.\frac{\partial}{\partial y^k}\right|_p = \sum_{k=1}^d \left[[D(y \circ x^{-1})](x(p))\right]_{jk} \left.\frac{\partial}{\partial y^k}\right|_p. \tag{12.3}$$

Hence, if A is the linear transformation of T_pM defined so that $A(\frac{\partial}{\partial y^j}|_p) = \frac{\partial}{\partial x^j}|_p$ for $j = 1, \dots, d$, (12.3) says that the matrix of A in the basis $\{\frac{\partial}{\partial y^k}|_p\}_{k=1}^d$ is the transpose $[D(y \circ x^{-1})](x(p))^\top$. Therefore $|\det A| = |\det A^\top| = |\det [D(y \circ x^{-1})](x(p))| = |\frac{\partial y}{\partial x}(p)|$. Thus, using the fact that δ_p is a density on T_pM, we have

$$\begin{aligned}\rho^x(p) = \delta_p\left(\left.\frac{\partial}{\partial x^1}\right|_p, \dots, \left.\frac{\partial}{\partial x^d}\right|_p\right) &= \delta_p\left(A\left(\left.\frac{\partial}{\partial y^1}\right|_p\right), \dots, A\left(\left.\frac{\partial}{\partial y^d}\right|_p\right)\right) \\ &= |\det A|\delta_p\left(\left.\frac{\partial}{\partial y^1}\right|_p, \dots, \left.\frac{\partial}{\partial y^d}\right|_p\right) = \left|\frac{\partial y}{\partial x}(p)\right| \rho^y(p).\end{aligned}$$

□

Proposition 12.19 shows how to construct a smooth measure from a smooth density field. In fact, every smooth measure arises this way, and the resulting correspondence between smooth measures and smooth density fields is a bijection.

THEOREM 12.20. *To each smooth measure μ there corresponds a unique smooth density field δ related to μ as in Proposition 12.19, i.e. $\mu = \mu_\delta$. Hence, this yields a bijection between smooth measures and smooth density fields.*

PROOF. Let $\{(U_j, x_j)\}$ be a countable collection of charts on the smooth manifold M such that $\bigcup_j U_j = M$. At each point $p \in U_j$, T_pM has $\{\frac{\partial}{\partial x_j^1}|_p, \dots, \frac{\partial}{\partial x_j^d}|_p\}$ as a basis; let δ_p^j be the density on T_pM determined by this basis, cf. (12.2). It is left as an exercise to check that δ^j is a smooth density field on U_j.

By assumption, $\mu|_{U_j}$ has a smooth density $\frac{d\mu}{dx_j}$ in U_j. Let $\{\psi_j\}$ be a partition of unity subordinate to the cover $\{U_j\}$. Define

$$\delta = \sum_j \psi_j \frac{d\mu}{dx_j} \delta^j. \tag{12.4}$$

Since ψ_j is supported inside the open set U_j, and since the partition of unity is locally finite, the sum is well defined and yields a smooth density field on all of M.

Now, if (V, y) is any chart on M, the local density ρ^y on V induced by δ is

$$\rho^y(p) = \delta_p\left(\frac{\partial}{\partial y^1}\bigg|_p, \dots, \frac{\partial}{\partial y^d}\bigg|_p\right) = \sum_j \psi_j(p)\frac{d\mu}{dx_j}(p)\delta_p^j\left(\frac{\partial}{\partial y^1}\bigg|_p, \dots, \frac{\partial}{\partial y^d}\bigg|_p\right).$$

Now, for $p \in U \cap V$, the value of δ_p^j evaluated on the y basis vector fields is, by definition, the absolute value of the determinant of the matrix whose (k, ℓ)-entry is the kth component of $\frac{\partial}{\partial y^\ell}|_p$ in the x_j basis of vector fields. Referring to (12.3) (with the roles of x and y reversed), this matrix is $[D(x_j \circ y^{-1})] \circ y(p)$, and hence the determinant is $|\frac{\partial x_j}{\partial y}(p)|$. Hence

$$\rho^y(p) = \sum_j \psi_j(p)\frac{d\mu}{dx_j}(p)\left|\frac{\partial x_j}{\partial y}(p)\right|.$$

But from Lemma 12.8, $\frac{d\mu}{dy} = \frac{d\mu}{dx_j}\left|\frac{\partial x_j}{\partial y}\right|$ on $U \cap V$. Thus, for $p \in U \cap V$,

$$\rho^y(p) = \sum_j \psi_j(p)\frac{d\mu}{dy}(p) = \frac{d\mu}{dy}$$

where the last equality follows from the partition of unity property. By Theorem 12.11, this shows that R induces the smooth measure μ, as claimed. Together with Proposition 12.19, this shows that there is a bijective correspondence. Between smooth measures and smooth density fields. □

REMARK 12.21. The above proof shows something computationally useful. The sum (12.4) used to define a density field corresponding to a given smooth measure μ is, a priori, highly dependent on the choice of charts and the partition of unity used. But as the proof shows, with any such choices, the resulting density field δ yields the original measure μ; and since (by Theorem 12.11 and Proposition 12.19) each smooth measure corresponds to exactly one density field, it follows that the sum (12.4) actually defines a density field that is independent of the choice of charts and partition of unity. This will be important in the sequel.

The last element we need before constructing Haar measure is a global change-of-variables formula for smooth measures. First we must see how densities push forward under smooth maps.

LEMMA 12.22. *Let M and N be smooth manifolds of dimension d, and let $F\colon M \to N$ be a local diffeomorphism. Let δ be a smooth density field on N. Define a new function $F^*\delta\colon \bigsqcup (T_pM)^d \to \mathbb{R}_+$ as follows:*

$$(F^*\delta)_p(X_1, \dots, X_d) = \delta_{F(p)}(dF_p(X_1), \dots, dF_p(X_d)).$$

Then $F^*\delta$ is a smooth density field on M.

PROOF. Let $(X_1, \dots, X_d)$ be a local frame of smooth vector fields on an open subset U in M. Since F is a local diffeomorphism, it is an open map, and $dF_p\colon T_pM \to T_{F(p)}N$ is an isomorphism for each p; hence $(dF_p(X_1|_p), \dots, dF_p(X_d|_p))$ is a frame on the open set $F(U)$ in N. The vector fields are also smooth: if $f \in C^\infty(N)$, then $f \circ F \in C^\infty(M)$, and so

$$p \mapsto dF_p(X_j|_p)(f) = X_j|_p(f \circ F) \text{ is smooth.}$$

This verifies item (2) in Definition 12.17 of a smooth density field; we must now verify that $(F^*\delta)_p$ is a density on T_pM for each $p \in M$.

Let $p \in M$ and fix $A \in \mathrm{End}(T_pM)$. We use the isomorphism $dF_p\colon T_pM \to T_{F(p)}N$ to conjugate A to produce an endomorphism A' of $T_{F(p)}N$:

$$A' = dF_p \circ A \circ dF_p^{-1} \in \mathrm{End}(T_{F(p)}N).$$

Note that $dF_p \circ A = A' \circ dF_p$. Thus, for $X_1, \dots, X_d \in T_pM$,

$$\begin{aligned}(F^*\delta)_p(A(X_1), \dots, A(X_d)) &= \delta_{F(p)}(dF_pA(X_1), \dots, dF_pA(X_d)) \\ &= \delta_{F(p)}(A'dF_p(X_1), \dots, A'dF_p(X_d)) \\ &= |\det A'|\delta_{F(p)}(dF_p(X_1), \dots, dF_p(X_d)) \\ &= |\det A|(F^*\delta)_p(X_1, \dots, X_d).\end{aligned}$$

In the last equality, $\det A' = \det A$ because A and A' are conjugate by an isomorphism. This verifies item (1) in Definition 12.17, and concludes the proof. □

REMARK 12.23. The proof of the transformation property above does not require F to be a diffeomorphism, just smooth. But smoothness can fail at points p where dF_p is not surjective; at such points $(F^*\delta)_p \equiv 0$, allowing the determinant to change sign, introducing a cusp due to the absolute value.

We now come to the change-of-variables theorem for smooth measures / smooth density fields.

THEOREM 12.24. *Let M and N be smooth d-dimensional manifolds, and let $F\colon M \to N$ be a diffeomorphism. Let δ be a smooth density field on M, and let μ_δ be the associated smooth measure. Then the push-forward $F_*\mu_\delta$ is a smooth measure on N, and its density is $(F^{-1})^*\delta$. In other words: for $f \in C_c(M)$,*

$$\int_M f d\mu_\delta = \int_N f \circ F^{-1}\, d(F_*\mu_\delta) = \int_N f \circ F^{-1} d\mu_{(F^{-1})^*\delta}.$$

PROOF. We will prove the equivalent statement (reversing the roles of M and N, thus replacing F^{-1} with F) that if δ is a smooth density field on N, then for $f \in C_c(N)$

$$\int_M f \circ F\, d\mu_{F^*\delta} = \int_N f\, d\mu_\delta. \tag{12.5}$$

To evaluate the integrals on the two sides, we will use (12.4); as spelled out in Remark 12.21, this is well-defined independent of which choice of charts and partition of unity is used. Thus we will choose different charts and partitions of unity on the two sides to aid in the proof.

Fix a countable family $\{(U_j, x_j)\}$ of charts covering N, and a subordinate partition of unity $\{\psi_j\}$. Then

$$\int_N f\, d\mu_\delta = \sum_j \int_{U_j} f\psi_j \rho^{x_j}\, dx_j \tag{12.6}$$

where for $q \in U_j$, $\rho^{x_j}(q) = \delta_q\left(\frac{\partial}{\partial x_j^1}|_q, \dots, \frac{\partial}{\partial x_j^d}|_q\right)$.

Now, for each chart (U_j, x_j), let $V_j = F^{-1}(U_j)$ and $y_j = x_j \circ F$; then $\{\psi_j \circ F\}$ is a partition of unity subordinate to the $\{V_j\}$. So we may compute the integral on the left-hand-side of (12.5) as

$$\int_M f \circ F\, d\mu_{F^*\delta} = \sum_j \int_{V_j} (f \circ F)(\psi_j \circ F)\, (F^*\rho)^{y_j}\, dy_j$$

where, for $p \in V_j$, $(F^*\rho)^{y_j}(p) = (F^*\delta)_p\left(\frac{\partial}{\partial y_j^1}|_p, \dots, \frac{\partial}{\partial y_j^d}|_p\right)$. From the definition,

$$(F^*\delta)_p\left(\left.\frac{\partial}{\partial y_j^1}\right|_p, \dots, \left.\frac{\partial}{\partial y_j^d}\right|_p\right) = \delta_{F(p)}\left(dF_p\left(\left.\frac{\partial}{\partial y_j^1}\right|_p\right), \dots, dF_p\left(\left.\frac{\partial}{\partial y_j^d}\right|_p\right)\right).$$

It follows immediately from Proposition 3.17 that $dF_p(\frac{\partial}{\partial y_j^k}|_p) = \frac{\partial}{\partial x_j^k}|_{F(p)}$ (note that the roles of x and y are reversed here from that Proposition). Hence

$$(F^*\rho)^{y_j}(p) = \delta_{F(p)}\left(\left.\frac{\partial}{\partial x_j^1}\right|_{F(p)}, \dots, \left.\frac{\partial}{\partial x_j^d}\right|_{F(p)}\right) = \rho^{x_j}(F(p)).$$

Finally, since $y_j = x_j \circ F$, we have $dy_j = (F^{-1})_*(dx_j)$, and so in total

$$\int_M f \circ F d\mu_{F^*\delta} = \sum_j \int_{F^{-1}(U_j)} (f \circ F)(\psi_j \circ F)(\rho^{x_j} \circ F)\, (F^{-1})_*(dx_j).$$

Applying the measure theoretic change of variables formula, this becomes the right-hand-side of (12.6), thus proving (12.5), concluding the proof. □

2.3. Haar Measure. We will now construct two smooth measures on each Lie group that are invariant under (left and right) translations by group elements. These are the **Haar measures** on the group. Appealing to Theorem 12.20, it suffices to construct a smooth density field with appropriate invariance properties.

DEFINITION 12.25. *Let G be a Lie group. Fix any non-zero density δ_e on its tangent space at the identity T_eG. Define two functions*

$$\delta^L\colon \bigsqcup_{g\in G} T_gG \to \mathbb{R}_+ \qquad \textit{and} \qquad \delta^R\colon \bigsqcup_{g\in G} T_gG \to \mathbb{R}_+$$

as follows: for $g \in G$ and $X_1, \dots, X_d \in T_gG$, the vectors $dL_{g^{-1}}|_g(X_1), \dots, dL_{g^{-1}}|_g(X_d)$ and $dR_{g^{-1}}|_g(X_1), \dots, dR_{g^{-1}}|_g(X_d)$ are in T_eG. Set

$$\delta_g^L(X_1, \dots, X_d) := \delta_e\left(dL_{g^{-1}}|_g(X_1), \dots, dL_{g^{-1}}|_g(X_d)\right)$$
$$\delta_g^R(X_1, \dots, X_d) := \delta_e\left(dR_{g^{-1}}|_g(X_1), \dots, dR_{g^{-1}}|_g(X_d)\right).$$

PROPOSITION 12.26. *The functions δ^L, δ^R in Definition 12.25 are smooth density fields on G.*

PROOF. The proof here is very similar to the proof of Lemma 12.22, with the added wrinkle that the diffeomorphism F being applied here is $F = L_g$ or $F = R_g$ which also depends on the variable g of the putative density field. We will go through the details again, to prove that δ^L is a smooth density field; the proof for δ^R is completely analogous. Let $U \subseteq M$ be open, and let $X_1, \dots, X_d \in \mathscr{X}(U)$ be a smooth local frame of vector fields. For each g and j,

$$dL_{g^{-1}}|_g(X_j(g))$$

is a vector in T_eG; its action on a test function $f \in C^\infty(G)$ is

$$[dL_{g^{-1}}|_g(X_1(g))](f) = [X_j(g)](f \circ L_{g^{-1}}).$$

The two-variable function $(g, h) \mapsto f \circ L_{g^{-1}}(h) = f(g^{-1}h)$ is smooth, owing to the smoothness of group multiplication, and the smoothness of f. A calculation in local coordinates then shows that $g \mapsto [X_j(g)](f \circ L_{g^{-1}})$ is smooth on U (see the proof of Lemma 5.23). This shows that $dL_{g^{-1}}|_g(X_j(g))$ are smooth T_eG-valued functions. It follows that, for any basis $\{e_i\}_{i=1}^d$ of T_eG, the matrix-valued function

$$g \mapsto [e_i^*(dL_{g^{-1}}|_g(X_j(g))]_{i,j=1}^d$$

is smooth, and of full-rank for all g (since the $X_j(g)$ are a basis at each point, and $dL_{g^{-1}}|_g$ is a linear isomorphism for each g). Thus, the absolute value of its determinant is a smooth function. Since the chosen density δ_e has this form for some basis $\{e_j\}$ (cf. Lemma 12.15 and Remark 12.16), we conclude that δ satisfies the smoothness condition in Definition 12.17 of a smooth density field. Hence, it remains to show that δ_g^L is a density on T_gG for each $g \in G$; but this is *exactly* proved in the second half of the proof of Lemma 12.22, with $F = L_{g^{-1}}$ and $p = e$.

□

This finally brings us to the Haar measures.

THEOREM 12.27. *Let G be a Lie group. There are smooth positive measures λ^L and λ^R on G that are left and right translation invariant:*

$$(L_g)_*\lambda^L = \lambda^L \qquad \textit{and} \qquad (R_g)_*\lambda^R = \lambda^R \qquad \textit{for all } \ g \in G.$$

They are called the (left and right) **Haar measures** *on G. Any two left Haar measures are scalar multiples of each other, and the same holds for any two right Haar measures.*

PROOF. We work with left Haar measure here; the case for right Haar measure is analogous. Let δ^L be a left-invariant density field on G, cf. Definition 12.25 and Proposition 12.26. By Proposition 12.19, there is a unique smooth measure $\lambda^L = \mu_{\delta^L}$ on G corresponding to the density field δ^L. If $g \in G$, then since L_g is a diffeomorphism of G, by the change of variables Theorem 12.24,

$$(L_g)_*\lambda^L = (L_g)_*\mu_{\delta^L} = \mu_{(L_g)_*^{-1}\delta^L}.$$

But δ^L was designed to be left-invariant. That is: since $(L_g)^{-1} = L_{g^{-1}}$, for any $h \in G$ and $X_1, \dots, X_d \in T_hG$, from Lemma 12.22 we have

$$\left((L_{g^{-1}})^*\delta^L\right)_h(X_1, \dots, X_d) = \delta^L_{g^{-1}(h)}(dL_{g^{-1}}|_h(X_1), \dots, dL_{g^{-1}}|_h(X_d)).$$

By definition of δ^L, this is equal to

$$\delta_e(dL_{(g^{-1}h)^{-1}}|_{g^{-1}h} \circ dL_{g^{-1}}|_h(X_1), \ldots, dL_{(g^{-1}h)^{-1}}|_{g^{-1}h} \circ dL_{g^{-1}}|_h(X_d)).$$

By the chain rule,

$$dL_{(g^{-1}h)^{-1}}|_{g^{-1}h} \circ dL_{g^{-1}}|_h = dL_{h^{-1}g}|_{g^{-1}h} \circ dL_{g^{-1}}|_h = d(L_{h^{-1}g} \circ L_{g^{-1}})|_h.$$

But $L_a \circ L_b = L_{ab}$, and so $L_{h^{-1}g} \circ L_{g^{-1}} = L_{h^{-1}}$. The upshot is therefore that

$$\left((L_{g^{-1}})^*\delta^L\right)_h (X_1, \ldots, X_d) = \delta_e(dL_{h^{-1}}|_h(X_1), \ldots, dL_{h^{-1}}|_h(X_d)) = \delta^L_h(X_1, \ldots, X_d).$$

That is: $(L_{g^{-1}})^*\delta^L = \delta^L$, as expected. Thus $(L_g)_*\lambda^L = \mu_{\delta^L} = \lambda^L$ as claimed.

Now, suppose μ is another left-invariant smooth measure on G. Since μ is smooth, by Theorem 12.20 it has a smooth density, $\mu = \mu_{\overline{\delta}}$ for some smooth density field $\overline{\delta}$. By assumption μ is left-invariant, so $(L_g)_*\mu = \mu$ for any $g \in G$. Again by the change of variables Theorem 12.24, $\mu_{\overline{\delta}} = (L_g)_*\mu_{\overline{\delta}} = \mu_{(L_{g^{-1}})^*\overline{\delta}}$. By the bijective correspondence between smooth measures and densities, it follows that $\overline{\delta} = (L_{g^{-1}})^*\overline{\delta}$ for all $g \in G$. In particular, that means that, for $X_1, \ldots, X_d \in T_gG$,

$$\overline{\delta}_g(X_1, \ldots, X_d) = \left((L_{g^{-1}})^*\overline{\delta}\right)_g (X_1, \ldots, X_d) = \overline{\delta}_{g^{-1}g}(dL_{g^{-1}}(X_1), \ldots, dL_{g^{-1}}(X_d)).$$

In other words: $\overline{\delta} = \overline{\delta}^L$ starting with a (potentially different) density $\overline{\delta}_e$ on T_eG. But by Lemma 12.15, there is a constant c so that $\overline{\delta}_e = c \cdot \delta_e$, and backtracking, we find therefore that $\mu = c \cdot \lambda^L$, as claimed. □

REMARK 12.28. The proof above shows that there is a unique up to scale *smooth* left (resp. right) Haar measure. It does not preclude the possibility that there could be another non-smooth left- (resp. right-)invariant measure on the group. However, this is not the case, as the following clever exercise due to Terry Tao demonstrates.

EXERCISE 12.28.1. *Let λ and ν be two left-invariant measures on a Lie group G. Since $\lambda \ll \lambda + \mu$, there is a Radon-Nikodym derivative $f = \frac{d\lambda}{d(\lambda+\mu)}$. Show that $0 \leq f \leq 1$, and use the left-invariance of the measures (and some averaging) to prove that f is constant a.e. with respect to $\lambda + \mu$. Conclude that λ and μ are proportional measures.*

EXAMPLE 12.29.
(1) On the additive Lie group $\mathbb{R}^d$, Lebesgue measure is the (unique up to scale) left *and* right Haar measure. Note that it is an infinite measure.
(2) On the Lie group S^1, the (again bi-invariant) Haar measure is also given by Lebesgue integration (wrapped around):

$$\int_{S^1} f\, d\lambda = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\, d\theta.$$

In this case, the measure is finite, and so it is customary to use the unique scaling that makes it a probability measure.

As these two examples demonstrate, Haar measures may be finite or infinite; the distinction exactly corresponds to whether the group is compact or not.

THEOREM 12.30. *If G is a Lie group, then its (left and right) Haar measures are all finite if and only if G is compact; otherwise, all Haar measures are infinite.*

PROOF. As constructed above, left and right Haar measures are smooth measures. Any smooth measure on a compact manifold is finite (since the manifold may be covered by finitely many charts, and the measure has a smooth density in each chart which can be shown to be bounded, e.g. by the shrinking lemma).

Conversely, suppose G is not compact. Let U be a neighborhood of the identity with compact closure, and let $V \subseteq U$ be an open neighborhood of the identity with the property that $gh^{-1} \in U$ whenever $g, h \in V$. (The existence of such a neighborhood is a homework exercise.) Since $\overline{U}$ is compact but G is not, no finite collection of left-translates of U can cover G. Thus, we can inductively find a sequence $g_1, g_2, g_3 \dots$ of points in G such that $g_n \notin \bigcup_{j<n} g_j U$.

Claim. *The left translates $\{g_j V\}_{j \in \mathbb{N}}$ are all disjoint.* We prove this by contradiction: suppose there are $i < j$ with $g_i V \cap g_j V \neq \varnothing$. Hence there are $u, v \in V$ with $g_i u = g_j v$. But that means $g_j = g_i u v^{-1}$; by construction $uv^{-1} \in U$. This contradicts the construction of g_j which is in particular not in $g_i U$.

Thus, if λ^L is a left Haar measure on G, we have

$$\lambda^L\left(\bigsqcup_j g_j V\right) = \sum_j \lambda^L(g_j V) = \sum_j \lambda^L(V)$$

by left-invariance. **Claim.** $\lambda^L(V) > 0$. To prove this, consider the set $\{hV \colon h \in G\}$ of all left translates of V; this clearly covers G. By second-countability, there is a countable subcover, hence there are $\{h_j\} \in G$ with $G = \bigcup_j h_j V$. Hence

$$\lambda^L(G) \leq \sum_j \lambda^L(h_j V) = \sum_j \lambda^L(V).$$

If $\lambda^L(V) = 0$, it then follows that $\lambda^L \equiv 0$, contradicting the definition of Haar measure.

Hence, since $\lambda^L(V) > 0$, it follows that $\lambda^L(\bigsqcup_j g_j V) = \infty$. The analogous argument shows the same result for any right Haar measure. □

In Example (12.29), the Haar measures were bi-invariant: both left- and right-translation invariant (evident in those cases, since the groups were abelian). This need not be the case: left Haar measures need not be right Haar measures, and vice versa. They are, however, connected in a nice way in general.

Suppose λ^R is a right Haar measure on G. Fix any element $g \in G$, and consider the push-forward $(L_g)_* \lambda^R$. If $f \in C_c(G)$, then

$$\int_G f \, d((L_g)_* \lambda^R) = \int_{L_g^{-1}(G)} f \circ L_g \, d\lambda^R = \int_G f \circ L_g \, d\lambda^R.$$

Because λ^R is right invariant, for any $h \in G$, $\lambda^R = (R_h)_* \lambda^R$. Thus

$$\int_G f \circ L_g \, d\lambda^R = \int_G f \circ L_g \, d((R_h)_* \lambda^R) = \int_G f \circ L_g \circ R_h \, d\lambda^R.$$

Since the left and right actions commute, we then have

$$\int_G f \circ L_g \circ R_h \, d\lambda^R = \int_G f \circ R_h \circ L_g \, d\lambda^R = \int_G f \circ R_h \, d((L_g)_* \lambda^R).$$

Putting these statements together, we see that $L_{g*} \lambda^R$ is another right-invariant measure. Because L_g is a diffeomorphism and λ^R is a smooth measure, $(L_g)_* \lambda^R$ is also a smooth measure, cf. Theorem

12.24. Hence, but the uniqueness of (smooth) right Haar measure up to scale, there is a positive constant $\Delta_G(g)$ such that

$$L_{g*}\lambda^R = \Delta_G(g)\lambda^R. \tag{12.7}$$

DEFINITION 12.31. *The* **modular function** *of a Lie group G is the function $\Delta_G\colon G \to \mathbb{R}_+$ defined as above: $(L_g)_*\lambda^R = \Delta_G(g)\lambda^R$ for any right Haar measure λ^R. Stated alternatively: Δ_G is characterized by the property*

$$\int_G f(gh)\,\lambda^R(dh) = \Delta_G(g)\int_G f(h)\,\lambda^R(dh), \qquad \textit{for } f \in C_c(G).$$

From here, it is easy to deduce that $\Delta_G(gh) = \Delta_G(g)\Delta_G(h)$, and $\Delta_G(e) = 1$; hence, Δ_G is a group homomorphism from G to $\mathbb{R}_+$.

A group is called **unimodular** *if $\Delta_G \equiv 1$; this is equivalent to all Haar measures on G being bi-invariant.*

Clearly all abelian Lie groups are unimodular. Another class of groups that are unimodular is compact groups.

PROPOSITION 12.32. *All compact Lie groups are unimodular.*

PROOF. The modular function Δ_G is defined by (12.7). For any given $g \in G$, evaluate this equation applied to the whole group G:

$$\Delta_G(g)\lambda^R(G) = (L_{g*}\lambda^R)(G) = \lambda^R(L_g(G)) = \lambda^R(G).$$

Since G is compact, $\lambda^R(G)$ is a positive, finite number; hence we can divide through by it to find $\Delta_G(g) = 1$. □

Hence, on compact Lie groups, we will usually refer to **the Haar measure**: it is both left and right invariant, and it is unique when scaled to be a probability measure.

EXAMPLE 12.33. $\mathrm{O}(n,\mathbb{R}) = F^{-1}(I_n)$ where $F(A) = AA^\top$ is continuous; hence $\mathrm{O}(n,\mathbb{R})$ is a closed subset of $M_n(\mathbb{R})$. Moreover, if $Q \in \mathrm{O}(n,\mathbb{R})$, then $n = \mathrm{Tr}(I_n) = \mathrm{Tr}(QQ^\top)$, and this is equal to the sum of the squares of all matrix entries; i.e. the (squared) Hilbert–Schmidt norm, which is the standard Euclidean norm on $M_n(\mathbb{R})$. Thus $\mathrm{O}(n,\mathbb{R})$ is contained in the sphere of radius $\sqrt{n}$, which is compact. So $\mathrm{O}(n,\mathbb{R})$ is a closed subset of a compact set, and is hence a compact Lie group. Thus it is unimodular. Analogous arguments apply to show the compactness of $\mathrm{SO}(n)$, $\mathrm{Sp}(n)$, $\mathrm{SU}(n)$, and $\mathrm{U}(n)$.

There are other, non-abelian, non-compact Lie groups that are unimodular (e.g. $\mathrm{GL}(n,\mathbb{F})$ and $\mathrm{SL}(n,\mathbb{F})$ for $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$). Later, we will see an explicit formula for the modular function on a Lie group, with useful characterizations of unimodularity. *A word of caution*: subgroups of unimodular groups need not be unimodular!

EXAMPLE 12.34. The *affine group* is the group of invertible affine transformations of $\mathbb{R}^d$; let's deal here with the case $d = 1$. That is, it is the set of functions $\mathbb{R} \to \mathbb{R}$ of the form $T(x) = ax + b$ for $a \neq 0$ and $b \in \mathbb{R}$. As a manifold it is diffeomorphic to $\mathbb{R}^* \times \mathbb{R}$; as a group it is a semidirect product of these. Perhaps the easiest way to view it is as a matrix group:

$$\left\{\begin{bmatrix} a & b \\ 0 & 1/a \end{bmatrix} : a \in \mathbb{R}\setminus\{0\}, b \in \mathbb{R}\right\}.$$

From this representation it is clear that the affine group is a Lie subgroup of $\mathrm{SL}(2,\mathbb{R})$, which is unimodular. However, the affine group is *not unimodular*. This could be computed explicitly in

this relatively simple case, but it would still take a page of calculations (first to compute the right Haar measure in these global coordinates, then to compute the modular function). We will defer this calculation until later, when we have studied the adjoint representations of Lie groups; we will prove a general formula for the modular function.

3. Lie Algebras

A Lie group G is a manifold with extra structure: the structure of a group. This is reflected in extra structure in the tangent spaces as well. This can be seen at the level of left-invariant vector fields, which form a *Lie algebra*, whose definition we now recall from Section 5.

DEFINITION 12.35. *A* **Lie algebra** *is a vector space* $\mathfrak{g}$ *equipped with an operation known as the* **bracket** $[\cdot,\cdot]\colon \mathfrak{g}\times\mathfrak{g}\to\mathfrak{g}$*, satisfying the following properties (as spelled out in Proposition 4.21): for* $a,b\in\mathbb{R}$ *and* $X,Y,Z\in\mathfrak{g}$*,*

(a) BILINEARITY*:*

$$[aX+bY,Z]=a[X,Z]+b[Y,Z],$$
$$[Z,aX+bY]=a[Z,X]+b[Z,Y].$$

(b) ANTISYMMETRY*:*

$$[X,Y]=-[Y,X].$$

(c) JACOBI IDENTITY*:*

$$[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]=0.$$

EXAMPLE 12.36. (a) Any vector space can be made into a Lie algebra by setting $[X,Y]=0$ for all X,Y. Such a Lie algebra is called **abelian**.
(b) The Lie bracket of Section 5 makes $\mathscr{X}(M)$ into a(n infinite-dimensional) Lie algebra.
(c) Let $\mathscr{A}$ be an (associative) $\mathbb{R}$-algebra. Then the **commutator bracket** $[a,b]=ab-ba$ defines a Lie algebra structure on $\mathscr{A}$.

We may now add another example to this list: the Lie bracket of vector fields makes $\mathscr{X}^L(G)$ into a Lie subalgebra of $\mathscr{X}(G)$.

PROPOSITION 12.37. *Let* $X,Y\in\mathscr{X}^L(G)$ *be left-invariant vector fields. Then* $[X,Y]$ *is also left-invariant.*

PROOF. This follows immediately from Corollary 4.23: for $g\in G$, since L_g is a diffeomorphism, and since $(L_g)_*X=X$ and $(L_g)_*Y=Y$ by assumption, we have

$$(L_g)_*[X,Y]=[(L_g)_*X,(L_g)_*Y]=[X,Y].$$

Thus $[X,Y]\in\mathscr{X}^L(G)$. □

This gives us the extra algebraic structure on T_eG: it is (up to its isomorphism with $\mathscr{X}^L(G)$) a Lie algebra.

DEFINITION 12.38. *Let* G *be a Lie group. Denote the Lie algebra of left-invariant vector fields* $\mathscr{X}^L(G)$ *as* $\mathrm{Lie}(G)$*, the* **Lie algebra of** G*. Common notation is to use lower-case Fraktur letters:* $\mathrm{Lie}(G)=\mathfrak{g}$*,* $\mathrm{H}=\mathfrak{h}$*, etc.*

So the tangent space at the identity of any Lie group G has the structure Lie algebra. In order to understand its Lie algebra structure, one in principle needs to identify T_eG with $\mathscr{X}^L(G)$ and take the bracket there. In practice, it is usually much easier to identify the bracket. To see how, we must first develop some basic properties of Lie algebras, beginning with their homomorphisms.

DEFINITION 12.39. *Let $\mathfrak{g}, \mathfrak{h}$ be two Lie algebras. A linear map $\phi\colon \mathfrak{g} \to \mathfrak{h}$ is called a* **Lie algebra homomorphism** *if*

$$[\phi(X), \phi(Y)]_{\mathfrak{h}} = \phi([X, Y]_{\mathfrak{g}}).$$

If ϕ is also a vector space isomorphism, we call it a **Lie algebra isomorphism**.

EXAMPLE 12.40. (a) If $\mathfrak{g}, \mathfrak{h}$ are abelian Lie algebras, then any linear map between them is a Lie algebra homomorphism.

(b) Let $\mathscr{A}, \mathscr{B}$ be a algebras, turned into Lie algebras via their commutator brackets. If $\phi\colon \mathscr{A} \to \mathscr{B}$ is an algebra homomorphism, then it is also a Lie algebra homomorphism:

$$[\phi(a), \phi(a')]_{\mathscr{B}} = \phi(a)\phi(a') - \phi(a')\phi(a) = \phi(aa' - a'a) = \phi([a, a']_{\mathscr{A}}.$$

Not all Lie algebra homomorphisms are of this form, as we will see.

When $\mathfrak{g}$ is the Lie algebra of a Lie group G, then Lie *group* homomorphisms of G yield Lie *algebra* homomorphisms of $\mathfrak{g}$, just by differentiating at the identity. To see how this works, we first need the following lemma, which says that in the category of left-invariant vector fields, we can push-forward by a Lie group homomorphism even if it is not a diffeomorphism.

LEMMA 12.41. *Let G, H be Lie groups and $F\colon G \to H$ a Lie group homomorphism. For every $X \in \mathrm{Lie}(G)$, there is a unique $Y \in \mathrm{Lie}(H)$ so that F-related to X: namely $Y = \widetilde{dF_e(X_e)}$. We denote this as $Y = F_*X$.*

PROOF. If X and Y are F-related then, by definition, $Y_{F(g)} = dF_g(X_g)$ for any $g \in G$; since $F(e) = e$, this shwos that $Y_e = dF_e(X_e)$. Thus, since Y_e is left-invariant, it must be true that $Y = \widetilde{dF_e(X_e)}$. So we need to verify that this vector field is, indeed, F related to X. To show this, we use the now familiar argument of differentiating the homomorphism property: since $F(L_g(g')) = F(gg') = F(g)F(g') = L_{F(g)}(F(g'))$, we have $F \circ L_g = L_{F(g)} \circ F$, and so

$$dF \circ d(L_g) = d(L_{F(g)}) \circ dF.$$

Thus, for any point $g \in G$,

$$dF(X_g) = dF(d(L_g)(X_e)) = d(L_{F(g)})(dF(X_e)) = d(L_{F(g)})(Y_e) = Y_{F(g)}$$

which shows that X and Y are F-related, as claimed. □

So we can consider the push-forward F_*X of any left-invariant vector field by a Lie group homomorphism. In this context, Corollary 4.23 still verifies that $[F_*X, F_*Y] = F_*[X, Y]$. This gives us the following.

PROPOSITION 12.42. *Let G, H be Lie groups, and let $F\colon G \to H$ be a Lie group homomorphism. Then $F_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ is a Lie algebra homomorphism.*

PROOF. Since F_* is a linear map, we need only verify that it preserves the bracket, which is the statement of Corollary 4.23. □

REMARK 12.43. We will see later that, remarkably, the converse is true if G is simply connected: in this case, any Lie algebra homomorphism $\mathrm{Lie}(G) \to \mathrm{Lie}(H)$ is of the form F_* for a unique Lie group homomorphism $G \to H$.

REMARK 12.44. Under the isomorphism $\lambda\colon T_eG \to \mathscr{X}^L(G)$, the map F_* becomes simple dF_e. So the statement is that, viewing the Lie algebra as T_eG, differentiating any Lie *group* homomorphism at the identity yields a Lie *algebra* homomorphism.

Here are some elementary properties of induces Lie algebra homomorphisms. The proof is simple and left to the reader.

PROPOSITION 12.45. *Let G, H, K be Lie groups, and let $F_1\colon G \to H$ and $F_2\colon H \to K$ be Lie group homomorphisms.*

(a) $(\mathrm{Id}_G)_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ *is the identity map.*
(b) $(F_2 \circ F_1)_* = (F_2)_* \circ (F_1)_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(K)$.

Consequently, isomorphic Lie groups have isomorphic Lie algebras.

REMARK 12.46. Again, remarkably, the converse is true in the simply connected case: if G, H are simply connected Lie groups and $\mathrm{Lie}(G) \cong \mathrm{Lie}(H)$, then $G \cong H$. We will prove this later.

It follows that we can identify the Lie algebra of a subgroup of G as a Lie subalgebra of $\mathrm{Lie}(G)$.

COROLLARY 12.47. *Let $H \subseteq G$ be a Lie subgroup, and let $\iota\colon H \hookrightarrow G$ be the inclusion map. There is a Lie subalgebra $\mathfrak{h} \subseteq \mathrm{Lie}(G)$ that is canonically isomorphic to $\mathrm{Lie}(H)$; it is given by*

$$\mathfrak{h} = \iota_*\left(\mathrm{Lie}(H)\right) = \{Y \in \mathrm{Lie}(G)\colon Y_e \in T_eH \subseteq T_eG\}.$$

PROOF. Since ι is a Lie group homomorphism, it follows that $\iota_*\colon \mathrm{Lie}(H) \to \mathrm{Lie}(G)$ is a Lie algebra homomorpism. Since the inclusion is an immersion (by definition of Lie subgroup), ι_* is injective, and so its image $\mathfrak{h}$ is linearly isomorphic to $\mathrm{Lie}(H)$; thus, as ι_* is a Lie algebra homomorphism, it is a Lie algebra isomorphism. We therefore need only to verify the second equality. Given $X \in \mathrm{Lie}(H)$, to say that $Y = \iota_* X$ is to say that X, Y are ι-related, meaning that $d\iota_h(X_h) = Y_{\iota(h)} = Y_h$. This is precisely how we identify T_hH as a subspace of T_hG, and so it follows that $Y = \iota_* X$ if and only if $Y_h \in T_hH$ for all $h \in H$. At $h = e$, this yields the forward inclusion: if $Y \in \mathfrak{h}$, then $Y_e \in T_eH$. Conversely, if $Y_e \in T_eH \subseteq T_eG$, this means there is a vector $X_e \in T_eH$ so that $Y_e = d\iota_e(X_e)$. Now Y is left-invariant, so for $h \in H$,

$$Y_h = d(L_h)_e(Y_e) = d(L_h)_e \circ d\iota_e(X_e) = d(L_h \circ \iota)_e(X_e).$$

For $h \in H$, we may easily verify $L_h \circ \iota = \iota \circ L_h$ as functions on G, and so

$$Y_h = d(\iota \circ L_h)_e(X_e) = d\iota_h(\widetilde{X}_h) = \iota_* \widetilde{X}_h.$$

Since the vector field $\widetilde{X}$ is left-invariant, it is in $\mathrm{Lie}(H)$, and this shows $Y \in \mathfrak{h}$. □

The Lie algebra of a Lie group G is the space of left-invariant vector fields on G. While this makes the action of the bracket clear, it is a challenge to understand the actual space. On the other hand, we have a linear isomorphism $\lambda\colon T_eG \to \mathrm{Lie}(G)$, so we would like to think of the Lie algebra as the tangent space at the identity, which is an object we can easily understand. The trick is then figuring out what λ does to the Lie bracket. In most of the cases we are interested in, this is not hard to understand. We begin with abelian groups.

PROPOSITION 12.48. *Let G be an abelian Lie group: $gh = hg$ for all $g, h \in G$. Then the Lie algebra $\mathrm{Lie}(G)$ is also abelian: $[X, Y] = 0$ for all $X, Y \in \mathrm{Lie}(G)$.*

PROOF. If G is abelian, then the map $\mathrm{inv}\colon G \to G$ given by $\mathrm{inv}(g) = g^{-1}$ is actually a group isomorphism. Thus, it induces a Lie algebra isomorphism $\mathrm{inv}_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$. Recall that $\mathrm{inv}_*(X)$ is the left-invariant extension of the vector $d\mathrm{inv}_e(X_e)$. We can compute $d\mathrm{inv}_e$ easily. In the proof of Lemma 11.2, we showed that, for the multiplication map $\mathrm{m}\colon G \times G \to G$,

$$d\mathrm{m}_{g,h}(X_g, Y_h) = d(R_h)_g(X_g) + d(L_g)_h(Y_h), \qquad X_g \in T_gG, Y_h \in T_hG.$$

At $g = h = e$, this says simply that $d\mathrm{m}_{(e,e)}(X_e, Y_e) = X_e + Y_e$ for $X_e, Y_e \in T_eG$. Now, $\mathrm{m} \circ (\mathrm{Id}, \mathrm{inv})\colon G \to G$ is the constant map $g \mapsto e$, and so

$$0 = d\left(\mathrm{m} \circ (\mathrm{Id}, \mathrm{inv})\right)_e = d\mathrm{m}_{(e,e)} \circ (\mathrm{Id}_{T_eG}, d\mathrm{inv}_e).$$

That is, for $X_e \in T_eG$,

$$0 = d\mathrm{m}_{(e,e)} \circ (\mathrm{Id}_{T_eG}, d\mathrm{inv}_e)(X_e) = d\mathrm{m}_{(e,e)} \circ (X_e, d\mathrm{inv}_e(X_e)) = X_e + d\mathrm{inv}_e(X_e).$$

Thus $d\mathrm{inv}_e(X_e) = -X_e$. It follows by linearity that $\mathrm{inv}_*(X) = -X$ on $\mathrm{Lie}(G)$. We showed above that this is therefore a Lie algebra isomorphism, meaning that

$$-[X, Y] = \mathrm{inv}_*[X, Y] = [\mathrm{inv}_*X, \mathrm{inv}_*Y] = [-X, -Y] = [X, Y].$$

Thus $[X, Y] = 0$ for all $X, Y \in \mathrm{Lie}(G)$. □

Most (and all interesting) Lie groups are not abelian, and likewise their Lie algebras won't be. In order to understand them better, we need to delve a little deeper into the structure of left-invariant vector fields, and in particular their flows. But in the special case (which covers almost all interesting examples) that the group is a *matrix Lie group* (i.e. a Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$ or $\mathrm{GL}(n, \mathbb{C})$), the Lie bracket coincides with what one would expect.

PROPOSITION 12.49. *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The Lie algebra $\mathfrak{gl}(n, \mathbb{F})$ of $\mathrm{GL}(n, \mathbb{F})$ is isomorphic to $M_n(\mathbb{F})$, equipped with its typical commutator Lie bracket $[X, Y] = XY - YX$.*

This will follow from some general considerations in the next chapter, but the basic idea is a simple computation and it will pay for future understanding to do it here directly.

PROOF. Since $\mathrm{GL}(n, \mathbb{F})$ is an open subset of $M_n(\mathbb{F})$, the tangent space at the identity $T_I\mathrm{GL}(n, \mathbb{F})$ is canonically identified with $T_IM_n(\mathbb{F}) \cong M_n(\mathbb{F})$, cf. Proposition 3.11. Our goal is to show that, for $X, Y \in M_n(\mathbb{F})$, the associated left-invariant vector fields $\widetilde{X}, \widetilde{Y}$ satisfy

$$[\widetilde{X}, \widetilde{Y}]_I = XY - YX. \tag{12.8}$$

To prove this, we recall the definition of $\widetilde{X}$: for any $A \in \mathrm{GL}(n, \mathbb{F})$,

$$\widetilde{X}_A = dL_A|_I(X).$$

In other words: for $f \in C^\infty(\mathrm{GL}(n, \mathbb{F}))$, the function $\widetilde{X}(f)$ is given by

$$\left(\widetilde{X}(f)\right)(A) = \widetilde{X}_A(f) = dL_A|_I(X)(f) = X(f \circ L_A).$$

To evaluate this, we choose a smooth curve $\alpha_X\colon \mathbb{R} \to \mathrm{GL}(n, \mathbb{F})$ satisfying $\alpha_X(0) = I$ and $\dot{\alpha}_X(0) = X$. For any such curve, we have

$$\left(\widetilde{X}(f)\right)(A) = X(f \circ L_A) = \frac{d}{dt}\bigg|_{t=0} f \circ L_A \circ \alpha_X(t) = \frac{d}{dt}\bigg|_{t=0} f(A\alpha_X(t)).$$

We can now do this computation in local (in this case global) coordinates: the matrix entries A_{ij}. For the curve $A\alpha_X(t) \in \mathrm{GL}(n, \mathbb{F}) \subset M_n(\mathbb{F})$,

$$(A\alpha_X(t))_{ij} = \sum_k A_{ik}[\alpha_X(t)]_{kj}$$

and so by ordinary multivariate calculus, we have

$$\left.\frac{d}{dt}\right|_{t=0} (A\alpha_X(t))_{ij} = \sum_k A_{ik}[\dot{\alpha}_X(0)]_{kj} = \sum_k A_{ik}X_{kj} = (AX)_{ij}.$$

Letting $\hat{f}$ denote the expression for f in coordinates as usual, we then apply multivariate calculus to the full expression to find

$$\left(\widetilde{X}(f)\right)(A) = \left.\frac{d}{dt}\right|_{t=0} \hat{f}(A\alpha_X(t)) = [D\hat{f}(A)](AX) = \sum_{i,j} \frac{\partial \hat{f}}{\partial A_{ij}}(A)(AX)_{ij}.$$

Now, we must evaluate $\widetilde{Y}(\widetilde{X}(f))(I)$. For the moment denoting $g = \widetilde{X}(f)$, following the above analysis we find that

$$\widetilde{Y}(\widetilde{X}(f))(A) = \sum_{n,m} \frac{\partial \hat{g}}{\partial A_{nm}}(A)Y_{nm}.$$

Combining, we have

$$\widetilde{Y}(\widetilde{X}(f))(I) = \sum_{i,j,n,m} \frac{\partial}{\partial A_{nm}} \left(\frac{\partial \hat{f}}{\partial A_{ij}}(A)(AX)_{ij} \right)\Bigg|_{A=I} Y_{nm}.$$

For the inside derivative, we apply the product rule:

$$\frac{\partial}{\partial A_{nm}} \left(\frac{\partial \hat{f}}{\partial A_{ij}}(A)(AX)_{ij} \right) = \frac{\partial^2 \hat{f}}{\partial A_{ij}\partial A_{nm}}(A)(AX)_{ij} + \frac{\partial \hat{f}}{\partial A_{ij}}(A)\frac{\partial}{\partial A_{nm}}(AX)_{ij}$$

and expanding the matrix product we have

$$\frac{\partial}{\partial A_{nm}}(AX)_{ij} = \sum_k \frac{\partial}{\partial A_{nm}} A_{ik}X_{kj} = \sum_k \delta_{ni}\delta_{mk}X_{kj} = \delta_{ni}X_{mj}.$$

Altogether, then, we have

$$\begin{aligned}\widetilde{Y}(\widetilde{X}(f))(I) &= \sum_{i,j,n,m} \left(\frac{\partial^2 \hat{f}}{\partial A_{ij}\partial A_{nm}}(I)(IX)_{ij} + \frac{\partial \hat{f}}{\partial A_{ij}}(I)\delta_{ni}X_{mj} \right) Y_{nm} \\ &= \sum_{i,j,m,n} \frac{\partial^2 \hat{f}}{\partial A_{ij}\partial A_{nm}}(I)X_{ij}Y_{nm} + \sum_{i,j,m} \frac{\partial \hat{f}}{\partial A_{ij}}(I)X_{mj}Y_{im} \qquad (12.9)\end{aligned}$$

and the second term is just the matrix product expansion of

$$\sum_{i,j} \frac{\partial \hat{f}}{\partial A_{ij}}(I)(YX)_{ij}.$$

Now, repeating all this to evaluate $\widetilde{X}(\widetilde{Y}(f))(I)$ simply means reversing the roles of X and Y in the computations. Note that the first term in (12.9) is manifestly symmetric in X and Y; hence, in the difference it will be canceled, and so we are left with

$$\widetilde{X}(\widetilde{Y}(f))(I) - \widetilde{Y}(\widetilde{X}(f))(I) = \sum_{i,j} \frac{\partial \hat{f}}{\partial A_{ij}}(I)(XY - YX)_{ij}. \tag{12.10}$$

This is precisely the local coordinate expression for $XY - YX$ interpreted as a derivation at I. That is: if $\beta\colon \mathbb{R} \to \mathrm{GL}(n, \mathbb{F})$ is a curve with $\beta(0) = I$ and $\dot{\beta}(0) = XY - YX$, then viewing $XY - YX \in M_n(\mathbb{F})$ as a derivation at I, we have

$$(XY - YX)(f) = \frac{d}{dt}\bigg|_{t=0} f \circ \beta(t)$$

and computing this in local coordinates yields precisely (12.10). This establishes the desired equality (12.8), concluding the proof. □

Combining this result with Corollary 12.47 shows that all matrix Lie groups have matrix Lie algebras (i.e. with bracket given by the commutator of matrices).

COROLLARY 12.50. *Let G be a matrix Lie group: a Lie subgroup of $\mathrm{GL}(n, \mathbb{F})$ for some $n \in \mathbb{N}$ and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. then the Lie algebra of G is isomorphic to a Lie subalgebra of $\mathfrak{gl}(n, \mathbb{F})$.*

EXAMPLE 12.51. (1) The Lie group $\mathrm{SL}(n, \mathbb{F})$ is the regular level set $\det^{-1}(1)$ of the smooth map $\det\colon \mathrm{GL}(n, \mathbb{F}) \to \mathbb{F}$. The tangent space at the identity, $T_I\mathrm{SL}(n, \mathbb{F})$, is then identified with the subspace $\ker D(\det)_I$ of $M_n(\mathbb{F}) \cong T_I\mathrm{GL}(n, \mathbb{F})$. As we computed in Example 11.5(3), $D(\det)_I(X) = \mathrm{Tr}\,(X)$. Thus

$$\mathfrak{sl}(n, \mathbb{F}) := \mathrm{Lie}(\mathrm{SL}(n, \mathbb{F})) \cong \{X \in M_n(\mathbb{F})\colon \mathrm{Tr}\,(X) = 0\}.$$

By Corollary 12.50, $\mathfrak{sl}(n, \mathbb{F})$ must be a Lie subalgebra of $M_n(\mathbb{F})$; i.e. it must be closed under the commutator bracket. This is easy to show directly. Indeed: if X, Y are *any* matrices in $M_n(\mathbb{F})$, then $\mathrm{Tr}\,(XY - YX) = 0$ due to the cyclic trace property $\mathrm{Tr}\,(XY) = \mathrm{Tr}\,(YX)$ and linearity of Tr. (This shows that $\mathfrak{sl}(n, \mathbb{F})$ is actually an ideal in the larger Lie algebra $\mathfrak{gl}(n, \mathbb{F})$.)

(2) The Lie group $\mathrm{O}(n, \mathbb{F})$ is the regular level set $F^{-1}(I)$ where $F(A) = AA^\top$ viewed as a map from $\mathrm{GL}(n, \mathbb{F})$ to the vector space $M_n^{\mathrm{sym}}(\mathbb{F})$ of symmetric (real or complex) matrices. As computed in Example 1.5 in the real case, $DF_I(X) = X + X^\top$. Hence, the tangent space at the identity, $T_I\mathrm{O}(n, \mathbb{F})$, is identified with $\ker(DF_I) = \{X \in M_n(\mathbb{F})\colon X^\top = -X\}$, the space of *skew-symmetric* matrices. This is the Lie algebra of $O(n, \mathbb{F})$:

$$\mathfrak{o}(n, \mathbb{F}) = \{X \in M_n(\mathbb{F})\colon X^\top = -X\}$$

where the Lie bracket is the usual commutator of matrices. As above, it follows from Corollary 12.50 that this space is a Lie subalgebra of $\mathfrak{gl}(n, \mathbb{F})$, and hence closed under commutator bracket; this is also easy to see directly: if $X^\top = -X$ and $Y^\top = -Y$,

$$\begin{aligned}[X, Y]^\top = (XY - YX)^\top &= Y^\top X^\top - X^\top Y^\top \\ &= (-Y)(-X) - (-X)(-Y) = -[X, Y].\end{aligned}$$

Now, consider the subgroup $\mathrm{SO}(n, \mathbb{F}) \subset \mathrm{O}(n, \mathbb{F})$, i.e. $\mathrm{SO}(n, \mathbb{F}) = \mathrm{O}(n, \mathbb{F}) \cap \mathrm{SL}(n, \mathbb{F})$. If $Q \in \mathrm{O}(n, \mathbb{F})$ then $1 = \det(I) = \det(QQ^\top) = \det(Q)^2$, and so $\det Q = \pm 1$. Hence, we can also express $\mathrm{SO}(n, \mathbb{F}) = \mathrm{O}(n, \mathbb{F}) \cap \det^{-1}(B(1, 1))$ (where $D(1, 1)$ is the open ball of

radius 1 around 1 in $\mathbb{F}$. Thus $\mathrm{SO}(n,\mathbb{F})$ is an open subset of $O(n,\mathbb{F})$ which contains I, and thus $T_I\mathrm{SO}(n,\mathbb{F}) = T_I\mathrm{SO}(n,\mathbb{F})$. As such, the two groups have the same Lie algebra:

$$\mathfrak{so}(n,\mathbb{F}) = \mathfrak{o}(n,\mathbb{F}).$$

This can also be seen on the Lie algebra side directly: since $\mathrm{SO}(n,\mathbb{F})$ is the regular level set of $(F,\det)$ at the value $(I,1)$, its tangent space at the identity is the kernel of $D(F,\det)_{I,I}$, which is the intersection of the kernels of DF_I and $D(\det)_I$: i.e.

$$\mathfrak{so}(n,\mathbb{F}) = \{X \in M_n(\mathbb{F})\colon X^\top = -X\} \cap \{X \in M_n(\mathbb{F})\colon \mathrm{Tr}\,(X) = 0\}.$$

However: if $X^\top = -X$, for diagonal entries we have $-X_{jj} = (X^\top)_{jj} = X_jj$, and thus all diagonal entires are 0. Hence, $X \in \mathfrak{o}(n,\mathbb{F})$ automatically implies that $\mathrm{Tr}\,(X) = 0$.

(3) Very similar considerations show that the unitary group $U(n)$ has Lie algebra

$$\mathfrak{u}(n) = \{X \in M_n(\mathbb{C})\colon X^* = -X\}$$

the *skew-Hermitian* matrices. In this case, $\mathrm{SU}(n) = \mathrm{U}(n) \cap \mathrm{SL}(n,\mathbb{C})$ is *not* an open subgroup: as discussed in Exampled 11.5(5), $\mathrm{SU}(n)$ is a subgroup of codimension 1. This can also be seen at the level of Lie algebras: following (2) above, we have

$$\mathfrak{su}(n) = \mathfrak{u}(n) \cap \{X \in M_n(\mathbb{C})\colon \mathrm{Tr}\,(X) = 0\}.$$

If $X \in \mathfrak{u}(n)$, then $X^* = -X$ and so the diagonal entries satisfie $-X_{jj} = (X^*)_{jj} = \overline{X_{jj}}$; hence, they are purely imaginary. As such, the constraint $0 = \mathrm{Tr}\,(X) = X_{11} + \cdots + X_{nn}$ is a 1-dimensional constraint.

(4) Recall the Heisenberg group $\mathrm{H}(n,\mathbb{F}) = \{I + T\colon T \in M_n(\mathbb{F})$ is strictly upper-triangular$\}$. It therefore has the global parametrization $f\colon \mathbb{T}_n(\mathbb{F}) \to \mathrm{H}(n,\mathbb{F})$ where $T_n(\mathbb{F})$ is the vector space of strictly upper-triangular matrices and $f(X) = I + X$. Then the tangent space at I is identified with the *image* of Df_I. Note that $Df_A(X) = X$ is the identity map for any point A, and hence the Lie algebra of $\mathrm{H}(n,\mathbb{F})$ is just the space of strictly upper-triangular matrices

$$\mathfrak{h}(n,\mathbb{F}) = T_n(\mathbb{F}).$$

As the product of strictly upper triangular matrices is strictly upper-triangular, so is the commutator, and hence we see directly (what already followed from Corollary 12.50 above) that $T_n(\mathbb{F})$ is closed under the commutator bracket.

CHAPTER 13

The Exponential Map

1. One-Parameter Subgroups

Let G be a Lie group. Let us begin by noting that left-invariant vector fields in $\mathscr{X}^L(G)$ are necessarily complete.

LEMMA 13.1. *Let $X \in \mathscr{X}^L(G)$ be a left-invariant vector field. Then X is complete.*

PROOF. Here we use the Uniform Time Lemma 5.19: if there is a uniform interval $(-\epsilon, \epsilon)$ so that the integral curve θ^g of X starting at any point $g \in G$ exists on $(-\epsilon, \epsilon)$, then X is actually complete. Now, we assume that X is left-invariant, which means that X is L_g-related to itself. By Lemma 5.17 (naturality of flows), it follows that $L_g \circ \theta^e$ is an integral curve of X starting at g, and therefore is equal to θ^g. Thus, each integral curve is defined for at least the interval that θ^e is defined, which gives the uniformity, concluding the proof. $\square$

Hence, all left-invariant vector fields are complete, so their flows are global. The question is: what do their maximal integral curves (which are defined for all time) look like? To answer this, we define the following.

DEFINITION 13.2. *Let G be a Lie group. A group homomorphism $\alpha\colon \mathbb{R} \to G$ (where $\mathbb{R}$ has the usual additive Lie group structure) is called a* **one-parameter subgroup**.

This is strange but universal terminology, as it means a one-parameter subgroup is not a subgroup *per se*, although it will turn out that the image of such a homomorphism $\alpha(\mathbb{R})$ is a Lie subgroup (isomorphic to either $\mathbb{R}$, $\mathbb{S}^1$, or $\{e\}$). To be clear: a one-parameter subgroup is a smooth map $\alpha\colon \mathbb{R} \to G$ with the properties that $\alpha(0) = e$ and

$$\alpha(s)\alpha(t) = \alpha(s+t), \qquad s, t \in \mathbb{R}.$$

Our main theorem is that these are the (maximal) integral curves of left-invariant vector fields.

THEOREM 13.3. *Let G be a Lie group. If $\alpha\colon \mathbb{R} \to G$ is a one-parameter subgroup, then it is the maximal integral curve of the left-invariant vector field which is equal to $\dot{\alpha}(0)$ at e. Conversely, if $X \in \mathscr{X}^L(G) = \mathrm{Lie}(G)$ is any left-invariant vector field, then its maximal integral curve starting at e is a one-parameter subgroup.*

PROOF. First, suppose α is a one-parameter subgroup. As it is a Lie group homomorphism, there is an induced Lie algebra homomorphism $\alpha_*\colon \mathrm{Lie}(\mathbb{R}) \to \mathrm{Lie}(G)$. Note that the vector field $\frac{d}{dt} \in \mathscr{X}(\mathbb{R})$ is left-invariant. Define $X = \alpha_*(\frac{d}{dt})$, which is by definition in $\mathrm{Lie}(G)$. We will show that X is the infinitesimal generator of α. Indeed, by definition (cf. Lemma 12.41 and Proposition 12.42), $X = \alpha_*(\frac{d}{dt})$ is the unique left-invariant vector field on G that is α-related to $\frac{d}{dt}$, meaning that, for each $t_0 \in \mathbb{R}$,

$$\dot{\alpha}(t_0) = d\alpha_{t_0}\left(\left.\frac{d}{dt}\right|_{t=t_0}\right) = X_{\alpha(t_0)}.$$

That is precisely to say that α is an integral curve of X; and since it is defined for all time, it is maximal.

Conversely, suppose $X \in \mathscr{X}^L(G)$, and let α be its maximal integral curve starting at e. By Lemma 13.1, α is defined on all of $\mathbb{R}$. Since X is left-invariant, it is L_g-related to itself, and as in the proof of the afore-mentioned lemma, by naturality of flows, it follows that $L_g \circ \alpha$ is also an integral curve of X. Apply this with $g = \alpha(s)$ for some $s \in \mathbb{R}$; then $t \mapsto \alpha(s)\alpha(t)$ is an integral curve of X starting at $\alpha(s)$. But the Translation Lemma 5.7, $t \mapsto \alpha(t+s)$ is also an integral curve starting at $\alpha(s)$, and so it follows that $\alpha(s+t) = \alpha(s)\alpha(t)$. Together with the fact that $\alpha(0) = e$, this shows α is indeed a one-parameter subgroup. $\square$

Given $X \in \mathrm{Lie}(G)$, we denote the corresponding integral curve starting at e as the **one-parameter subgroup generated by** X. This gives us a correspondence

$$T_eG \cong \mathrm{Lie}(G) \cong \{\text{one-parameter subgroups of } G\}.$$

EXAMPLE 13.4. Let $G = \mathrm{GL}(n, \mathbb{F})$ where $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The collection of all Lie group homomorphisms $\alpha\colon \mathbb{R} \to G$ is well-known in this case. Note that, since G is an open submanifold of the vector space $\mathbb{M}_n(\mathbb{F})$, we have global coordinates (the matrix entries) to work with, so we can differentiate in the classical way. Thus, holding t fixed, and differentiating with respect to s, the group homomorphism property $\alpha(s+t) = \alpha(s)\alpha(t)$ yields $\dot\alpha(s+t) = \dot\alpha(s)\alpha(t)$. Now taking $s = 0$, this gives the ODE

$$\dot\alpha(t) = \dot\alpha(0)\alpha(t).$$

By the uniqueness of solutions to ODEs, we need only find *a* solution to this equation with $\alpha(0) = I_n$, then we know it is *the* solution. As with the case $n = 1$, this is just given by the exponential map: for $A \in \mathbb{M}_n(\mathbb{F})$, we define as usual

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

Note that $\mathbb{M}_n(\mathbb{F})$ is a Banach space under the operator norm, and since $\|AB\| \leq \|A\|\|B\|$, it follows that

$$\left\| \sum_{k=\ell}^{m} \frac{1}{k!} A^k \right\| \leq \sum_{k=\ell}^{m} \frac{1}{k!} \|A\|^k.$$

This is the ℓ-up-to-m segment of the infinite series for the real number $e^{\|A\|}$, and since that series converges, the right-hand-side tends to 0 as $\ell, m \to \infty$. Thus, the series of partial sums is Cauchy, and since $\mathbb{M}_n(\mathbb{F})$ is complete, the series converges.

What's more, let $\epsilon(t) = e^{tA}$. That is

$$\epsilon(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k.$$

The same argument above shows that this power-series converges uniformly on compact t-sets, and so it follows that it is differentiable and its derivative is given term-by-term:

$$\dot\epsilon(t) = \sum_{k=1}^{\infty} \frac{kt^{k-1}}{k!} A^k = A \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} A^{k-1} = A \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} A^\ell = Ae^{tA}.$$

In particular, this means that the unique solution of $\dot\alpha(t) = \dot\alpha(0)\alpha(t)$ with $\alpha(0) = I_n$ and $\dot\alpha(0) = A$ is $\alpha(t) = e^{tA}$. We now must verify that this indeed is a one-parameter group (we only checked above that *if* there is a one-parameter group *then* it must satisfy this ODE).

First note that $\epsilon(0) = I_n$ by definition. Also, the proof that $\epsilon(s+t) = \epsilon(s)\epsilon(t)$ is exactly the same as the standard power-series proof of this fact for $n = 1$. In particular, this means that $I_n = \epsilon(t-t) = \epsilon(t)\epsilon(-t)$, which shows that $\epsilon(t)$ is invertible, with inverse $\epsilon(-t)$. Thus, ϵ is a one-parameter subgroup. Moreover, we have therefore characterized all one-parameter subgroups of $\mathrm{GL}(n, \mathbb{F})$: they are all functions of the form $\epsilon(t) = e^{tA}$ for some $A \in \mathbb{M}_n(\mathbb{F})$.

Finally, note that invertible matrices form a dense open subset of $\mathbb{M}_n(\mathbb{F})$, so the tangent space $T_e\mathrm{GL}(n, \mathbb{F})$ is (up to the usual inclusion identification) equal to $\mathbb{M}_n(\mathbb{F})$. Thus, for any left-invariant vector field $X \in \mathrm{Lie}(\mathrm{GL}(n, \mathbb{F}))$, its initial vector X_e is a matrix in $\mathbb{M}_n(\mathbb{F})$. Thus, the one-parameter group generated by X is simply $\alpha(t) = e^{tX_e}$.

Example 13.4 actually shows us what all one-parameter subgroups look like in all linear Lie groups (Lie subgroups of $\mathrm{GL}(n, \mathbb{F})$), due to the following general lemma.

LEMMA 13.5. *Let G be a Lie group, and let $H \subseteq G$ be a Lie subgroup. The one-parameter subgroups of H are precisely those one-parameter subgroups α of G for which $\dot{\alpha}(0) \in T_eH$.*

PROOF. First, suppose $\alpha\colon \mathbb{R} \to H$ is a one-parameter subgroup of H. Then the composition

$$\mathbb{R} \xrightarrow{\alpha} H \hookrightarrow G$$

is a Lie group homomorphism and so is a one-parameter subgroup of G, which clearly satisfies $\dot{\alpha}(0) \in T_eH$.

Conversely, suppose α is a one-parameter subgroup of G which happens to satisfy $\dot{\alpha}(0) \in T_eH$. Let $\widetilde{\alpha}$ be the one-parameter subgroup of H with $\dot{\widetilde{\alpha}}(0) = \dot{\alpha}(0)$. Again, we can compose with the inclusion map into G, and then view $\widetilde{\alpha}$ as a one-parameter subgroup of G; but it has the same initial velocity vector as α, and since both are maximal integral curves of the left-invariant vector field on G which is equal to $\dot{\alpha}(0)$ at e (cf. Theorem 13.3), it follows that they are equal. Thus, α actually takes values in H, and is a one-parameter subgroup of H. □

EXAMPLE 13.6. Consider the Lie group $\mathrm{O}(n, \mathbb{R})$, which is the I_n-level set of the function $F(A) = AA^\top$. By Proposition 9.34, the tangent space $T_{I_n}\mathrm{O}(n, \mathbb{R})$ (viewed now as a subspace of $T_{I_n}\mathbb{M}_n(\mathbb{R}) = \mathbb{M}_n(\mathbb{R})$ since $\mathrm{O}(n, \mathbb{R})$ is an embedded submanifold of $\mathbb{M}_n(\mathbb{R})$) is equal to $\ker(dF_{I_n})$. As we've computed, $dF_{I_n}(H) = H + H^\top$, and so $T_{I_n}\mathrm{O}(n, \mathbb{R}) = \{H \in \mathbb{M}_n(\mathbb{R})\colon H^\top = -H\}$ (the so-called *skew-symmetric matrices*). Thus, by Lemma 13.5, the one-parameter subgroups of $\mathrm{O}(n, \mathbb{R})$ are all of the form $t \mapsto e^{tH}$ for some skew-symmetric H. In particular, this shows that every matrix of the form e^H with $H^\top = -H$ is in fact orthogonal. (This is easy to see directly from the fact that $(e^H)^\top = e^{H^\top}$; thus $(e^H)^{-1} = e^{-H} = e^{H^\top} = (e^H)^\top$.)

2. The Exponential Map

In light of Example 13.4 and Lemma 13.5, at least in linear Lie groups $G \subseteq \mathrm{GL}(n, \mathbb{R})$, all one-parameter groups are of the form $t \mapsto e^{tA}$ for some $A \in T_eG \subseteq \mathbb{M}_n(\mathbb{R})$. We therefore borrow the terminology *exponential map* to refer to one-parameter groups in general.

DEFINITION 13.7. *Let G be a Lie group. Define a map $\exp\colon \mathrm{Lie}(G) \to G$, called the* **exponential map of** *G, as follows: for $X \in \mathrm{Lie}(G)$, let α_X be the one-parameter subgroup generated by X. Then*

$$\exp(X) \equiv \alpha_X(1).$$

That is: $\exp(X)$ is the value of the maximal integral curve of X starting at e at time 1.

It might make sense to denote the exponential map by $\exp_G$ as it in principle depends on G, but this added notation is universally left out. As we will see in the following propositions, it makes good sense to call this an exponential map in general. To begin, we note that, as in the linear case, $\exp$ maps the line $\{tX : t \in \mathbb{R}\}$ in $\mathrm{Lie}(G)$ to the one-parameter subgroup generated by X.

PROPOSITION 13.8. *Let G be a Lie group. For any $X \in \mathrm{Lie}(G)$, $\alpha(s) = \exp sX$ is the one-parameter subgroup generated by X.*

PROOF. Let $\widetilde{\alpha} : \mathbb{R} \to G$ be the one parameter subgroup generated by X, i.e. the maximal integral curve of X starting at e. By the Dilation Lemma 5.6, for any fixed $s \neq 0$ the dilated curve $\widetilde{\alpha}_s(t) = \widetilde{\alpha}(st)$ is the integral curve of sX starting at e, and so by definition

$$\exp sX \equiv \widetilde{\alpha}_s(1) = \alpha(s).$$

□

Let us now summarize and prove all the major results about the exponential map that we will use.

THEOREM 13.9. *Let G be a Lie group. Denote by $\mathrm{Lie}(G)$ the Lie algebra of left-invariant vector fields as usual, and let $\mathfrak{g}$ denote T_eG.*

(a) *The exponential map $\exp : \mathrm{Lie}(G) \to G$ is smooth (giving $\mathrm{Lie}(G)$ the standard smooth structure of a vector space – i.e. identifying it with $\mathfrak{g}$).*
(b) *For $X \in \mathrm{Lie}(G)$ and $s, t \in \mathbb{R}$, $\exp(s+t)X = \exp sX \exp tX$.*
(c) *For any $X \in \mathrm{Lie}(G)$, $(\exp X)^{-1} = \exp(-X)$.*
(d) *For any $X \in \mathrm{Lie}(G)$ and any $n \in \mathbb{Z}$, $(\exp X)^n = \exp(nX)$.*
(e) *Identifying $T_0\mathrm{Lie}(G) \cong \mathfrak{g}$, the differential $d\exp_p : T_0\mathrm{Lie}(G) \to \mathfrak{g}$ is the identity map.*
(f) *The exponential map restricts to a diffeomorphism from some neighborhood of 0 in $\mathrm{Lie}(G)$ onto a neighborhood of e in G.*
(g) *If H is another Lie group and $F : G \to H$ is a Lie group homomorphism, then with $F_* : \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ the corresponding Lie algebra homomorphism, we have $\exp \circ F_* = F \circ \exp$; i.e. the following diagram commutes:*

$$\begin{tikzcd} \mathrm{Lie}(G) \arrow[r, "F_*"] \arrow[d, "\exp"'] & \mathrm{Lie}(H) \arrow[d, "\exp"] \\ G \arrow[r, "F"'] & H \end{tikzcd}$$

(h) *The flow θ^X of $X \in \mathrm{Lie}(G)$ is given by $\theta^X_t = R_{\exp tX}$; i.e. $\theta^X_t(g) = g \exp tX$.*

PROOF. The only part here that is a little tricky is (a). We need to show that $\exp X \equiv \theta_X(1, e)$ depends smoothly on X. This does not follow directly from the Fundamental Theorem on Flows (Theorem 5.15), but we can reduce it to be so-covered with the following trick. Define a vector field Ξ on $G \times \mathrm{Lie}(G)$ as follows:

$$\Xi_{g,X} = (X_g, 0) \in T_gG \oplus T_X\mathrm{Lie}(G) \cong T_{(g,X)}(G \times \mathrm{Lie}(G)).$$

In fact, $\Xi \in \mathscr{X}(G \times \mathrm{Lie}(G))$ (it is smooth), which we can see as follows. Choose any basis $X_1, \ldots, X_n$ for $\mathrm{Lie}(G)$; then we have global coordinates (x^j) for $\mathrm{Lie}(G)$ (where the point (x^j)

specifies the vector field $\sum_{j=1}^n x^j X_j$). Let (y^i) be a coordinate chart for G, and let $f \in C^\infty(G \times \mathrm{Lie}(G))$. Let $\mathbf{x} = (x^1, \dots, x^n)$ and $\mathbf{y} = (y^1, \dots, y^n)$. Then we have

$$(\Xi f)(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^n x^j X_j f(\mathbf{x}, \mathbf{y})$$

where X_j only differentiates f in the $\mathbf{y}$-directions. As X_j are smooth vector fields, this is a smooth function, and so it follows that Ξ is smooth. Let Θ denote the flow of Ξ; it is easy to check that

$$\Theta_t(g, X) = (\theta^X(t, g), X).$$

By Theorem 5.15, Θ is smooth. Thus its component functions are smooth, and this shows that

$$X \mapsto \exp X = \theta^X(1, e) = \pi_1 \circ \Theta_1(e, X)$$

is smooth, as claimed (here $\pi_1 \colon G \times \mathrm{Lie}(G) \to G$ is the projection onto the first factor). This proves (a).

Since $\exp tX = \alpha_X(t)$ where α_X is the one parameter group generated by X (cf. Proposition 13.8), it follows from the group homomorphism property that $\exp(s+t)X = \alpha_X(s+t) = \alpha_X(s)\alpha_X(t) = \exp sX \exp tX$, proving (b); (c) and (d) follow immediately. For (e), fix $X \in \mathrm{Lie}(G)$, and let $\sigma(t) = tX$ be the straight line curve in $\mathrm{Lie}(G)$; then $\dot\sigma(0) = X$. Thus, again by Proposition 13.8, we have

$$d\exp_0(X) = d\exp_0(\dot\sigma(0)) = \frac{d}{dt} \exp \circ \sigma(t)\bigg|_{t=0} = \frac{d}{dt}\bigg|_{t=0} \exp tX = \dot\alpha_X(0) = X.$$

Part (f) now follows immediately from (e) and the inverse function theorem.

For part (g), we actually show the (nominally) stronger property that $\exp(tF_*X) = F(\exp tX)$ for all $t \in \mathbb{R}$. By Proposition 13.8, $t \mapsto \exp(tF_*X)$ is the one-parameter subgroup generated by F_*X; so it suffices to show that $\beta(t) = F(\exp tX)$ is a Lie algebra homomorphism $\mathbb{R} \to H$ with $\dot\beta(0) = (F_*X)_e$. That it is a Lie group homomorphism follows from (b): β is the composition of $t \mapsto \exp tX$ which is a Lie group homomorphism $\mathbb{R} \to G$, and the Lie group homomorphism $F \colon G \to H$. The initial velocity can be computed thus:

$$\dot\beta(0) = \frac{d}{dt}\bigg|_{t=0} F(\exp tX) = dF_e\left(\frac{d}{dt}\bigg|_{t=0} \exp tX\right) = dF_e(X_e) = (F_*X)_e$$

as claimed.

Finally, for part (h), since X is left invariant, it follows that for any $g \in G$ and any integral curve α of X, $L_g\alpha$ is also an integral curve of X. Thus $t \mapsto L_g \exp tX$ is the integral curve of X starting at g, meaning $L_g \exp tX = \theta^X(t, g)$. Thus

$$R_{\exp tX} g = g \exp tX = L_g \exp tX = \theta^X(t, g) = \theta_t^X(g)$$

as claimed. □

REMARK 13.10. It is very important to realize that Theorem 13.9 does *not* imply that $\exp(X + Y) = \exp(X)\exp(Y)$ for any two vector fields $X, Y \in \mathrm{Lie}(G)$. This generally only happens when $[X, Y] = 0$, as we will soon see.

3. The Adjoint Maps

Let G be a Lie group. For any $g \in G$, we have the associated left-action of G on itself by conjugation:

$$C_g \colon G \to G, \quad C_g(h) = ghg^{-1}.$$

The map C_g is a Lie group homomorphism from G to itself (it is an inner automorphism). It thus induces a Lie algebra homomorphism $(C_g)_* \colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$. This is called the **Adjoint map**:

$$\mathrm{Ad}_g = (C_g)_* \colon \mathrm{Lie}(G) \to \mathrm{Lie}(G).$$

Since C_g is a group isomorphism, Ad_g is a Lie algebra isomorphism. In particular, Ad_g is a linear isomorphism. In other words, for each g, $\mathrm{Ad}_g \in \mathrm{GL}(\mathrm{Lie}(G))$. We canonically identify $\mathrm{Lie}(G)$ with $\mathfrak{g} = T_eG$, and this makes $\mathrm{GL}(\mathrm{Lie}(G))$ into a Lie group in its own right. So we have a map

$$\mathrm{Ad} \colon G \to \mathrm{GL}(\mathrm{Lie}(G))$$

where $\mathrm{Ad}(g) = \mathrm{Ad}_g$. In fact, this is a Lie group homomorphism.

PROPOSITION 13.11. *For any Lie group G, the Adjoint map $\mathrm{Ad} \colon G \to \mathrm{GL}(\mathrm{Lie}(G))$ is a Lie group homomorphism.*

PROOF. First note that $C_{g_1g_2}(h) = g_1g_2h(g_1g_2)^{-1} = g_1g_2hg_2^{-1}g_1^{-1} = C_{g_1} \circ C_{g_2}(h)$; thus $C_{g_1g_2} = C_{g_1} \circ C_{g_2}$. It then follows that

$$\mathrm{Ad}_{g_1g_2} = (C_{g_1g_2})_* = (C_{g_1} \circ C_{g_2})_* = (C_{g_1})_* \circ (C_{g_2})_* = \mathrm{Ad}_{g_1} \circ \mathrm{Ad}_{g_2}.$$

Since $C_e = \mathrm{Id}_G$, we also have $\mathrm{Ad}_e = \mathrm{Id}_{\mathrm{Lie}(G)}$, and hence Ad is a group homomorphism from G into $\mathrm{GL}(\mathrm{Lie}(G))$. (The inverse of Ad_g is, of course, $\mathrm{Ad}_{g^{-1}}$.) Hence, to complete the proof, we need only show that Ad is smooth.

Let $C \colon G \times G \to G$ be the two-variable conjugation map $C(g,h) = ghg^{-1}$ (so $C_g = C(g,\cdot)$). Let $X \in \mathrm{Lie}(G)$ and $g \in G$. By definition, $X = \frac{d}{dt}|_{t=0} \exp tX$, and so we have

$$(\mathrm{Ad}_g X)_e = \left.\frac{d}{dt}\right|_{t=0} C_g(\exp tX) = \left.\frac{d}{dt}\right|_{t=0} C(g, \exp tX) = dC_{g,e}(0, X_e).$$

As C is smooth, this depends smoothly on both g and X (as can readily be seen in local coordinates). Thus, the left-translation of it, which is (by definition) $\mathrm{Ad}_g X$, is also smooth in g and X.

Now, choose a basis $\{X_1, \ldots, X_n\}$ for $\mathrm{Lie}(G)$, and let $\{\Lambda_1, \ldots, \Lambda_n\}$ be the dual basis. Then for any linear operator $A \in \mathrm{End}(\mathrm{Lie}(G))$, its matrix entries in terms of the basis $\{X_1, \ldots, X_n\}$ are $A_i^j = \Lambda^j(AX_i)$. This gives global coordinates on $\mathrm{GL}(\mathrm{Lie}(G))$, and in these coordinates, the above computation shows that the matrix entries of Ad_g are

$$[\mathrm{Ad}_g]_i^j = \Lambda^j(\mathrm{Ad}_g X_i)$$

are smooth functions of g, concluding the proof. □

Thus, since $\mathrm{Ad} \colon G \to \mathrm{GL}(\mathrm{Lie}(G))$ is a Lie group homomorphism, it too induces a Lie algebra homomorphism $\mathrm{Ad}_* \colon \mathrm{Lie}(G) \to \mathrm{Lie}\,(\mathrm{GL}(\mathrm{Lie}(G)))$. The codomain is a complicated looking space, but we will simplify it by invoking the canonical identification of $\mathrm{Lie}(H)$ with T_eH which, in this direction, is just given by $X \mapsto X_e$. Note that the tangent space $T_e\mathrm{GL}(V)$ is precisely the full space of endomorphisms of V, which is usually denoted $\mathfrak{gl}(V)$.

DEFINITION 13.12. *The* **adjoint map** $\mathrm{ad}\colon \mathrm{Lie}(G) \to \mathfrak{gl}(\mathrm{Lie}(G))$ *is the induced Lie algebra homomorphism composed with the evaluation map at the identity:* $\mathrm{ad}(X) = (\mathrm{Ad}_*(X))_e$. *In other words:* $\mathrm{ad}(X)$ *is the left invariant vector field whose value at* e *is* $(d\mathrm{Ad})_e(X_e)$.

Thus, for each $X \in \mathrm{Lie}(G)$, $\mathrm{ad}(X)$ is an endomorphism of $\mathrm{Lie}(G)$: it is a linear map $\mathrm{Lie}(G) \to \mathrm{Lie}(G)$. In fact, it is a familiar one.

THEOREM 13.13. *For any Lie group* G *and any* $X \in \mathrm{Lie}(G)$*, the linear operator* $\mathrm{ad}(X)$ *on* $\mathrm{Lie}(G)$ *is given by*

$$\mathrm{ad}(X)Y = [X, Y].$$

PROOF. Since $\frac{d}{dt}|_{t=0} \exp tX = X_e$, we have

$$(d\mathrm{Ad})_e(X_e)Y = \left(\frac{d}{dt}\bigg|_{t=0} \mathrm{Ad}(\exp tX)\right) Y = \frac{d}{dt}\bigg|_{t=0} (\mathrm{Ad}(\exp tX)Y)\,. \tag{13.1}$$

Now, $\mathrm{Ad}(g) = (C_g)_*$, and $C_g(h) = ghg^{-1} = R_{g^{-1}} \circ L_g(h)$, thus $\mathrm{Ad}(g) = (R_{g^{-1}})_* \circ (L_g)_*$. Applying this with $g = \exp tX$ (and so $g^{-1} = \exp(-tX)$) gives, at the identity,

$$(\mathrm{Ad}(\exp tX)Y)_e = d(R_{\exp(-tX)}) \circ d(L_{\exp tX})(Y_e) = d(R_{\exp(-tX)})(Y_{\exp tX})$$

where the last equality follows from the left-invariance of Y. Now, as shown in Theorem 13.9(h), the flow θ^X of X is given by $\theta^X_t = R_{\exp tX}$, and so we have

$$(\mathrm{Ad}(\exp tX)Y)_e = d(\theta^X_{-t})(Y_{\theta^X_t(e)}).$$

Thus,

$$(\mathrm{ad}(X)Y)_e = \frac{d}{dt}\bigg|_{t=0} d(\theta^X_{-t})(Y_{\theta^X_t(e)}) = (\mathscr{L}_X(Y))_e,$$

i.e. it is the Lie derivative, cf. Definition 5.22. The result now follows from Proposition 5.24: $(\mathrm{ad}(X)Y)_e = [X, Y]_e$, and since both $\mathrm{ad}(X)Y$ and $[X, Y]$ are left-invariant, this concludes the proof. □

So, what have we gained? We have recast the bracket on $\mathrm{Lie}(G)$ in terms of a complicated second derivative $\mathrm{ad}\colon \mathrm{Lie}(G) \to \mathfrak{gl}(\mathrm{Lie}(G))$. That might seem counterproductive; but, in fact, it allows us to finally see what the bracket really looks like when though of as an operation not on $\mathrm{Lie}(G)$ but on T_eG – at least in the case of a matrix group G.

THEOREM 13.14. *Let* $G \subseteq \mathrm{GL}(n, \mathbb{R})$ *be a matrix Lie group. Then* $\mathfrak{g} = T_eG$ *is a subspace of* $\mathbb{M}_n(\mathbb{R})$. *We identify* $\mathfrak{g} \to \mathrm{Lie}(G)$ *in the usual way* $A \mapsto \widetilde{A}$. *Then for* $A, B \in \mathfrak{g}$,

$$[\widetilde{A}, \widetilde{B}]_{\mathrm{Lie}(G)} = \widetilde{[A, B]}_{\mathfrak{gl}(n)}.$$

That is: the Lie bracket on the tangent space $\mathfrak{g}$ *is the commutator bracket of matrices.*

REMARK 13.15. The same thing holds for complex matrix Lie groups, since we can view $\mathrm{GL}(n, \mathbb{C})$ as a Lie subgroup of $\mathrm{GL}(2n, \mathbb{R})$, cf. Example 11.16(2).

PROOF. First note that it suffices to prove the theorem for the group $\mathrm{GL}(n, \mathbb{R})$ itself, by Corollary 12.47. We now proceed to compute the Ad and ad maps directly on $\mathrm{GL}(n, \mathbb{R})$. Fix $g \in \mathrm{GL}(n, \mathbb{R})$. Then $C_g\colon \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ is the restriction of the same map $C_g\colon \mathbb{M}_n(\mathbb{R}) \to \mathbb{M}_n(\mathbb{R})$ to the open subset $\mathrm{GL}(n, \mathbb{R})$. Note that the map $C_g(A) = gAg^{-1}$ is *linear* in $A \in \mathbb{M}_n(\mathbb{R})$. Consequently, the differential of C_g is the map itself:

$$(dC_g)_e(A) = gAg^{-1}.$$

Thus, $\mathrm{Ad}(g)$ is the linear operator on $\mathrm{Lie}(\mathrm{GL}(n,\mathbb{R}))$ given by

$$\mathrm{Ad}_g(\widetilde{A}) = \widetilde{gAg^{-1}}.$$

Now, we use (13.1). Let $X = \widetilde{A}$ and $Y = \widetilde{B}$ for $A, B \in \mathfrak{gl}(n)$. Then $\exp(tX) = e^{tA}$, and so We have

$$\mathrm{Ad}(\exp tX)Y = \mathrm{Ad}(e^{tA}))\widetilde{B} = \widetilde{e^{tA}Be^{-tA}}$$

and so

$$(\mathrm{ad}(X)Y)_e = \frac{d}{dt}\bigg|_{t=0} e^{tA}Be^{-tA}.$$

We now use regular calculus: applying the product rule,

$$\frac{d}{dt}\left(e^{tA}Be^{-tA}\right) = \left(\frac{d}{dt}e^{tA}\right)Be^{-tA} + e^{tA}B\left(\frac{d}{dt}e^{-tA}\right) = Ae^{tA}Be^{-tA} + e^{tA}B(-A)e^{-tA}.$$

Setting $t = 0$ yields $AB - BA$, and so we have

$$(\mathrm{ad}(X)Y)_e = AB - BA.$$

Combining this with Theorem 13.13, we have $[\widetilde{A}, \widetilde{B}]_e = AB - BA$, and using left-invariance of $[\widetilde{A}, \widetilde{B}]$ yields the result. □

Henceforth, especially when dealing with matrix Lie groups, we will de-emphasize the distinction between the Lie algebra $\mathrm{Lie}(G)$ and the tangent space $T_eG = \mathfrak{g}$ at the identity; the Lie bracket is easy to understand on $\mathfrak{g}$, as it is always just the commutator bracket of matrices. Hence, we may be imprecise and refer to $\mathfrak{g}$ as the Lie algebra of G

Before we continue, let's point out a few properties of the Ad and ad maps that were hidden in the preceding proofs.

COROLLARY 13.16. *Let G be a Lie group, and let $X \in \mathrm{Lie}(G)$. Then* $\mathrm{Ad}(\exp X) = \exp(\widetilde{\mathrm{ad}X})$.

PROOF. Since $\widetilde{\mathrm{ad}(X)} = \mathrm{Ad}_*(X)$, this follows immediately from Theorem 13.9(g). □

COROLLARY 13.17. *Let $G \subseteq \mathrm{GL}(n,\mathbb{R})$ be a matrix Lie group, and let $\mathfrak{g} \subseteq \mathbb{M}_n(\mathbb{R})$ be its Lie algebra. Then for any $A \in \mathfrak{g}$ and any $g \in G$, $gAg^{-1} \in \mathfrak{g}$*

PROOF. We computed above that $gAg^{-1} = (dC_g)_e(A)$ which is, by definition, a vector in $T_{C_g(e)}G = T_eG = \mathfrak{g}$. □

EXAMPLE 13.18. Let $G = \mathrm{SL}(n,\mathbb{R})$. This is the 1-level set of the Lie group homomorphism $\det\colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}^*$, and so its Lie algebra / tangent space at the identity $\mathfrak{sl}(n,\mathbb{R})$ is $\ker((d\det)_e)$. As we computed in Example 11.5(3), $(d\det)_e(A) = \mathrm{Tr}\,A$, and so

$$\mathfrak{sl}(n,\mathbb{R}) = \{A \in \mathbb{M}_n(\mathbb{R})\colon \mathrm{Tr}\,A = 0\}.$$

Now, if $g \in G$ (in fact if g is any invertible matrix whatsoever), and $A \in \mathfrak{sl}(n,\mathbb{R})$, we have

$$\mathrm{Tr}\,(gAg^{-1}) = \mathrm{Tr}\,(g^{-1}gA) = \mathrm{Tr}\,A = 0$$

and so indeed $gAg^{-1} \in \mathfrak{sl}(n,\mathbb{R})$. Also, note that $\mathrm{Tr}\,[A,B] = \mathrm{Tr}\,(AB - BA) = \mathrm{Tr}\,(AB) - \mathrm{Tr}\,(BA) = 0$ by the trace property, for *any* matrices A, B; so, indeed, $\mathfrak{sl}(n,\mathbb{R})$ is closed under the commutator bracket, making into a Lie algebra; as we saw above, it is *the* Lie algebra of $\mathrm{SL}(n,\mathbb{R})$.

EXAMPLE 13.19. Let $G = \mathrm{O}(n, \mathbb{R}) = \mathrm{O}(n)$. Then the Lie algebra / tangent space at the identity is $\mathfrak{o}(n) = \{A \in \mathbb{M}_n(\mathbb{R})\colon A + A^\top = 0\}$. We can compute directly that if $A, B \in \mathfrak{o}(n)$, then

$$\begin{aligned}[A,B]^\top = (AB - BA)^\top &= B^\top A^\top - A^\top B^\top \\ &= (-B)(-A) - (-A)(-B) = BA - AB = -(AB - BA) = -[A,B]\end{aligned}$$

so $[A, B] \in \mathfrak{o}(n)$, making $\mathfrak{o}(n)$ into a Lie algebra (and it is *the* Lie algebra of $\mathfrak{O}(n)$). Note that $\mathrm{SO}(n)$ has the same Lie algebra, as it is the identity component of $\mathrm{O}(n)$. Also note that we can check again directly that, for any $Q \in O(n)$ and any $A \in \mathfrak{o}(n)$,

$$(QAQ^{-1})^\top = (QAQ^\top)^\top = QA^\top Q^\top = -QAQ^\top,$$

and so indeed $QAQ^{-1} \in \mathfrak{o}(n)$.

EXAMPLE 13.20. Consider the Lie groups $\mathrm{U}(n)$ and $\mathrm{SU}(n)$. Then the Lie algebra of $\mathrm{U}(n)$ is $\mathfrak{u}(n) = \{A \in \mathbb{M}_n(\mathbb{C})\colon A^* + A = 0\}$, and calculations like those above confirm directly that this is closed under the commutator bracket. Now, the subgroup $\mathrm{SU}(n)$ actually has (real) codimension 1, since the determinant of a generic matrix in $\mathrm{U}(n)$ is a unit modulus complex number, not just ± 1. At the level of Lie algebras, this is reflected in the fact that $\mathrm{SU}(n)$ is the level set of the pair of functions $A \mapsto (AA^*, \det A) = (I, 1)$, and so the tangent space is the kernel of this map which is the intersection of the two kernels: $\mathfrak{su}(n) = \{A^* + A = 0\} \cap \{\operatorname{Tr} A = 0\}$. In the real case, the condition $A^\top + A = 0$ implies that $\operatorname{Tr} A = 0$, but this is not so in the complex case (the diagonal entries of a skew-Hermitian $A \in \mathfrak{u}(n)$ can be any purely imaginary numbers).

4. Normal Subgroups and Ideals

Recall that a subgroup $H \subseteq G$ of a group G is called *normal* if it is closed under conjugation by elements of G: $gHg^{-1} \subseteq H$ for all $g \in G$. These are the "best" subgroups: the kernel of any group homomorphism is a normal subgroup, and conversely, every normal subgroup is the kernel of a group homomorphism: namely the projection $\pi\colon G \to G/H$ (highlighting the fact that normal subgroups are the only ones for which the quotient G/H has a group structure).

There is an analogous concept for Lie algebras; as we will see, the connection is more than nominal in the case of Lie groups.

DEFINITION 13.21. *Let $\mathfrak{g}$ be a Lie algebra. A linear subspace $\mathfrak{I} \subseteq \mathfrak{g}$ is called an* **ideal** *if, for every $X \in \mathfrak{I}$ and $Y \in \mathfrak{g}$, $[X, Y] \in \mathfrak{I}$.*

For example, let $\mathfrak{g} = \mathfrak{gl}(n, \mathbb{F})$ for $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and let $\mathfrak{I} \subset \mathfrak{g}$ be the linear space of matrices X satisfying $\operatorname{Tr} X = 0$ (that is: $\mathfrak{I} = \mathfrak{sl}(n, \mathbb{F})$). If $Y \in \mathfrak{gl}(n, \mathbb{F})$ is any matrix, then $\operatorname{Tr}([X, Y]) = \operatorname{Tr}(XY - YX) = 0$; so $\mathfrak{I}$ is an ideal. Note that any ideal is a Lie subalgebra by definition.

PROPOSITION 13.22. *Let $\mathfrak{g}$ be a Lie algebra. If $\mathfrak{h}$ is another Lie algebra, and $\phi\colon \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism, then $\ker \phi$ is an ideal in $\mathfrak{g}$. Conversely, if $\mathfrak{I}$ is an ideal in $\mathfrak{g}$, then there is a unique Lie algebra structure on the quotient vector space $\mathfrak{g}/\mathfrak{I}$ for which the projection $\pi\colon \mathfrak{g} \to \mathfrak{g}/\mathfrak{I}$ is a Lie algebra homomorphism. Consequently, any ideal is the kernel of a Lie algebra homomorphism (namely π).*

PROOF. Let $X \in \ker \phi$ and $Y \in \mathfrak{g}$. Since ϕ is a Lie algebra homomorphism, we have

$$\phi([X, Y]) = [\phi(X), \phi(Y)] = [0, \phi(Y)] = 0,$$

and so $[X, Y] \in \ker \phi$ as well, proving that $\ker \phi$ is an ideal.

Now, suppose $\mathfrak{I}$ is an ideal. We wish to define a Lie bracket on $\mathfrak{g}/\mathfrak{I}$ so that the projection π is a homomorphism. It is clear that there is a unique way to do this: for $X, Y \in \mathfrak{g}$, we must have $[\pi(X), \pi(Y)] \equiv \pi([X, Y])$. We must check that this is well-defined. Suppose that $\pi(X) = \pi(X')$ and $\pi(Y) = \pi(Y')$. Thus $X' = X + Z$ and $Y' = Y + W$ where $Z, W \in \mathfrak{I}$. Thus

$$[X', Y'] = [X + Z, Y + W] = [X, Y] + [X, W] + [Z, Y] + [Z, W].$$

Since $\mathfrak{I}$ is an ideal, $[X, W] \in \mathfrak{I}$, $[Z, Y] = -[Y, Z] \in \mathfrak{I}$, and $[Z, W] \in \mathfrak{I}$; thus $[X', Y'] - [X, Y] \in \mathfrak{I}$, so $\pi([X', Y']) = \pi([X, Y])$, showing this bracket is well-defined. It is easy to check that it is a Lie bracket.

Finally, by design π is a Lie algebra homomorphism. Since $\ker \pi = \mathfrak{I}$, this shows $\mathfrak{I}$ is the kernel of a Lie algebra homomorphism. □

Now, suppose G is a Lie group and $H \subseteq G$ is a Lie subgroup. If both are connected, we will see that normality of H is equivalent to its Lie algebra being an ideal in $\mathrm{Lie}(G)$. First, we note that, to test normality from the definition, it suffices to do so on group elements in the image of the exponential map (which generates the group, since it is an open neighborhood of the identity in a connected group, cf. Proposition 11.12(c).

LEMMA 13.23. *Let G be a connected Lie group, and let $H \subseteq G$ be a connected Lie subgroup. Then H is normal in G iff for all $X \in \mathrm{Lie}(G)$ and $Y \in \mathrm{Lie}(H)$,*

$$\exp(X) \exp(Y) \exp(-X) \in H. \tag{13.2}$$

PROOF. Since $\exp(-X) = \exp(X)^{-1}$, if H is normal then (13.2) by definition of normality. Conversely, suppose (13.2) holds. Let $V \subseteq \mathrm{Lie}(G)$ be a neighborhood of 0 such that $\exp_G$ is a diffeomorphism from V onto $U = \exp_G(V)$. Now, since H is a Lie subgroup, $\exp_H = \exp_G|_H$ (by Theorem 13.9(g)), and so applying the same argument and shrinking V if necessary, without loss of generality $\exp_H$ is a diffeomorphism from $V \cap \mathrm{Lie}(H)$ onto a neighborhood U_0 of the identity in H. Shrinking V further if necessary, we may assume it is symmetric: $X \in V$ iff $-X \in V$. Then (13.2) says that $ghg^{-1} \in H$ whenever $g \in U$ and $h \in U_0$.

Now, as H is connected, Proposition 11.12(c) shows that U_0 generates H. Thus, any element $h \in H$ may be written in the form $h = h_1 \cdots h_m$ for some $h_1, \ldots, h_m \in U_0$, and so we have, for any $g \in U$,

$$ghg^{-1} = gh_1 \cdots h_m g^{-1} = (gh_1g^{-1}) \cdots (gh_mg^{-1}) \in H.$$

Similarly, as G is connected, Proposition 11.12(c) shows that U generates G. Thus, any $g \in G$ may be written as $g = g_1 \cdots g_k$ with $g_1, \ldots, g_k \in U$. Now, for any $h \in H$, we've seen that $ghg^{-1} \in H$ for $g \in U$; thus, if $g_1, g_2 \in U$ then $g_2hg_2^{-1} \in H$ and therefore $(g_1g_2)h(g_1g_2)^{-1} = g_1(g_2hg_2^{-1})g_1^{-1} \in H$. Proceeding by induction, we see that $ghg^{-1} \in H$ for any $g \in G$, concluding the proof that H is normal. □

Before connecting ideals in $\mathrm{Lie}(G)$ with normal subgroups of G, we need one more lemma: we can use the exponential map to identify a Lie subalgebra (this will be useful in the next section as well).

LEMMA 13.24. *Let G be a Lie group, and let $H \subseteq G$ be a Lie subgroup. With $\mathrm{Lie}(H)$ considered as a subalgebra of $\mathrm{Lie}(G)$ as in Corollary 12.47,*

$$\mathrm{Lie}(H) = \{X \in \mathrm{Lie}(G) \colon \exp tX \in H \textit{ for all } t \in \mathbb{R}\}.$$

PROOF. By the characterization of $\mathrm{Lie}(H) \subseteq \mathrm{Lie}(G)$ in Corollary 12.47, we simply need to establish that

$$\exp tX \in H \text{ for all } t \in \mathbb{R} \quad \Longleftrightarrow \quad X_e \in T_eH.$$

If $X_e \in T_eH$, then $\exp tX \in H$ for all $t \in \mathbb{R}$ by Lemma 13.5. Conversely, if $\exp tX \in H$ for all t, then the map $t \mapsto \exp tX$ is a smooth map $\mathbb{R} \to H$, and thus $X_e = \frac{d}{dt}\big|_{t=0} \exp tX \in T_eH$. $\square$

THEOREM 13.25. *Let G be a Lie connected group, and let H be a connected Lie subgroup. Then H is normal in G if and only if* $\mathrm{Lie}(H)$ *is an ideal in* $\mathrm{Lie}(G)$.

PROOF. For any $g \in G$, recall that $\mathrm{Ad}(g)\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ is the Lie algebra homomorphism induced by the conjugation group homomorphism C_g; thus, by Theorem 13.9(g), the following diagram commutes

$$\begin{tikzcd} \mathrm{Lie}(G) \arrow[r, "\mathrm{Ad}(g)"] \arrow[d, "\exp"'] & \mathrm{Lie}(G) \arrow[d, "\exp"] \\ G \arrow[r, "C_g"'] & G \end{tikzcd}$$

So $\exp \circ \mathrm{Ad}(g) = C_g \circ \exp$. Let $Y \in \mathrm{Lie}(H)$, and apply this with $g = \exp X$ for some $X \in \mathrm{Lie}(G)$; thus

$$\exp\left(\mathrm{Ad}(\exp X)Y\right) = C_{\exp X}(\exp Y) = \exp X \exp Y \exp(-X).$$

On the other hand, by Corollary 13.16, we have

$$\mathrm{Ad}(\exp X)Y = \exp(\widetilde{\mathrm{ad}X})Y$$

Now $\widetilde{\mathrm{ad}X}$ acts on the matrix Lie algebra $\mathrm{Lie}(GL(\mathrm{Lie}(G)))$, and so the exponential map can be identified as the matrix exponential of the evaluation at the identity:

$$(\widetilde{\mathrm{ad}X})Y = e^{\mathrm{ad}X}Y = \sum_{k=0}^{\infty} \frac{1}{k!}(\mathrm{ad}X)^k Y.$$

Now, assume $\mathrm{Lie}(H)$ is an ideal in $\mathrm{Lie}(G)$. Then $[X, Y] \in \mathrm{Lie}(H)$, and therefore $(\mathrm{ad}X)^2 Y = \mathrm{ad}(X)([X, Y]) = [X, [X, Y]] \in \mathrm{Lie}(H)$, and so forth by induction we find that $(\mathrm{ad}X)^k Y \in \mathrm{Lie}(H)$. Thus, we see that

$$\exp X \exp Y \exp(-X) = \mathrm{Ad}(\exp X)Y = \exp(\widetilde{\mathrm{ad}X})Y = \sum_{k=0}^{\infty} \frac{1}{k!}(\mathrm{ad}X)^k Y \in \mathrm{Lie}(H).$$

By Lemma 13.23, it follows that H is normal in G.

Conversely, suppose H is normal in G. Let $X \in \mathrm{Lie}(G)$ and $Y \in \mathrm{Lie}(H)$; then Lemma 13.23 shows that, for any $s, t \in \mathbb{R}$,

$$\exp tX \exp sY \exp(-tX) \in H.$$

Applying the same reasoning as above we can express this in terms of Ad as

$$H \ni \exp(\mathrm{Ad}(\exp tX)sY) = \exp(s\mathrm{Ad}(\exp tX)Y)$$

where we have used the fact that $\mathrm{Ad}(\exp tX)$ is a linear operator on $\mathrm{Lie}(G)$. Hence, the vector field $Z = \mathrm{Ad}(\exp tX)Y$ in $\mathrm{Lie}(G)$ satisfies $\exp sZ \in H$ for all $s \in \mathbb{R}$; it follows from Lemma 13.24 that $Z \in \mathrm{Lie}(H)$. This holds true for all $t \in \mathbb{R}$, and so we also have

$$[X, Y] = \mathrm{ad}(X)Y = \frac{d}{dt}\bigg|_{t=0} \mathrm{Ad}(\exp tX)Y \in \mathrm{Lie}(H).$$

This shows that $\mathrm{Lie}(H)$ is an ideal. $\square$

CHAPTER 14

The Baker-Campbell-Hausdorff Formula

Let $\mathfrak{g}$ be a Lie algebra, and let $X, Y \in \mathfrak{g}$. Unless $[X, Y] = 0$, it is not true that $\exp(X + Y) = \exp(X)\exp(Y)$. This is, however, true "to first order." The Baker-Campbell-Hausdorff formula is a general expression for $\log(\exp(X)\exp(Y))$ that, amazingly, is expressed entirely in terms of the Lie bracket. That is: if G is a Lie group, then at least on a neighborhood of the identity where $\exp$ is a diffeomorphism, the group operation can be expressed *entirely* in terms of the Lie bracket on $\mathrm{Lie}(G)$. This fact has many important consequences that we will explore.

REMARK 14.1. As we'll see below, the work of Baker, Campbell, and Hausdorff in the late 19th and early 20th Century did not give an explicit formula, but rather showed that one exists: that there is a way to express $\log(\exp(X)\exp(Y))$ entirely in terms of brackets, and brackets of brackets, etc. of X, Y. The first *explicit* formula was given by Dynkin in 1947. For this reason, many people call it the Baker-Campbell-Hausdorff-Dynkin formula.

1. First-Order Terms and the Lie-Trotter Product Formula

We begin with the lowest-order terms in the Baker-Campbell-Hausdorff formula.

PROPOSITION 14.2. *Let G be a Lie group. Then for any $X, Y \in \mathrm{Lie}(G)$, there is some $\epsilon > 0$ and a smooth function $Z\colon (-\epsilon, \epsilon) \to \mathrm{Lie}(G)$ so that*

$$\exp(tX)\exp(tY) = \exp(t(X + Y) + t^2 Z(t)), \qquad |t| < \epsilon.$$

PROOF. By Theorem 13.9(f), there are neighborhoods U of 0 in $\mathrm{Lie}(G)$ and V of e in G so that $\exp$ is a diffeomorphism from U onto V. Let $\beta(t) = \exp(tX)\exp(tY)$; then $\beta(0) = e \in V$. Since V is open and β is continuous, there is some $\epsilon > 0$ so that $\beta(t)$ in V for all $|t| < \epsilon$. Thus, we may define $\gamma\colon (-\epsilon, \epsilon) \to \mathrm{Lie}(G)$ by

$$\gamma(t) = \exp^{-1}(\beta(t)) = \exp^{-1}(\exp(tX)\exp(tY)), \quad |t| < \epsilon,$$

and this defines a smooth function. By definition $\gamma(0) = 0$, and

$$\exp(tX)\exp(tY) = \exp\gamma(t), \quad |t| < \epsilon.$$

We now calculate the first-order Taylor expansion of γ. Since it is smooth, Taylor's theorem with integral remainder (Theorem 0.9) shows that $\gamma(t) = 0 + t\gamma'(0) + t^2 Z(t)$ for some smooth function Z, so to conclude the proof we need only compute $\gamma'(0)$. By the inverse function theorem,

$$\gamma'(0) = (d\exp_0)^{-1} \left.\frac{d}{dt}\right|_{t=0} \exp(tX)\exp(tY) = \left.\frac{d}{dt}\right|_{t=0} \exp(tX)\exp(tY)$$

because $d\exp_0 = \mathrm{Id}$ by Theorem 13.9(e). Continuing, using the multiplication map m,

$$\gamma'(0) = \left.\frac{d}{dt}\right|_{t=0} \mathrm{m}(\exp(tX), \exp(tY)) = d\mathrm{m}_{(e,e)}\left(\left.\frac{d}{dt}\right|_{t=0} \exp(tX), \left.\frac{d}{dt}\right|_{t=0} \exp(tY)\right) = d\mathrm{m}_{(e,e)}(X, Y).$$

We have computed that $d\mathrm{m}_{(e,e)}(X, Y) = X + Y$, and this concludes the proof. $\square$

REMARK 14.3. There is nothing sophisticated going on here, just Taylor's theorem. We can easily go further and compute higher-order terms: if you do that up to order 4, you can compute that

$$\gamma(t) = t(X+Y) + \frac{t^2}{2}[X,Y] + \frac{t^3}{12}([X,[X,Y]] + [Y,[Y,X]]) + O(t^4).$$

The amazing fact is that all the terms can be expressed as iterated brackets of X and Y (this is not at all clear from Taylor's theorem). We will address this later in the section.

COROLLARY 14.4. *Let G be a Lie group. For $X, Y \in \mathrm{Lie}(G)$, $t \in \mathbb{R}$,*

$$\lim_{n\to\infty} \left(\left(\exp \frac{t}{n} X \right) \left(\exp \frac{t}{n} Y \right) \right)^n = \exp t(X+Y).$$

This is called the Lie product formula, or the Lie-Trotter product formula. It was proved by Lie (in a less rigorous sense) in the 19th Century; this formula was extended to the context of some unbounded operators on Hilbert space by Trotter in the 1950s.

PROOF. By Proposition 14.2, there is some $\epsilon > 0$ and a smooth function $Z\colon (-\epsilon, \epsilon) \to \mathrm{Lie}(G)$ so that $\exp(sX)\exp(sY) = \exp(s(X+Y) + s^2 Z(s))$ for $|s| < \epsilon$. For any $t \in \mathbb{R}$, for all sufficiently large n, $|\frac{t}{n}| < \epsilon$, and so

$$\left(\exp \frac{t}{n} X \right) \left(\exp \frac{t}{n} Y \right) = \exp \left(\frac{t}{n}(X+Y) + \frac{t^2}{n^2} Z\left(\frac{t}{n}\right) \right).$$

Taking nth powers on both sides and using Theorem 13.9(d), this gives

$$\left(\left(\exp \frac{t}{n} X \right) \left(\exp \frac{t}{n} Y \right) \right)^n = \exp \left(t(X+Y) + \frac{t^2}{n} Z\left(\frac{t}{n}\right) \right).$$

As Z is continuous at 0, as $n \to \infty$, $Z(t/n) \to Z(0)$, and so multiplying by t^2/n sends this term to 0, completing the proof. □

2. The Closed Subgroup Theorem

We have seen (Lemma 11.11 and Proposition 11.12) that the group structure of a Lie subgroup imposes strong geometric requirements: for example, an open subgroup is also closed. The main theorem of this section is the ultimate result of this variety: it turns out that it is impossible for any subgroup of a Lie group to be closed unless it is an embedded Lie subgroup.

THEOREM 14.5 (Closed Subgroup Theorem). *Suppose G is a Lie group, and $H \subseteq G$ is a subgroup that is closed in G. Then H is an embedded Lie subgroup.*

We will prove the theorem in a sequence of lemmas. First, note by Lemma 11.4, it suffices to show that H is an embedded submanifold. To do this, we will find slice charts for H, and to do *this* we will identify what the tangent spaces of of the subgroup are. Using left multiplication, it suffices to find the tangent space at e, and so we begin by determining what subspace of $\mathfrak{g} = \mathrm{Lie}(G)$ is actually the Lie algebra of H.

We are motivated here by Lemma 13.24, which identifies the Lie algebra of a known Lie subgroup in terms of the image of the exponential map. That will be our definition here. Define $\mathfrak{h} \subseteq \mathfrak{g}$ as follows:

$$\mathfrak{h} = \{X \in \mathfrak{g} \colon \exp tX \in H \text{ for all } t \in \mathbb{R}\}.$$

LEMMA 14.6. *The subset $\mathfrak{h} \subseteq \mathrm{Lie}(G)$ is a linear subspace.*

PROOF. It is not clear from the definition that $\mathfrak{h}$ is even a subspace; it follows trivially from the definition that $\mathfrak{h}$ is closed under scalar multiplication, but closure under addition is not so clear. This is where the Lie product formula of Corollary 14.4 comes into play. For any fixed $t \in \mathbb{R}$, if $X, Y \in \mathfrak{h}$ then, by definition, $\exp(\frac{t}{n}X)$ and $\exp(\frac{t}{n}Y)$ are in H for all $n \in \mathbb{Z}$. Thus

$$\left(\left(\exp \frac{t}{n}X\right)\left(\exp \frac{t}{n}Y\right)\right)^n \in H \quad \text{for any } t \in \mathbb{R}.$$

Thus, taking limits, since H is closed, it follows from the Lie product formula that $\exp t(X+Y) \in H$ as well. Thus $X + Y \in \mathfrak{h}$. □

We will eventually show that this subspace $\mathfrak{h}$ is $\mathrm{Lie}(H)$. En route, we first establish that the exponential map (of G) is, when restricted to a neighborhood of 0 in $\mathfrak{h}$, a diffeomorphism onto a neighborhood of e in H. Of course, to say it is a diffeomorphism already implies we have a smooth structure, and so we will use this to help define the smooth structure.

We need one small technical result before we proceed.

LEMMA 14.7. *Let G be a Lie group and let $\mathfrak{g} = \mathrm{Lie}(G)$. Let $\mathfrak{h} \subseteq \mathfrak{g}$ be a linear subspace, and let $\mathfrak{b}$ be a complementary subspace (so $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{b}$). The map $\Phi\colon \mathfrak{g} \to G$ defined by*

$$\Phi\colon \mathfrak{h} \oplus \mathfrak{b} \ni X + Y \mapsto \exp X \exp Y$$

is a diffeomorphism from some neighborhood of 0 in $\mathfrak{g}$ onto its image.

PROOF. First note that Φ is well-defined: we may write it as $\Phi = \widetilde{\Phi} \circ \psi$ where $\psi\colon \mathfrak{g} = \mathfrak{h} \oplus \mathfrak{b} \to \mathfrak{h} \times \mathfrak{b}$ is the linear isomorphism $X + Y \mapsto (X, Y)$, and $\widetilde{\Phi}(X, Y) = \exp X \exp Y$. It therefore suffices to show that $\widetilde{\Phi}$ restricts to a diffeomorphism from some neighborhood of $(0, 0) \in \mathfrak{h} \times \mathfrak{b}$ into its image in G. We can compute that

$$d\widetilde{\Phi}_{(0,0)}(X, Y) = d\mathrm{m}_{(e,e)}(d\exp_0(X), d\exp_0(Y)) = d\mathrm{m}_{(e,e)}(X, Y) = X + Y.$$

The image of $d\widetilde{\Phi}_{(0,0)}$ is in $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{b}$, and so composing with the linear isomorphism ψ^{-1} we have $\psi^{-1} \circ d\widetilde{\Phi}_{(0,0)} = \mathrm{Id}_{\mathfrak{h}\times\mathfrak{b}}$. Thus, $d\Phi_0 = \psi^{-1} \circ d\widetilde{\Phi}_{(0,0)} \circ \psi$ is a linear isomorphism, and the result follows from the inverse function theorem. □

Now, let $U \subseteq \mathfrak{g}$ be a neighborhood of 0 such that $\exp\colon U \to \exp(U) \subseteq G$ is a diffeomorphism. Since $\exp(\mathfrak{h}) \subseteq H$, it follows that $\exp(U \cap \mathfrak{h}) \subseteq (\exp U) \cap H$. In fact, for small enough U, this containment is an equality. (This shows $\exp$ is a homeomorphism from $U \cap \mathfrak{h}$ onto $(\exp U) \cap \mathfrak{h}$.)

PROPOSITION 14.8. *There is an open neighborhood U of $0 \in \mathfrak{g}$ such that $\exp(U \cap \mathfrak{h}) = (\exp U) \cap H$.*

PROOF. As noted above, the forward containment is clear from the definitions. We will prove the reverse containment holds for all small enough U by contradiction. Assume there is no neighborhood U of 0 such that $\exp$ is a diffeomorphism from U to its image and $(\exp U) \cap H \subseteq \exp(U \cap \mathfrak{h})$.

Choose a subspace $\mathfrak{b} \subseteq \mathfrak{g}$ complementary to $\mathfrak{h}$, and define $\Phi(X + Y) = \exp X \exp Y$ as in Lemma 14.7. By that lemma, shrinking if necessary, we can find a neighborhood U_0 of $0 \in \mathfrak{g}$ so that $\exp|_{U_0}$ and $\Phi|_{U_0}$ are diffeomorphisms onto their images.

Now, let $\{U_j\}$ be a sequence of open balls in $\mathfrak{g}$ whose radii converge to 0. Then let $V_j = \exp(U_j)$, and $\widetilde{U}_j = \Phi^{-1}(V_j)$. For all sufficiently large j, $U_j \subset U_0$. Now to our contradiction argument: our assumption implies that for all large j there is some $h_j \in (\exp U_j) \cap H$ such that

$h_j \notin \exp(U_j \cap \mathfrak{h})$. In particular, this means $h_j = \exp Z_j$ for some $Z_j \in U_j$. Now, $\exp(U_j) = \Phi(\widetilde{U}_j)$ by definition, and therefore there is also some pair $(X_j, Y_j) \in \widetilde{U}_j \subseteq \mathfrak{h} \oplus \mathfrak{b}$ such that $h_j = \exp X_j \exp Y_j$. If Y_j were 0, this would mean $\exp Z_j = \exp X_j \in \exp(\mathfrak{h})$; but exp is injective on $U_j \subseteq U_0$, which would mean $X_j = Z_j \in U_j \cap \mathfrak{h}$, contradicting our assumption that $h_j \notin \exp(U_j \cap \mathfrak{h})$. Thus $Y_j \neq 0$. But since $(X_j, X_j) \in \widetilde{U}_j = \Phi^{-1} \circ \exp(U_j)$ and the radius of U_j tends to 0, it follows by continuity that $Y_j \to 0$ as $j \to \infty$. Note also that $\exp X_j \in \exp(\mathfrak{h}) \subseteq H$, and therefore $\exp Y_j = (\exp X_j)^{-1} h_j \in H$ as well.

Now, choose any norm $\|\cdot\|$ on $\mathfrak{b}$, and let $c_j = \|Y_j\|$; then $c_j \to 0$ as $j \to \infty$, but $c_j \neq 0$. Now Y_j/c_j is in the unit sphere in the finite-dimensional normed space $(\mathfrak{b}, \|\cdot\|)$, which is compact, and thus there is a subsequence that converges; reindexing, wlog we have some limit $Y \in \mathfrak{b}$ with $\|Y\| = 1$ so that $Y_j/c_j \to Y$ as $j \to \infty$. In particular, $Y \neq 0$. We will now show that $Y \in \mathfrak{h}$; since $\mathfrak{h}$ and $\mathfrak{b}$ are complementary, this is a contradiction.

To see that $Y \in \mathfrak{h}$, let $t \in \mathbb{R}$, and for each j let $n_j = \lfloor t/c_j \rfloor \in \mathbb{Z}$; in particular, $|n_j - tc_j| \leq 1$. This means that $|n_j c_j - t| \leq c_j \to 0$, and so $n_j c_j \to t$. Thus

$$n_j Y_j = (n_j c_j) \frac{Y_j}{c_j} \to tY$$

and by continuity we have $\exp n_j Y_j \to \exp tY$. Note that $\exp n_j Y_j = (\exp Y_j)^{n_j}$, and since $\exp Y_j \in H$, it follows that $\exp n_j Y_j \in H$. As H is a closed subgroup, the limit $\exp tY = \lim_{j\to\infty} \exp n_j Y_j \in H$ as well. Thus $\exp tY \in H$ for all $t \in \mathbb{R}$, which means $Y \in \mathfrak{h}$. This gives us the desired contradiction, and concludes the proof. □

We are now ready to prove the theorem.

PROOF OF THE CLOSED SUBGROUP THEOREM. Choose a linear isomorphism $E\colon \mathfrak{g} \to \mathbb{R}^n$ such that $E(\mathfrak{h}) = \mathbb{R}^k \times \{0\}$. Let U be a neighborhood as in Proposition 14.8. Then the composite map $\varphi = E \circ \exp^{-1}\colon \exp U \to \mathbb{R}^n$ is a smooth chart for G, and $\varphi\left((\exp U) \cap H\right) = E(U \cap \mathfrak{h})$ is the slice chart obtained by setting the last $n-k$ coordinates in φ to 0. This gives us a slice chart for points near e, but we can translate in the group to do the same at any point: if $h \in H$, L_h is a diffeomorphism from $\exp U$ onto a neighborhood of h in G, and since H is a subgroup, $L_h(H) = H$. Thus

$$L_h\left((\exp U) \cap H\right) = L_h(\exp U) \cap H$$

and so $\varphi \circ L_h^{-1}$ is a slice chart for H in a neighborhood of h. It follows that H is an embedded submanifold, concluding the proof. □

Let us also note that there is a (much simpler) converse: embedded Lie subgroups are always closed.

PROPOSITION 14.9. *Let G br a Lie group, and suppose $H \subseteq G$ is an embedded Lie subgroup. Then H is closed in G.*

PROOF. Let $\overline{H}$ denote the closure of H in G, and let $g \in \overline{H}$. Let (h_n) be a sequence in H with $h_n \to g$. Let U be a neighborhood of e in G that is the domain for a slice chart for H, and shrink U a little to a neighborhood W containing e so that $\overline{W} \subset U$. There exists a neighborhood V of e in G so that, for all $g_1, g_2 \in V$, $g_1 g_2^{-1} \in W$ (this is an assigned homework exercise [**3**, Problem 7-6]). Note that $h_n g^{-1} \to e$, and thus for all large n $h_n g^{-1} \in V$. Then for such large n, m,

$$h_n h_m^{-1} = (h_n g^{-1})(g h_m^{-1}) = (h_n g^{-1})(h_m g^{-1})^{-1} \in W.$$

Sending $n \to \infty$ and using continuity of the multiplication map shows that $gh_m^{-1} = \lim_n h_n h_m^{-1} \in \overline{W} \subset U$. Since U is a slice chart, and $H \cap U$ is a slice, it is closed in U, and therefore $gh_m^{-1} \in H$ for all large m. But $h_m \in H$, and so $g \in H$. Thus H is closed. □

Thus, it is impossible to have a Lie subgroup that is the interior of a manifold with boundary in the ambient Lie group. If a Lie subgroup is not closed, then it is not embedded. For example, the image of the one parameter group $\alpha(t) = (e^{2\pi it}, e^{2\pi iat})$ for $a \in \mathbb{R} \setminus \mathbb{Q}$ is a Lie subgroup of $\mathbb{T}^2$ which is dense, but not all of $\mathbb{T}^2$. Note that the closure is all of $\mathbb{T}^2$, which is of course embedded. That is our final (immediate) corollary to the closed subgroup theorem.

COROLLARY 14.10. *If H is any Lie subgroup of G, then the closure $\overline{H}$ of H in G is an embedded Lie subgroup.*

PROOF. By the closed subgroup theorem, we simply need to show that $\overline{H}$ is a subgroup. But this follows by continuity of the multiplication map: if $g, h \in \overline{H}$, let $(g_n), (h_n)$ be sequences in H with $g_n \to g$ and $h_n \to h$. Then $H \ni g_n h_n \to gh$, and so $gh \in \overline{H}$. Thus, $\overline{H}$ is a subgroup, and it is closed, thus it is an embedded Lie subgroup. □

3. The Heisenberg Group: a Case Study

Recall the Heisenberg Group $\mathrm{H}(n, \mathbb{R})$ of Example 11.5(7), consisting of matrices in $\mathbb{M}_n(\mathbb{R})$ of the form $I + T$ where T is strictly upper-triangular. To be concrete, let us take $n = 3$, so

$$\mathrm{H}(3, \mathbb{R}) = \left\{ \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} : a, b, c \in \mathbb{R} \right\}. \tag{14.1}$$

Since $\mathrm{H}(3, \mathbb{R})$ is an affine subspace of $\mathbb{M}_3(\mathbb{R})$, as an embedded submanifold its tangent space is simply the strictly upper triangular matrices $\mathfrak{h}(3, \mathbb{R})$

$$\mathfrak{h}(3, \mathbb{R}) = \left\{ \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} : u, v, w \in \mathbb{R} \right\}.$$

It follows from Theorem 13.13 that this subspace must be closed under the commutator bracket; it is an easy (and illustrative) calculation to see it directly:

$$\begin{bmatrix} 0 & x_1 & z_1 \\ 0 & 0 & y_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & x_2 & z_2 \\ 0 & 0 & y_2 \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & x_2 & z_2 \\ 0 & 0 & y_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & x_1 & z_1 \\ 0 & 0 & y_1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & x_1 y_2 - x_2 y_1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

and this is certainly in $\mathfrak{h}(3, \mathbb{R})$. We can frame this another way. Notice that the linear space $\mathfrak{h}(3, \mathbb{R})$ is 3-dimensional, spanned by the three vectors $X = E_{12}$, $Y = E_{23}$, and $Z = E_{33}$. Then the statement that $\mathfrak{h}(3, \mathbb{R})$ is closed under the bracket means that all brackets of these three vectors must be linear combinations of them. In fact, the preceding calculation shows that

$$[X, Y] = Z, \quad [X, Z] = [Y, Z] = 0. \tag{14.2}$$

In particular, this shows that, as a Lie algebra, $\mathfrak{h}(3, \mathbb{R})$ is generated by the two-element set $\{X, Y\}$.

Now, consider the exponential map $\exp\colon \mathfrak{h}(3,\mathbb{R}) \to \mathrm{H}(3,\mathbb{R})$. As this is a matrix Lie group, we know that it is simply given by matrix exponentiation, and this is easy to compute from the power series. Indeed, we have

$$\begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 & xy \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix}^3 = 0.$$

Thus, for any $W \in \mathfrak{h}(3,\mathbb{R})$, $W^k = 0$ for $k \geq 3$, and thus

$$\exp W = \sum_{k=0}^{\infty} \frac{1}{k!} W^k = I + W + \frac{1}{2} W^2.$$

Explicitly, this means

$$\exp \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} = I + \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 & 0 & xy \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & x & z + \frac{xy}{2} \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}.$$

From here, we can see explicitly that:

PROPOSITION 14.11. *The exponential map* $\exp\colon \mathfrak{h}(3,\mathbb{R}) \to \mathrm{H}(3,\mathbb{R})$ *is a diffeomorphism.*

PROOF. This is simply a matter of solving equations. Given any element $g \in \mathrm{H}(3,\mathbb{R})$ (given in the form (14.1)), we need to find a unique $W \in \mathfrak{h}(3,\mathbb{R})$ with $\exp W = g$. That is:

$$\begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} = \exp \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & x & z + \frac{xy}{2} \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}.$$

Immediately we see that we must have $x = a$ and $y = b$, which shows that $z = c - \frac{xy}{2} = c - \frac{ab}{2}$. This is a real number, and so indeed there is a unique solution W for each g. What's more, note that the expression for $\exp^{-1}$ is a polynomial in the entries of the matrix, and so it is smooth, concluding the proof. □

Because the form of the exponential map is so simple in this example, we can hope to fully understand all the terms in this Baker-Campbell-Hausdorff formula, and indeed that is the case. We could fool around directly with power series, but since we will use this example to motivate the general case, we want to work directly with Lie algebraic terms. To that end, we note from (14.2) that $\mathfrak{h}(3,\mathbb{R})$ has the property that all second- and higher-order commutators vanish (that is: $[X,[X,Y]] = [X,Z] = 0$, etc.) When this happens, the Baker-Campbell-Hausdorff formula takes a very simple form.

PROPOSITION 14.12. *Let* $X, Y \in \mathbb{M}_n(\mathbb{R})$ *be elements that commute with their commutator:* $[X,[X,Y]] = [Y,[X,Y]] = 0$. *Then*

$$e^X e^Y = e^{X+Y+\frac{1}{2}[X,Y]}.$$

PROOF. We introduce a parameter $t \in \mathbb{R}$ scaling X and Y, so that the desired conclusion becomes

$$e^{tX} e^{tY} = e^{tX+tY+\frac{t^2}{2}[X,Y]}.$$

By assumption, $[X,Y]$ commutes with X and Y, and so it commutes with $X+Y$; thus, the standard power-series approach to exponentials shows that

$$e^{tX+tY+\frac{t^2}{2}[X,Y]} = e^{t(X+Y)}e^{\frac{t^2}{2}[X,Y]}.$$

Hence, the desired property is equivalent to the statement that

$$A(t) \equiv e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]} = e^{t(X+Y)} \equiv B(t).$$

Note that the two functions $A, B\colon \mathbb{R} \to \mathbb{M}_n$ are smooth. Our strategy to prove they are equal is to show that they both satisfy the same ODE with the same initial condition; by Theorem 0.22(2) (the uniqueness part of the Picard-Lindelöf Theorem), it then follows that $A = B$.

We begin with the right-hand-side $B(t)$ which is the one-parameter subgroup generated by $X+Y$; thus, it satisfies the ODE

$$\dot{B}(t) = B(t)(X+Y), \quad B(0) = I.$$

Moving to the left-hand-side, we apply the product rule to find that

$$\dot{A}(t) = e^{tX}Xe^{tY}e^{-\frac{t^2}{2}[X,Y]} + e^{tX}e^{tY}Ye^{-\frac{t^2}{2}[X,Y]} - te^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}[X,Y].$$

Since Y commutes with $[X,Y]$, it also commutes with all powers of $[X,Y]$, and therefore with $e^{-\frac{t^2}{2}[X,Y]}$; hence, the last two terms maybe written as

$$e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}(Y - t[X,Y]) = A(t)(Y - t[X,Y]). \tag{14.3}$$

For the first term, we must commute X past e^{tY}. To do this, we involve the adjoint map, following Corollary 13.16:

$$Xe^{tY} = e^{tY}e^{-tY}Xe^{tY} = e^{tY}\mathrm{Ad}(e^{-tY})X = e^{tY}e^{\mathrm{ad}(-tY)}X = e^{tY}e^{-t\mathrm{ad}(Y)}X.$$

Recall that $\mathrm{ad}(Y)X = [Y,X]$. Since $[Y,[Y,X]] = -[Y,[X,Y]] = 0$, we therefore have $\mathrm{ad}(Y)^2X = 0$, and so $\mathrm{ad}(Y)^kX = 0$ for $k \geq 2$. Thus, expanding the power-series,

$$e^{-t\mathrm{ad}(Y)}X = X - t\mathrm{ad}(Y)X = X - t[Y,X] = X + t[X,Y].$$

So, finally, the first term in $\dot{A}(t)$ is

$$e^{tX}e^{tY}(X + t[X,Y])e^{-\frac{t^2}{2}[X,Y]} = e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}(X + t[X,Y]) = A(t)(X + t[X,Y]) \tag{14.4}$$

where the last equalities result from commuting $X + t[X,Y]$ past $e^{-\frac{t^2}{2}[X,Y]}$, valid since X and $[X,Y]$ both commute with $[X,Y]$. Combining (14.3) and (14.4) yields

$$\dot{A}(t) = A(t)(X+Y)$$

which is the same ODE satisfied by B. Since the initial condition is again $A(0) = I$, we now conclude that $A(t) = B(t)$ for all $t \in \mathbb{R}$. □

Thus, since any two vectors in $\mathfrak{h}(3,\mathbb{R})$ satisfy the conditions of Proposition 14.12, it follows that the Baker-Campbell-Hausdorff formula for the Heisenberg group can be stated thus.

COROLLARY 14.13 (the Baker-Campbell-Hausdorff Formula for $\mathrm{H}(3,\mathbb{R})$). *For any* $X, Y \in \mathfrak{h}(3,\mathbb{R})$*, the exponential map* $\exp\colon \mathrm{h}(3,\mathbb{R}) \to \mathrm{H}(3,\mathbb{R})$ *satisfies*

$$\exp tX \exp tY = \exp\left(t(X+Y) + \frac{t^2}{2}[X,Y]\right), \qquad t \in \mathbb{R}.$$

REMARK 14.14. Note: the X, Y in this corollary do not refer only to the basis elements $X = E_{12}$ and $Y = E_{23}$ in $\mathfrak{h}(3, \mathbb{R})$ as in (14.2), but rather to any two elements.

This points the way to how we will prove the Baker-Campbell-Hausdorff formula in general: by solving ODEs. To that end, we will need a formula for the differential of the exponential map; as noted above, it will involve that adjoint map. That is the goal of the next section.

Before we get there, we will complete the story of the Heisenberg group's connection to its Lie algebra by proving the Lie correspondence in this example. Recall from Proposition 12.42 that, if H, G are two Lie groups and $F\colon H \to G$ is a Lie group homomorphism, it induces a Lie algebra homomorphism $F_*\colon \mathrm{Lie}(H) \to \mathrm{Lie}(G)$. If we canonically identify $\mathrm{Lie}(H) \cong T_eH = \mathfrak{h}$ and $\mathrm{Lie}(G) \cong T_eG = \mathfrak{g}$, then the induced map F_* becomes $dF_e\colon \mathfrak{h} \to \mathfrak{g}$.

The question is: can we reverse this process? Given a Lie *algebra* homomorphism $\phi\colon \mathfrak{h} \to \mathfrak{g}$, is there a (unique) Lie group homomorphism F so that $dF_e = \phi$? (Spoiler alert: the answer, in general, is no.) To get a sense of what this F might look like, we refer back to Theorem 13.9(g), which asserts (in this language) that $\exp_G \circ dF_e = F \circ \exp_H$. This, of course, does not uniquely specift F in terms of dF_e, except in the special case when $\exp_H$ happens to be invertible.

THEOREM 14.15. *Let G be a matrix Lie group, with Lie algebra $\mathfrak{g}$, and let $\phi\colon \mathfrak{h}(3, \mathbb{R}) \to \mathfrak{g}$ be a Lie algebra homomorphism. Then $F = \exp_G \circ \phi \circ \exp^{-1}_{\mathrm{H}(3,\mathbb{R})}$ is the unique Lie group homomorphism $F\colon \mathrm{H}(3, \mathbb{R}) \to G$ such that $dF_e = \phi$.*

Thus, there is an exact correspondence between Lie *group* homomorphisms from $\mathrm{H}(3, \mathbb{R})$ to some matrix Lie group, and Lie *algebra* homomorphisms from $\mathfrak{h}(3, \mathbb{R})$ to its Lie algebra.

PROOF. By Proposition 14.11, $\exp_{\mathrm{H}(3,\mathbb{R})}$ is a diffeomorphism, so F is well-defined and smooth. Since $d\exp_e = \mathrm{Id}$, it is easy to check that $dF_e = \phi$. So what remains to see is that F is indeed a group homomorphism. This is where the BCH formula comes into play. Fix $g, h \in \mathrm{H}(3, \mathbb{R})$; then there are unique $X, Y \in \mathfrak{h}(3, \mathbb{R})$ so that $g = \exp X$ and $h = \exp Y$. Thus

$$F(gh) = F(\exp X \exp Y) = F(\exp(X + Y + \tfrac{1}{2}[X, Y])) = \exp(\phi(X + Y + \tfrac{1}{2}[X, Y]))$$

by Corollary 14.13 and the definition of F. Now ϕ is a Lie algebra homomorphism, and so

$$\phi(X + Y + \tfrac{1}{2}[X, Y]) = \phi(X) + \phi(Y) + \tfrac{1}{2}[\phi(X), \phi(Y)]).$$

Now, since $X, Y \in \mathfrak{h}(3, \mathbb{R})$, they both commute with their commutators. Since ϕ is a Lie algebra homomorphism, it follows that $\phi(X)$ and $\phi(Y)$ also commute with *their* commutators. Thus, by Proposition 14.12, it follows that

$$\exp(\phi(X)+\phi(Y)+\tfrac{1}{2}[\phi(X), \phi(Y)]) = \exp(\phi(X)) \exp(\phi(Y)) = F(\exp X)F(\exp Y) = F(g)F(h).$$

This shows that F is a group homomorphism, completing the proof. □

The point is: the BCH formula expresses $\exp^{-1}(\exp X \exp Y)$ entirely in terms of Lie algebraic operations, and so we may pass ϕ through to give a Lie group homomorphism, modulo the well-definedness of $\exp^{-1}$. That is the general way in which the BCH formula will be used; the question of how to make sense of $\exp^{-1}$ in general is a topology question we will tackle after we have proved the full BCH formula.

4. The Differential of the Exponential Map

We have already seen, and used repeatedly, the fact that $d\exp_0 = \mathrm{Id}$. We will presently need the differential of the exponential map at an arbitrary point of the Lie algebra, not just 0.

THEOREM 14.16. *Let G be a Lie group, and let $\exp\colon \mathrm{Lie}(G) \to G$ be the exponential map. Then for any $X \in \mathrm{Lie}(G)$,*

$$d\exp_X = (L_{\exp X})_* \int_0^1 e^{-t\,\mathrm{ad}X}\,dt. \tag{14.5}$$

To be clear: $\mathrm{ad}X\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ is the linear operator $\mathrm{ad}X(Y) = [X, Y]$; we may take the usual matrix exponential of this operator, and moreover (as we are in finite dimensions) there is nothing fancy about integrating an operator-valued function. To be clear, for any $Y \in \mathrm{Lie}(G)$,

$$\left(\int_0^1 e^{-t\,\mathrm{ad}X} dt\right)(Y) \equiv \int_0^1 e^{-t\,\mathrm{ad}X} Y\,dt.$$

This is then followed by the linear operator $(L_{\exp X})_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$. If we are identifying $\mathrm{Lie}(G)$ with T_eG as usual, then this should read $dL_{\exp X}|_e$ instead of $(L_{\exp X})_*$.

Before we prove the theorem (preceded by a few lemmas needed in the calculation), let's first expand the integral as a power series instead. In general, if A is any linear operator on a finite dimensional vector space, we can integrate the exponential power series term-by-term, and we get

$$\int_0^1 e^{-tA}\,dt = \sum_{k=0}^{\infty} \frac{(-A)^k}{k!} \int_0^1 t^k\,dt = \sum_{k=0}^{\infty} (-1)^k \frac{A^k}{(k+1)!}.$$

We might choose to write the function of $\mathrm{ad}X$ in (14.5) in this way instead of as the integral; they are equivalent. Another somewhat incorrect way that it is often written is as follows. Suppose that $a \in \mathbb{C}^*$ is a nonzero (i.e. invertible) complex number. Then

$$\int_0^1 e^{-ta}\,dt = \frac{1 - e^{-a}}{a}.$$

It then follows using functional calculus that, if A is an invertible operator,

$$\int_0^1 e^{-tA}\,dt = A^{-1}(I - e^{-A}) = \frac{I - e^{-A}}{A}.$$

This only makes sense if A is invertible, but the above formulas (as a power series or integral) are often used to give meaning to this expression even when A is not invertible:

$$A^{-1}(I - e^{-A}) \equiv \sum_{k=0}^{\infty} (-1)^k \frac{A^k}{(k+1)!} = \int_0^1 e^{-tA}\,dt.$$

In particular, (14.5) is often written in the form

$$d\exp_X = (\exp X)_* \frac{I - e^{-\mathrm{ad}X}}{\mathrm{ad}X}.$$

This does not make literal sense, since $\mathrm{ad}X$ is *never* invertible: $\ker(\mathrm{ad}X) \ni X$ which is nonzero unless $X = 0$ in which case $\mathrm{ad}X = \mathrm{ad}0 = 0$ is certainly not invertible. Nevertheless, this is the way you will often see it written.

The following limit will be needed in the computation of the differential of $\exp$.

LEMMA 14.17. *Let A be a linear operator on a finite dimensional vector space V. Then*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{\ell=0}^{m-1} \left(e^{-A/m}\right)^{\ell} = \int_0^1 e^{-tA}\, dt.$$

PROOF. Rewriting the left-hand side as the limit of

$$\sum_{\ell=0}^{m-1} e^{-\frac{\ell}{m}A} \cdot \frac{1}{m}$$

we recognize this as a Riemann sum for the stated integal, with partition $\{0, \frac{1}{m}, \frac{2}{m}, \cdots, 1\}$ of constant width $\frac{1}{m}$, and left-end-point evaluations. Thus, the result is merely the statement that $t \mapsto e^{-tA}$ is Riemann integrable on $[0, 1]$ (which is immediate from its continuity).

□

We will also need a generalization of the iterated product rule.

LEMMA 14.18. *Let M be a manifold and let G be a Lie group. Let $F_1, F_2\colon M \to G$ be smooth maps, and define $F(p) = F_1(p) \cdot F_2(p)$ to be the product. Then*

$$dF_p = dL_{F_1(p)}|_{F_2(p)} \circ dF_2|_p + dR_{F_2(p)}|_{F_1(p)} \circ dF_1|_p.$$

PROOF. We write F as $F = \mathrm{m} \circ (F_1, F_2)$; then by the chain rule

$$dF_p(X) = d\mathrm{m}_{(F_1(p),F_2(p))} \circ d(F_1, F_2)_p(X) = d\mathrm{m}_{(F_1(p),F_2(p))}(dF_1|_p X, dF_2|_p X)$$

Now, in the proof of Lemma 11.2, we saw that $d\mathrm{m}_{g,h}(Y, Z) = dR_h|_g(Y) + dL_g|_h(Z)$; in the present case, this means

$$dF_p(X) = dR_{F_2(p)}|_{F_1(p)} \circ dF_1|_p X + dL_{F_1(p)}|_{F_2(p)} \circ dF_2|_p X.$$

□

COROLLARY 14.19. *Let M be a manifold and let G be a Lie group. Let $n \geq 2$ be an integer, let $F_1, \ldots, F_n\colon M \to G$ be smooth maps, and define $F(p) = F_1(p) \cdots F_n(p)$ to be the product. Then*

$$dF_p = \sum_{k=1}^{n} dL_{F_1(p)\cdots F_{k-1}(p)}|_{F_k(p)\cdots F_n(p)} \circ dR_{F_{k+1}(p)\cdots F_n(p)}|_{F_k(p)} \circ dF_k|_p. \tag{14.6}$$

It may be easier to read what (14.6) says if we write it in terms of the tangent map without evaluation at basepoints explicit:

$$dF = \sum_{k=1}^{n} dL_{F_1\cdots F_{k-1}} \circ dR_{F_{k+1}\cdots F_n} \circ dF_k.$$

Also, since $L_{gh} = L_g \circ L_h$ while $R_{gh} = R_h \circ R_g$, we can apply the chain rule and express this instead as

$$dF = \sum_{k=1}^{n} dL_{F_1} \circ \cdots \circ dL_{F_{k-1}} \circ dR_{F_n} \circ \cdots \circ dR_{F_{k+1}} \circ dF_k.$$

This is the generalization of the iterated product rule for functions taking values in a real algebra (such as $\mathbb{M}_n(\mathbb{R})$), where

$$d(F_1 \cdots F_n) = \sum_{k=1}^{n} F_1 \cdots F_{k-1} \circ dF_k \circ F_{k+1} \cdots F_n.$$

Indeed, in the case that G is a matrix Lie group, since $L_g\colon G \to G$ is the restriction of the linear map $A \mapsto gA$ on $\mathbb{M}_n(\mathbb{R})$, its derivative is given by $dL_g(X) = gX$ (at any point), and so (14.6) matches this more elementary product rule. The proof of Corollary 14.19 is a simple induction on Lemma 14.18, and is left to the (bored) reader.

We now stand ready to prove Theorem 14.16.

PROOF OF THEOREM 14.16. Fix $X, Y \in \mathrm{Lie}(G)$. We will use the linear curve $\alpha(t) = X + tY$ which passes through X at $t = 0$ and satisfies $\dot\alpha(0) = Y$. Then

$$d\exp_X(Y) = \frac{d}{dt}\bigg|_{t=0} \exp\alpha(t) = \frac{d}{dt}\bigg|_{t=0} \exp(X + tY).$$

Now, for any $m \in \mathbb{N}$,

$$\exp(X + tY) = \left(\exp\left(\frac{X}{m} + t\frac{Y}{m}\right)\right)^m.$$

We now apply the iterated product rule of Corollary 14.19, which yields that

$$\begin{aligned}&\frac{d}{dt}\bigg|_{t=0} \left(\exp\left(\frac{X}{m} + t\frac{Y}{m}\right)\right)^m \\ &= \sum_{k=1}^{m} dL_{\exp(X/m+tY/m)^{k-1}}\big|_{t=0} \circ dR_{\exp(X/m+tY/m)^{m-k}}\big|_{t=0} \left(\frac{d}{dt}\bigg|_{t=0} \exp\left(\frac{X}{m} + t\frac{Y}{m}\right)\right).\end{aligned}$$

We now reindex $\ell = m - k$, giving

$$\sum_{\ell=0}^{m-1} dL_{\exp(X/m)^{m-\ell-1}} \circ dR_{\exp(X/m)^{\ell}} \left(\frac{d}{dt}\bigg|_{t=0} \exp\left(\frac{X}{m} + t\frac{Y}{m}\right)\right).$$

From the chain rule, $dL_{\exp(X/m)^{m-\ell-1}=dL_{\exp(X/m)^{m-1}}\circ dL_{\exp(X/m)^{-\ell}}}$. We also use the linearity of the inside derivative in Y/m to give

$$dL_{\exp(X/m)^{m-1}} \left[\frac{1}{m}\sum_{\ell=0}^{m-1} dL_{\exp(X/m)^{-\ell}} \circ dR_{\exp(X/m)^{\ell}} \left(\frac{d}{dt}\bigg|_{t=0} \exp\left(\frac{X}{m} + tY\right)\right)\right].$$

Now, set $g = \exp(X/m)^{-\ell}$; then $dL_g \circ dR_{g^{-1}} = dC_g$ and (under the standard identification of $\mathrm{Lie}(G)$ with T_eG) this is the definition of $\mathrm{Ad}(g)$. Also using the fact that Ad is a homomorphism, this gives

$$d\exp_X(Y) = dL_{\exp(X/m)^{m-1}} \circ \frac{1}{m}\sum_{\ell=0}^{m-1} \mathrm{Ad}(\exp(-X/m))^{\ell} \left(\frac{d}{dt}\bigg|_{t=0} \exp\left(\frac{X}{m} + tY\right)\right).$$

But

$$\mathrm{Ad}(\exp(-X/m)) = e^{\mathrm{ad}(-X/m)} = e^{-\frac{\mathrm{ad}X}{m}}.$$

So, all together, we have for any $m \in \mathbb{N}$

$$d\exp_X(Y) = dL_{\exp\frac{m-1}{m}X} \circ \frac{1}{m}\sum_{\ell=0}^{m-1} \left(e^{-\frac{\mathrm{ad}X}{m}}\right)^{\ell} \left(\frac{d}{dt}\bigg|_{t=0} \exp\left(\frac{X}{m} + tY\right)\right).$$

We now take the limit as $m \to \infty$. First, for the inside derivative, by smoothness we may take $m \to \infty$ inside, which just gives us $\frac{d}{dt}|_{t=0} \exp(tY) = Y$. The normalized sum converges to $\int_0^1 e^{-t\mathrm{ad}X}\,dt$ by Lemma 14.17. Finally, $\exp\frac{m-1}{m}X$ converges to $\exp X$, and under the standard identification of $\mathrm{Lie}(G)$ with T_eG, $(L_{\exp X})_*$ is identified with $dL_{\exp X}|_e$; this concludes the proof. $\square$

Let us conclude by combining with the chain rule to state the following version of Theorem 14.16; the proof is immediate.

COROLLARY 14.20. *Let G be a Lie group, and let $s \mapsto X(s)$ be a smooth curve in* $\mathrm{Lie}(G)$. *Then*

$$\frac{d}{ds} \exp X(s) = (L_{\exp X(s)})_* \int_0^1 e^{-t\,\mathrm{ad}X(s)} \frac{dX}{ds}\, dt.$$

5. The Baker-Campbell-Hausdorff(-Poincaré-Dynkin) formula

We can now state an integral form of the Baker-Campbell-Hausdorff formula. We need a certain holomorphic function. First, let $\log z$ denote the standard branch of the complex logarithm. If we restrict it to the disk $|z-1|<1$, it has the convergent power series

$$\log z = \sum_{n=1}^{\infty} \frac{(-1)^n}{n}(z-1)^n.$$

Now, define

$$g(z) = \frac{\log z}{1-\frac{1}{z}},$$

which is then manifestly holomorphic on the annulus $0<|z-1|<1$. In fact, 1 is a removable singularity: by l'Hôpital's rule,

$$\lim_{z\to 1} g(z) = \lim_{z\to 1} \frac{\frac{d}{dz}\log z}{\frac{d}{dz}(1-\frac{1}{z})} = \lim_{z\to 1} \frac{\frac{1}{z}}{\frac{1}{z^2}} = 1.$$

Thus, g is actually holomorphic on the disk $|z-1|<1$, and so it has a convergent power series

$$g(z) = \sum_{n=0}^{\infty} g_n (z-1)^n, \quad |z-1|<1.$$

We will return to the power series coefficients g_n later in this section; for now it is good enough to know they exist.

We can then use holomorphic functional calculus to define $g(A)$ for any linear operator A on a finite dimensional vector space with operator norm satisfying $\|A-I\|<1$; it is defined simply as the power series

$$g(A) = \sum_{n=0}^{\infty} g_n (A-I)^n.$$

We may now state the first form of the Baker-Campbell-Hausdorff formula, which is due to Poincaré.

THEOREM 14.21 (Baker-Campbell-Hausdorff-Poincaré Formula). *Let G be a Lie group, with Lie algebra $\mathfrak{g}$. There is an open neighborhood U of $0 \in \mathfrak{g}$ so that, for all $X, Y \in U$,*

$$\exp X \exp Y = \exp\left(X + \int_0^1 g(e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y})(Y)\, dt\right). \tag{14.7}$$

This is an admittedly *awful* formula, and no one could reasonably use it to actually compute with. Rather, its use is in its form: it manifestly expresses the product $\exp X \exp Y$ as the exponential of something built explicitly out of the Lie algebraic objects $\mathrm{ad}X$ and $\mathrm{ad}Y$. Thus, it shows that, near $0 \in \mathfrak{g}$ at least, the group operation on G is completely determined by the Lie bracket on $\mathfrak{g}$.

PROOF. We know that $\exp$ is a diffeomorphism on some neighborhood U_0 of $0 \in \mathfrak{g}$; thus, we can choose a neighborhood U_1 so that, for all $X, Y \in U_1$ and $t \in [0,1]$, the element $\exp X \exp tY$ is in $\exp(U_0)$, and thus there is a unique element $Z(t) \in U_0$ with $\exp Z(t) = \exp X \exp tY$. Moreover, as $\exp$ is a diffeomorphism on U_0, Z is a smooth function of $t \in [0,1]$. Our goal is to compute $Z(1)$; since Z is a smooth function, we can compute this as

$$Z(1) = Z(0) + \int_0^1 \frac{d}{dt} Z(t)\, dt = X + \int_0^1 \frac{d}{dt} Z(t)\, dt.$$

Thus, it bhooves us to calculate the derivative of $Z(t)$. Now, on the one hand, by definition we have

$$\frac{d}{dt} \exp Z(t) = \frac{d}{dt} \exp X \exp tY = dL_{\exp X} \frac{d}{dt} \exp tY = dL_{\exp X} \circ dL_{\exp tY}(Y)$$

where we have used the product rule of Lemma 14.18, and the definition of $\exp tY$ as the integral curve of the left-invariant vector field Y. By the chain rule, and the fact that $L_{gh} = L_g \circ L_h$, we can write this as $dL_{\exp X \exp tY}(Y) = dL_{\exp Z(t)} Y$, and so we have

$$dL_{\exp(-Z(t))} \frac{d}{dt} \exp Z(t) = Y.$$

On the other hand, by Corollary 14.20, we have

$$dL_{\exp(-Z(t))} \frac{d}{dt} \exp Z(t) = \int_0^1 e^{-s\,\mathrm{ad}Z(t)} \frac{d}{dt} Z(t)\, ds.$$

That is, if we let $h(z) = \int_0^1 e^{-sz}\, ds = \frac{1-e^{-z}}{z}$ (with a removable singularity at $z = 0$), then

$$h(\mathrm{ad}Z(t)) \frac{d}{dt} Z(t) = Y.$$

Now $h(0) = 1$ so $h(A)$ is invertible for all sufficiently small operators A. Shrinking U_1 if necessary, we may assume that X, Y are small enough to make $\mathrm{ad}Z(t)$ small enough (for all $t \in [0,1]$) that $h(\mathrm{ad}Z(t))$ is invertible. That is:

$$\frac{d}{dt} Z(t) = h(\mathrm{ad}Z(t))^{-1} Y.$$

It remains to see that $h(\mathrm{ad}Z(t))^{-1}$ has the desired form. We use the relationship between Ad and ad to compute that

$$e^{\mathrm{ad}Z(t)} = \mathrm{Ad}(\exp Z(t)) = \mathrm{Ad}(\exp X \exp tY) = \mathrm{Ad}(\exp X)\mathrm{Ad}(\exp tY) = e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y}.$$

Thus

$$h(\mathrm{ad}Z(t)) g(e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y}) = h(\mathrm{ad}Z(t)) g(e^{\mathrm{ad}Z(t)}).$$

For any sufficiently small complex number w,

$$h(w) g(e^w) = \int_0^1 e^{-sw}\, ds \frac{w}{1-e^{-w}} = \frac{1}{1-e^{-w}} \int_0^1 w e^{-sw}\, ds = 1.$$

It then follows that, for all sufficiently small operator A, $h(A)g(e^A) = I$. Thus $h(\mathrm{ad}Z(t))^{-1} = g(e^{\mathrm{ad}X}e^{t\,\mathrm{ad}Y})$, concluding the proof. $\square$

Now, to derive a series expansion, we must expand g as a power series, and also expand $e^{\mathrm{ad}X}e^{t\,\mathrm{ad}Y}$, combining appropriately. While the computations are elementary, they become fierce quickly. Let us content ourselves for now with the expansion to order 2. By differentiation we can quickly verify that

$$g(z) = 1 + \frac{1}{2}(z-1) - \frac{1}{6}(z-1)^2 + O((z-1)^3).$$

At the same time, we have

$$\begin{aligned} e^{\mathrm{ad}X}e^{t\,\mathrm{ad}Y} - I &= \left(I + \mathrm{ad}X + \frac{1}{2}(\mathrm{ad}X)^2 + \cdots\right)\left(I + t\,\mathrm{ad}Y + \frac{t^2}{2}(\mathrm{ad}Y)^2 + \cdots\right) - I \\ &= \mathrm{ad}X + t\,\mathrm{ad}Y + t\,\mathrm{ad}X\,\mathrm{ad}Y + \frac{1}{2}(\mathrm{ad}X)^2 + \frac{t^2}{2}(\mathrm{ad}Y)^2 + \cdots \end{aligned}$$

where the $\cdots$ mean terms of degree ≥ 3 in $\mathrm{ad}X$ and $\mathrm{ad}Y$. Since there is no degree 0 term in this expansion, taking the nth power will result in terms of degree $\geq n$, and so composing the Taylor series we get

$$\begin{aligned} g(e^{\mathrm{ad}X}e^{t\,\mathrm{ad}Y}) &= I + \frac{1}{2}\left(\mathrm{ad}X + t\,\mathrm{ad}Y + t\,\mathrm{ad}X\,\mathrm{ad}Y + \frac{1}{2}(\mathrm{ad}X)^2 + \frac{t^2}{2}(\mathrm{ad}Y)^2\right) \\ &\qquad - \frac{1}{6}\left(\mathrm{ad}X + t\,\mathrm{ad}Y + t\,\mathrm{ad}X\,\mathrm{ad}Y + \frac{1}{2}(\mathrm{ad}X)^2 + \frac{t^2}{2}(\mathrm{ad}Y)^2\right)^2 + \cdots \\ &= I + \frac{1}{2}\left(\mathrm{ad}X + t\,\mathrm{ad}Y + t\,\mathrm{ad}X\,\mathrm{ad}Y + \frac{1}{2}(\mathrm{ad}X)^2 + \frac{t^2}{2}(\mathrm{ad}Y)^2\right) \\ &\qquad - \frac{1}{6}\left((\mathrm{ad}X)^2 + t^2(\mathrm{ad}Y)^2 + t\,\mathrm{ad}X\mathrm{ad}Y + t\,\mathrm{ad}Y\mathrm{ad}X\right) + \cdots \\ &= I + \frac{1}{2}\mathrm{ad}X + \frac{t}{2}\mathrm{ad}Y + \frac{1}{12}(\mathrm{ad}X)^2 + \frac{t^2}{12}(\mathrm{ad}Y)^2 + \frac{t}{3}\mathrm{ad}X\,\mathrm{ad}Y - \frac{t}{6}\mathrm{ad}Y\,\mathrm{ad}X + \cdots \end{aligned}$$

Applying this to Y, many of the terms cancel (as $\mathrm{ad}Y(Y) = [Y,Y] = 0$), and we find

$$g(e^{\mathrm{ad}X}e^{t\,\mathrm{ad}Y})(Y) = Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{t}{6}[Y,[X,Y]] + \cdots$$

Integrating,

$$\int_0^1 g(e^{\mathrm{ad}X}e^{t\,\mathrm{ad}Y})(Y)\,dt = Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \cdots$$

Thus, Theorem 14.21 shows that

$$\exp X \exp Y = \exp\left(X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \cdots\right)$$

justifying the formula in Remark 14.3.

Following this procedure, being very methodical, it is possible to find the complete expansion. This was done by Dynkin, which is why we attach his name to the formula. We state it here for completeness, but we do not prove it (as the proof would take several pages, and would not really teach us anything).

THEOREM 14.22 (Baker-Campbell-Hausdorff-Dynkin Formula). *Let G be a Lie group, with Lie algebra $\mathfrak{g}$. There is an open neighborhood U of $0 \in \mathfrak{g}$ so that, for all $X, Y \in U$, $\exp X \exp Y = \exp(\mu(X,Y))$, where*

$$\mu(X,Y) = X+Y+\sum_{k=1}^{\infty} \frac{(-1)^k}{k+1} \sum_{\substack{m_1,\dots,m_k\geq 0\\ n_1,\dots,n_k\geq 0\\ m_j+n_j>0}} \frac{1}{m_1+\cdots+m_k+1} \frac{(\mathrm{ad}X)^{m_1}}{m_1!}\frac{(\mathrm{ad}Y)^{n_1}}{n_1!}\cdots\frac{(\mathrm{ad}X)^{m_k}}{m_k!}\frac{(\mathrm{ad}Y)^{n_k}}{n_k!}(X).$$

In particular, we can check the 4th order term from here. The $\mathrm{ad}(\cdot)$ that immediately hits X must be an $\mathrm{ad}Y$ (as $\mathrm{ad}X(X) = 0$)

6. Interlude: Riemannian Distance

Our next goal is to generalize Theorem 14.15 (as far as possible) to other Lie groups. To do so, we will need a few ideas from Riemannian geometry that we have so far avoided. Recall that a *Riemannian metric* on a manifold M is a symmetric tensor field $\mathrm{g} \in \mathscr{T}^2(M)$ that is positive-definite at each point: for each $p \in M$, $\mathrm{g}_p \colon T_pM \times T_pM \to \mathbb{R}$ is an inner product. We then denote the length of a vector $X_p \in T_pM$ by

$$|X_p|_{\mathrm{g}} \equiv \mathrm{g}(X_p, X_p)^{1/2}.$$

One of the key uses of Riemannian metrics is to give a metric structure to manifolds. The way to do this is to use them to measure the lengths of curves.

DEFINITION 14.23. *Let (M, g) be a Riemannian manifold. Let $a < b$ in $\mathbb{R}$, and let $\alpha \colon [a,b] \to M$ be a piecewise smooth curve. The* **Riemannian length** *of α is defined to be*

$$\mathrm{L}_{\mathrm{g}}(\alpha) \equiv \int_a^b |\dot\alpha(t)|_{\mathrm{g}}\, dt.$$

This precisely mirrors the usual calculus definition of lengths of curves. It is easy to verify standard properties of this length definition; for example,

$$\mathrm{L}_{\mathrm{g}}(\alpha|_{[a,c]}) = \mathrm{L}_{\mathrm{g}}(\alpha|_{[a,b]}) + \mathrm{L}_{\mathrm{g}}(\alpha|_{[b,c]}), \qquad a \leq b \leq c. \tag{14.8}$$

It is important to note that L_{g} is independent of parametrization.

LEMMA 14.24. *Let (M, g) be a Riemannian manifold, and let $\alpha \colon [a,b] \to M$ be a piecewise smooth curve. Let β be a reparametrixation of α (see the discussion preceding Proposition 6.25). Then $\mathrm{L}_{\mathrm{g}}(\alpha) = \mathrm{L}_{\mathrm{g}}(\beta)$.*

The proof is a standard calculation involving the change-of-variables from calculus, and is left to the reader.

We can now use lengths of curves to define distances between points, by using curves as "measuring tapes" (and then choosing the shortest one).

DEFINITION 14.25. *Let (M, g) be a connected Riemannian manifold. The* **Riemannian distance** *$d_{\mathrm{g}} \colon M \times M \to \mathbb{R}_+$ is the minimal length of any piecewise smooth curve connecting two points:*

$$d_{\mathrm{g}}(p,q) \equiv \inf_{\alpha\colon p\to q} L_{\mathrm{g}}(\alpha).$$

It is well-defined since, for any two $p, q \in M$ there exists a piecewise smooth curve connecting p to q (cf. Lemma 6.28).

EXAMPLE 14.26. if $\overline{\mathrm{g}}$ is the usual Euclidean inner product on $\mathbb{R}^n$, then (as straight lines are the shortest paths) $d_{\overline{\mathrm{g}}}(p,q) = |p-q|_{\overline{\mathrm{g}}}$, the usual Euclidean metric (in topological sense).

In fact, d_{g} is always a metric in the topological sense. Some of the defining properties of a metric are immediate from the definition; others (such as nondegeneracy) require a little work. First we prove the following lemma, which shows that it is always locally equivalent to the Euclidean inner product.

LEMMA 14.27. *Let U be an open subset of $\mathbb{R}^n$, and let g be a Riemannian metric on U. Let $\overline{\mathrm{g}}$ denote the Euclidean inner product on $\mathbb{R}^n$. If $K \subset U$ is compact, then there are positive constants $c, C > 0$ so that, for all $x \in K$ and $v \in T_xU$,*

$$c|v|_{\overline{\mathrm{g}}} \le |v|_{\mathrm{g}} \le C|v|_{\overline{\mathrm{g}}}.$$

PROOF. As usual, we identify $T_xU \cong \mathbb{R}^n$, and so $TU \cong U \times \mathbb{R}^n$. Consider first the subset $K \times \mathbb{S}^{n-1} \subset K \times \mathbb{R}^n \subset TU$, which is compact. By definition of $\mathbb{S}^{n-1}$, $|(x,v)|_{\overline{\mathrm{g}}} = 1$ for $(x,v) \in K \times \mathbb{S}^{n-1}$. The function $(x,v) \mapsto |(x,v)|_{\mathrm{g}}$ is continuous and strictly positive, and so by compactness there are constants $c, C > 0$ such that $\le |(x,v)|_{\mathrm{g}} \le C$ for $(x,v) \in K \times \mathbb{S}^{n-1}$. Now, for any $(x,v) \in K \times \mathbb{R}^n$, either $v = 0$ (in which case $|(x,v)|_{\mathrm{g}} = |(x,v)|_{\overline{\mathrm{g}}} = 0$) or there is a unique $\lambda > 0$ (equal to $|(x,v)|_{\overline{\mathrm{g}}}$ so that $\lambda^{-1}v \in \mathbb{S}^{n-1}$. Thus, by the homogeneity of the norm $|\cdot|_{\mathrm{g}}$ at the point x,

$$|(x,v)|_{\mathrm{g}} = \lambda|(x,\lambda^{-1}v)|_{\mathrm{g}} \le \lambda C = C|(x,v)|_{\overline{\mathrm{g}}}.$$

This is half of the desired inequality; the other half is proved the same way. □

THEOREM 14.28. *Let (M, g) be a connected Riemannian manifold. The the Riemannian distance function $d_{\mathrm{g}}\colon M \times M \to \mathbb{R}$ is a (topological) metric, and the metric space (M, d_{g}) has the same topology as M.*

PROOF. By definition $d_{\mathrm{g}} \ge 0$. It is symmetric since, for any curve $\alpha\colon p \to q$, the reversed curve $-\alpha\colon q \to p$ is a reparametrization, and so by Lemma 14.24 $\mathrm{L}_{\mathrm{g}}(\alpha) = \mathrm{L}_{\mathrm{g}}(-\alpha)$, which (taking $\inf_\alpha$ of both sides) shows that $d_{\mathrm{g}}(p,q) = d_{\mathrm{g}}(q,p)$. The triangle inequality follows from (14.8). Thus, to show d_{g} is a metric, it suffices to show that it is nondegenerate: $d_{\mathrm{g}}(p,q) > 0$ if $p \ne q$.

Fix $p \ne q$, and let (U, φ) be a chart containing p by not q. As usual, let $\widehat{U} = \varphi(U)$. We can then define a Riemannian metric $\widehat{\mathrm{g}}$ on $\widehat{U}$ by $\widehat{\mathrm{g}} = (\varphi^{-1})_*\mathrm{g}$,

$$\widehat{\mathrm{g}}(v,w) = \mathrm{g}(d(\varphi^{-1})_{\varphi(r)}(v), d(\varphi^{-1})_{\varphi(r)}(w)), \qquad r \in U,\ v, w \in T_{\varphi(r)}\widehat{U}.$$

This is a Riemannian metric since φ^{-1} is a diffeomorphism, so $d(\varphi^{-1})_{\varphi(r)}$ is a linear isomorphism at each point $r \in U$.

Fix some $\epsilon > 0$ so that $\overline{B_\epsilon(\varphi(p))} \subset \widehat{U}$; let $V = \varphi^{-1}(B_\epsilon(\varphi(p)))$, so that $\overline{V} \subset U$ is compact. By Lemma 14.27, there are constants $c, C > 0$ so that

$$c|v|_{\overline{\mathrm{g}}} \le |v|_{\widehat{\mathrm{g}}} \le C|v|_{\overline{\mathrm{g}}}, \qquad \forall\, x \in B_\epsilon(\varphi(p)),\ v \in T_xB_\epsilon(\varphi(p))$$

where $\overline{\mathrm{g}}$ is the Euclidean metric.

Now, let $\alpha\colon [a,b] \to \overline{V}$ be a piecewise smooth curve contained in $\overline{V}$, and let $\widehat{\alpha}\colon [a,b] \to \mathbb{R}^n$ be the coordinate curve $\widehat{\alpha} = \varphi \circ \alpha$. Then, integrating, we have

$$c\mathrm{L}_{\overline{\mathrm{g}}}(\widehat{\alpha}) \le \mathrm{L}_{\widehat{\mathrm{g}}}(\widehat{\alpha}) \le C\mathrm{L}_{\overline{\mathrm{g}}}(\widehat{\alpha}).$$

On the other hand, note that

$$|\widehat{\alpha}'(t)|^2_{\widehat{\mathrm{g}}} = \mathrm{g}\left(d(\varphi^{-1})_{\widehat{\alpha}(t)}\frac{d}{dt}\widehat{\alpha}(t), d(\varphi^{-1})_{\widehat{\alpha}(t)}\frac{d}{dt}\widehat{\alpha}(t)\right) = \mathrm{g}(\dot{\alpha}(t), \dot{\alpha}(t)) = |\dot{\alpha}(t)|_{\mathrm{g}}$$

by the chain rule. It follows that

$$\mathrm{L}_{\widehat{\mathrm{g}}}(\widehat{\alpha}) = \int_a^b |\widehat{\alpha}'(t)|_{\widehat{\mathrm{g}}}\, dt = \int_a^b |\dot{\alpha}(t)|_{\mathrm{g}}\, dt = \mathrm{L}_{\mathrm{g}}(\alpha). \tag{14.9}$$

Thus, we have

$$c\mathrm{L}_{\overline{\mathrm{g}}}(\widehat{\alpha}) \leq \mathrm{L}_{\mathrm{g}}(\alpha) \leq C\mathrm{L}_{\overline{\mathrm{g}}}(\widehat{\alpha}).$$

Now, let γ be a piecewise smooth curve from p to q. Let $t_0 = \inf\{t \in [a,b] \colon \gamma(t) \notin \overline{V}\}$; as $q \notin \overline{V}$, by continuity $t_0 \in (a,b)$, $\gamma(t) \in \overline{V}$ for $a \leq t < t_0$, and $\gamma(t_0) \in \partial V$. It follows that

$$\mathrm{L}_{\mathrm{g}}(\gamma) \geq \mathrm{L}_{\mathrm{g}}(\gamma|_{[a,t_0]}) \geq c\mathrm{L}_{\overline{\mathrm{g}}}(\widehat{\alpha}|_{[a,t_0]}).$$

We know that the $\overline{\mathrm{g}}$-shortest path from $\widehat{\gamma}(a) = p$ to $\widehat{\gamma}(t_0)$ in $\mathbb{R}^n$ is the straight line path, and so

$$\mathrm{L}_{\overline{\mathrm{g}}}(\widehat{\gamma}|_{[a,t_0]}) \geq d_{\overline{\mathrm{g}}}(p, \gamma(t_0)) = \epsilon$$

since $\gamma(t_0) \in \partial V$ and so $\widehat{\gamma}(t_0) \in \partial\varphi(V) = \partial B_\epsilon(p)$. Hence, any piecewise smooth curve from p to q has g-length $\geq c\epsilon$, and it follows taking the infimum that $d_{\mathrm{g}}(p,q) \geq c\epsilon > 0$. Ergo, d_{g} is a metric.

It remains to show that the d_{g} metric topology on M is the same as the manifold topology on M. First, suppose $U \subseteq M$ is open in the manifold topology. Let $p \in U$, and fix a a chart (V, φ) at p so that $\overline{V} \subset U$, and $\varphi(V)$ is a ball of some radius $\epsilon > 0$ centered at $\varphi(p)$. The preceding argument shows that there is a (V and therefore ϵ dependent) constant c so that $d_{\mathrm{g}}(p,q) \geq c\epsilon$ for all $q \notin \overline{V}$. The contrapositive is that, for any point q with $d_{\mathrm{g}}(p,q) < c\epsilon$, it follows that $q \in \overline{V} \subseteq U$. Thus, the g-ball of radius $c\epsilon$ around p is contained in U. As this holds for any $p \in U$, it follows that U is open in (M, g).

Conversely, Let W be open in (M, g), and let $p \in W$. Fix a chart (V, φ) such that $\widehat{V} = \varphi(V) = B_1(\varphi(p))$. Let $\overline{\mathrm{g}}$ be the Euclidean metric on $\widehat{V}$, and let $\widehat{\mathrm{g}} = (\varphi^{-1})_*\mathrm{g}$ as above. Then there are $c, C > 0$ so that

$$c|v|_{\overline{\mathrm{g}}} \leq |v|_{\widehat{\mathrm{g}}} \leq C|v|_{\overline{\mathrm{g}}}, \qquad \forall\, x \in \widehat{V},\, v \in T_x\widehat{V}. \tag{14.10}$$

Fix $\epsilon > 0$ small enough that $\epsilon < 1$ and such that the g-ball $B^{\mathrm{g}}_{C\epsilon}(p) = \{q \in M \colon d_{\mathrm{g}}(p,q) < C\epsilon\}$ is contained in W. At the same time, let V_ϵ denote the set of points $q \in \overline{V} \subseteq M$ such that $d_{\overline{\mathrm{g}}}(\varphi(p), \varphi(q)) < \epsilon$. If $q \in V_\epsilon$, let $\widehat{\alpha}$ be the straight line path from $\varphi(p)$ to $\varphi(q)$, and let $\alpha = \varphi^{-1} \circ \widehat{\alpha}$; then from (14.9) and integration of (14.10), we have

$$d_{\mathrm{g}}(p,q) \leq \mathrm{L}_{\mathrm{g}}(\alpha) = \mathrm{L}_{\widehat{\mathrm{g}}}(\widehat{\alpha}) \leq C\mathrm{L}_{\overline{\mathrm{g}}}(\hat{\alpha}) < C\epsilon.$$

This shows that $V_\epsilon \subseteq B^{\mathrm{g}}_{C\epsilon}(p) \subseteq W$. Since V_ϵ is a neighborhood of p in the manifold topology, this shows W is open in the manifold topology, completing the proof. □

REMARK 14.29. To summarize the idea of the last half of the proof (showing the manifold and metric topologies are the same): the manifold topology is locally Euclidean, which means it is locally the Euclidean metric topology. Lemma 14.27 shows that the Euclidean metric and the given Riemannian metric are locally equivalent metrics, and thus give the same topology.

In the context of Lie groups, the most relevant kind of Riemannian metric is a left-invariant metric.

DEFINITION 14.30. *Let G be a Lie group. A* **left-invariant Riemannian metric** *on G is a Riemannian metrc $\mathrm{g} \in \mathscr{T}^2(G)$ that is left-invariant: for any $g \in G$, $(L_g)_*(\mathrm{g}) = \mathrm{g}$. Equivalently: such a metric is of the form*

$$\mathrm{g}(X_g, Y_g) = \langle dL_g|_e(X_e), dL_g|_e(Y_e)\rangle_e, \qquad g \in G$$

for some given inner product $\langle \cdot, \cdot \rangle_e$ on the Lie algebra $\mathfrak{g} = T_eG$.

A left-invariant Riemannian metric gives rise to a left-invariant Riemannian distance.

PROPOSITION 14.31. *Let G be a Lie group, and let g be a left-invariant Riemannian metric on G. Then d_{g} is left-invariant: for any $g, h_1, h_2 \in G$,*

$$d_{\mathrm{g}}(gh_1, gh_2) = d_{\mathrm{g}}(h_1, h_2).$$

EXAMPLE 14.32. The Euclidean inner product on $\mathbb{R}^n$ is translation-invariant, and as a consequence the Euclidean distance $(x, y) \mapsto |x - y|$ is translation invariant. Proposition 14.31 is the generalization to arbitrary left-invariant Riemannian metrics on Lie groups.

PROOF OF PROPOSITION 14.31. Let $\mathrm{PS}_{h_1 \to h_2}(G)$ denote the set of piecewise linear curves from h_1 to h_2 in G. Then $\alpha \mapsto L_g \circ \alpha$ is a bijection from $\mathrm{PS}_{h_1 \to h_2}(G)$ to $\mathrm{PS}_{gh_1 \to gh_2}(G)$, and so

$$d_{\mathrm{g}}(gh_1, gh_2) = \inf\{\mathrm{L}_{\mathrm{g}}(L_g \circ \alpha) \colon \alpha \in \mathrm{PS}_{h_1 \to h_2}(G)\}. \tag{14.11}$$

We then compute that, for any $\alpha \colon [a, b] \to G$ in $\mathrm{PS}_{h_1 \to h_2}(G)$,

$$\mathrm{L}_{\mathrm{g}}(L_g \circ \alpha) = \int_a^b |(L_g \circ \alpha)'(t)|_{\mathrm{g}}\, dt = \int_a^b |(dL_g)_{\alpha'(t)} \alpha'(t)|_{\mathrm{g}}\, dt$$

and

$$\begin{aligned} |(dL_g)_{\alpha'(t)} \dot{\alpha}(t)|_{\mathrm{g}} &= \mathrm{g}\left((dL_g)_{\alpha'(t)} \alpha'(t), (dL_g)_{\alpha'(t)} \alpha'(t)\right) \\ &= ((L_g)_* \mathrm{g})\,(\alpha'(t), \alpha'(t)) \\ &= \mathrm{g}(\alpha'(t), \alpha'(t)) = |\alpha'(t)|_{\mathrm{g}} \end{aligned}$$

by definition 14.30. Thus $\mathrm{L}_{\mathrm{g}}(L_g \circ \alpha) = L_{\mathrm{g}}(\alpha)$, and this, together with (14.11), concludes the proof. □

To conclude this section, we give an application of the existence of a left-invariant distance function as in Proposition 14.31 that we will need in the next section. The result is: any continuous curve α has *multiplicative increments* $\alpha(s)^{-1}\alpha(t)$ that all stay within a given neighborhood of the identity when s, t are sufficiently small.

LEMMA 14.33. *Let G be a Lie group and let $\alpha \colon [a, b] \to G$ be a continuous path. Then for any neighborhood U of $e \in G$, there is a $\delta > 0$ so that, for any $s, t \in [a, b]$ with $|s - t| < \delta$, $\alpha(s)^{-1}\alpha(t) \in U$.*

PROOF. Fix some left-invariant Riemannian metric g on G. Since α is continuous $[a, b] \to M$, it is continuous as a map from $[a, b]$ to the metric space (M, d_{g}) by Theorem 14.28. Since $[a, b]$ is compact, it follows that α is uniformly continuous, so for any $\epsilon > 0$, there is some $\delta = \delta(\epsilon) > 0$ such that $|s - t| < \delta$ implies that $d_{\mathrm{g}}(\alpha(s), \alpha(t)) < \epsilon$. Now, let $\epsilon > 0$ be small enough that the metric ball $B^{\mathrm{g}}_{\epsilon}(e) = \{g \in G \colon d_{\mathrm{g}}(e, g) < \epsilon\}$ is contained in U. Then by left-invariance and Proposition 14.31, for $|s - t| < \delta(\epsilon)$,

$$d_{\mathrm{g}}(\alpha(s)^{-1}\alpha(t), e) = d_{\mathrm{g}}(\alpha(s) \cdot \alpha(s)^{-1}\alpha(t), \alpha(s) \cdot e) = d_{\mathrm{g}}(\alpha(t), \alpha(s)) = d_{\mathrm{g}}(\alpha(s), \alpha(t)) < \epsilon.$$

Thus $\alpha(s)^{-1}\alpha(t) \in B^{\mathrm{g}}_{\epsilon}(e) \subseteq U$ whenever $|s - t| < \delta$, as required. □

REMARK 14.34. As pointed out in class by Jonathan Conder, Lemma 14.33 can be proved without knowledge of the existence of a left-invariant distance function. We include the simple topology proof below. We may have use for left-invariant Riemannian metrics later on, so the material in this section is good to keep in mind.

ALTERNATE PROOF OF LEMMA 14.33. The function $A\colon [a,b]\times[a,b] \to G$ given by $A(s,t) = \alpha(s)^{-1}\alpha(t)$ is continuous. Thus $A^{-1}(U)$ is an open subset of $[a,b] \times [a,b]$. Since $A(t,t) = \alpha(t)^{-1}\alpha(t) = e$ for all $t \in [a,b]$, the diagonal $D = \{(t,t)\colon t \in [a,b]\}$ is contained in $A^{-1}(U)$, so each point (t,t) has a neighborhood contained in $A^{-1}(U)$. Fix a diagonal square S_t centered at each each point $t \in D$ that is contained in $A^{-1}(U)$; so S_t has the form

$$S_t = \{(s',t')\colon |s'-t'| < \delta',\ \&\ t - \tfrac{\delta'}{2} < s'+t' < t + \tfrac{\delta'}{2}\}.$$

Since D is compact, it is covered by a finite collection of such squares. Let δ be the minimum side-length of the covering squares. Then $|s-t| < \delta$ implies $(s,t) \in A^{-1}(U)$, as required. □

7. The Lie Correspondence

In this section, we prove the converse of Proposition 12.42: Lie algebra homomorphisms give rise to Lie group homomorphisms (when the domain group is simply-connected).

THEOREM 14.35. *Let G, H be Lie groups, with Lie algebra $\mathfrak{g}, \mathfrak{h}$, and suppose that G is simply-connected. If $\phi\colon \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism, then there is a unique Lie group homomorphism $\Phi\colon G \to H$ such that $\Phi(\exp X) = \exp(\phi(X))$ for all $X \in \mathfrak{g}$.*

Note that, since $d\exp_e = \mathrm{Id}$, it follows that $\Phi_* = \phi$. The condition that G is simply-connected is necessary – the next chapter will explore the extent to which Theorem 14.35 fails without this assumption. Note: simply-connected implies connected.

Let us begin with the following far less ambitious uniqueness lemma: on a connected group, there is *at most one* Lie group homomorphism which induces any given Lie algebra homomorphism.

LEMMA 14.36. *Let G, H be Lie groups, and let G be connected. Suppose $\Phi_1, \Phi_2\colon G \to H$ are Lie group homomorphism. If $(\Phi_1)_* = (\Phi_2)_*$, then $\Phi_1 = \Phi_2$.*

PROOF. Denote $\phi = (\Phi_1)_* = (\Phi_2)_*$. By Theorem 13.9(g), we know that, for $j = 1,2$, $\Phi_j(\exp X) = \exp(\phi(X))$ for any $X \in \mathrm{Lie}(G)$, and by part (e) of the same theorem, there is a neighborhood U of $0 \in \mathrm{Lie}(G)$ such that $\exp\colon U \to V = \exp(U) \subseteq G$ is a diffeomorphism. Since G is connected, V generates all of G by 11.12, and thus every element $g \in G$ can be written in the form $g = \exp(X_1)\cdots\exp(X_n)$ for some $X_1,\dots,X_n \in U$. Thus, for $j = 1,2$,

$$\Phi_j(g) = \Phi_j(\exp(X_1)\cdots\exp(X_n)) = \Phi_j(\exp(X_1))\cdots\Phi_j(\exp(X_n)) = \exp(\phi(X_1))\cdots\exp(\phi(X_n))$$

and so $\Phi_1(g) = \Phi_2(g)$ for all $g \in G$. □

Before embarking on the proof of Theorem 14.35, let us note the following Corollary, which is the main point of theoretical interest: in the simply-connected category, Lie group isomorphism and Lie algebra isomorphism are equivalent.

COROLLARY 14.37 (Lie Correspondence). *Let G, H be simply-connected Lie group, with Lie algebras $\mathfrak{g}, \mathfrak{h}$. Then G is isomorphic to H if and only if $\mathfrak{g}$ is isomorphic to $\mathfrak{h}$.*

PROOF. The 'only if' direction was proved in Proposition 12.45, so assume we know $\mathfrak{g} \cong \mathfrak{h}$, and let $\phi\colon \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra isomorphism. Let $\Phi\colon G \to H$ be the associated Lie group homomorphism as in Theorem 14.35. Similarly, $\phi^{-1}\colon \mathfrak{h} \to \mathfrak{g}$ is a Lie algebra isomorphism, and so it has an associated Lie group homomorphism $\Psi\colon H \to G$. We will show that Φ and Ψ are inverses.

The map $\Phi \circ \Psi \colon H \to H$ is a Lie group homomorphism, and by the chain rule, its differential is $\phi \circ \phi^{-1} = \mathrm{Id}_{\mathfrak{h}} = (\mathrm{Id}_H)_*$; similarly, $(\Psi \circ \Phi)_* = \mathrm{Id}_{\mathfrak{g}} = (\mathrm{Id}_G)_*$. The result now follows from Lemma 14.36. $\square$

Thus: determining whether two simply-connected Lie groups are isomorphic reduces to the *linear algebra* problem of deciding whether their Lie algebras are isomorphic.

To prove Theorem 14.35, we begin by mimicking the proof of Theorem 14.15, now using the full Baker-Campbell-Hausdorff formula instead of its special case Corollary 14.13 which holds in $\mathrm{H}(3, \mathbb{R})$. The key difference is: we no longer know that the exponential map on the domain group is a global diffeomorphism (indeed, this is almost never true). As such, the farthest the direct analog of the proof can get us is to construct a *local homomorphism*.

DEFINITION 14.38. *Let G, H be Lie groups. A* **local homomorphism** *from G to H is a pair (F, U) where $V \subseteq G$ is a connected neighborhood of the identity, and $F \colon V \to G$ is a smooth map which satisfies the following property: if $g, h \in V$ are such that $gh \in V$, then $F(gh) = F(g)F(h)$.*

For example, if $\Phi \colon G \to H$ is a Lie group homomorphism, then $\Phi|_V$ is a local homomorphism for any connected neighborhood V. The question is whether all local homomorphism are restrictions like this, and the answer depends on the topology of G, as we will see later in this section. First, let us now see how far the analog of Theorem 14.15 will take us.

PROPOSITION 14.39. *Let G, H be Lie groups with Lie algebra $\mathfrak{g}, \mathfrak{h}$. If $\phi \colon \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism, then there is a local Lie group homomorphism (F, V) such that $\exp$ is a diffeomorphism $\exp^{-1}(V) \to V$ and $F(\exp X) = \exp \phi(X)$ whenever $\exp X \in V$.*

PROOF. We apply Theorem 14.21, choosing U so that (14.7) holds both for (X, Y) and for $(\phi(X), \phi(Y))$, and so that $\exp$ is one-to-one on U and U is connected; then set $V = \exp(U)$. We may then define F on V by $F = \exp \circ \phi \circ \exp^{-1}|_V$, which is a composition of smooth maps. We must prove that F is a local homomorphism. Let $g, h \in U$; then there are unique $X, Y \in \mathfrak{g}$ so that $\exp X = g$ and $\exp Y = h$. Assume also that $gh \in U$. Applying the Baker-Campbell-Hausdorff formula, we have

$$F(gh) = F(\exp X \exp Y) = F\left(X + \int_0^1 g(e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y})Y\,dt\right).$$

By definition of F,

$$F(gh) = \exp\left[\phi\left(X + \int_0^1 g(e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y})Y\,dt\right)\right].$$

Since ϕ is linear and continuous, we may pass it through the sum and integral:

$$F(gh) = \exp\left[\phi(X) + \int_0^1 \phi\left(g(e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y})Y\right)\,dt\right].$$

Since ϕ is a Lie algebra homomorphism, the integrand satisfies

$$\phi\left(g(e^{\mathrm{ad}X} e^{t\,\mathrm{ad}Y})Y\right) = g(e^{\mathrm{ad}\phi(X)} e^{t\,\mathrm{ad}\phi(Y)})\phi(Y).$$

Indeed: if we expand $g(AB)$ as a (noncommutative) power series in two operators A, B, we see that the left-hand-side of the expression is a sum of terms each of which is a finite iterated bracket of Xs and Ys (indeed, the precise form of this sum is given in Dynkin's formula of Theorem 14.22). The statement that ϕ is a Lie algebra homomorphism is the statement that the image of

each such term under ϕ is the result of replacing (X, Y) with $(\phi(X), \phi(Y))$. Then recombining, we get the right-hand-side. Finally, we apply the Baker-Campbell-Hausdorff formula once morein the other direction, and this yield

$$F(gh) = \exp \phi(X) \exp \phi(Y) = F(\exp X)F(\exp Y) = F(g)F(h).$$

Thus, F is a local homomorphism, as desired. □

Now, to prove Theorem 14.35, it suffices to show that such a local Lie homomorphism actually extends to a Lie group homomorphism; if so, it will be unique by Lemma 14.36. The remainder of this section will be devoted to the proof; we state it as a proposition.

PROPOSITION 14.40. *Let G, H be Lie groups, with G simply connected. If (F, V) is a local homomorphism $G \to H$, then there exists a (global) Lie group homomorphism $\Phi\colon G \to H$ such that $\Phi|_V = F$.*

PROOF. We break the proof into several parts.

Step 1: define Φ along a path. As G is simply connected (which implies connected, which in a manifold implies path connected), for any $g \in G$ there is a piecewise smooth curve $\alpha\colon [0, 1] \to G$ so that $\alpha(0) = e$ and $\alpha(1) = g$. Fix any partition $P = \{0 = t_0 < t_1 < \cdots < t_n = 1\}$ of $[0, 1]$ so that, for any j and any $s, t \in [t_{j-1}, t_j]$, the increment $\alpha(s)^{-1}\alpha(t) \in V$; call such a partition **adapted to** (α, V). (The existence of adapted partitions is guaranteed by Lemma 14.33: choose a $\delta > 0$ so that $\alpha(s)^{-1}\alpha(t) \in V$ whenever $|s - t| < \delta$, and then select any partition with maximum width $< \delta$.) Hence $\alpha(t_1) = e^{-1}\alpha(t_1) = \alpha(t_0)^{-1}\alpha(t_1) \in V$, and we can in general express $g = \alpha(1)$ as a telescoping product

$$g = [\alpha(t_1)] \cdot [\alpha(t_1)^{-1}\alpha(t_2)] \cdots [\alpha(t_{n-1})^{-1}\alpha(t_n)],$$

a product of increments each of which is in V. We will now define a map $\Phi = \Phi_{\alpha,P}\colon G \to H$ as follows:

$$\Phi(g) = F(\alpha(t_1))F(\alpha(t_1)^{-1}\alpha(t_2)) \cdots F(\alpha(t_{n-1})^{-1}\alpha(t_n)).$$

This at least makes sense, but Φ a priori depends both on the continuous path α chosen to connect e to g, and on the given partition P that keeps the increments in V. The remainder of the proof will show that Φ is in fact independent of the path and partition, and moreover is a Lie group homomorphism that agrees with F on V.

Step 2: independence of the partition. Suppose P is a partition adapted to (α, V). Then it is immediate from the definition that any refinement of P is also adapted to (α, V). We claim that if P' is a refinement of P then $\Phi_{\alpha,P}(g) = \Phi_{\alpha,P'}(g)$. To prove this, it suffices by induction to prove it when P' is a refinement by a single point, say $s \in (t_{j-1}, t_j)$. Then

$$\Phi_{\alpha,P'}(g) = F(\alpha(t_1)) \cdots F(\alpha(t_{j-1})^{-1}\alpha(s))F(\alpha(s)^{-1}\alpha(t_j)) \cdots F(\alpha(t_{n-1})^{-1}\alpha(t_n)). \tag{14.12}$$

Now, since both increments $\alpha(t_{j-1})^{-1}\alpha(s)$ and $\alpha(s)^{-1}\alpha(t_j)$ are in V (since P is adapted to (α, V)), where F is a local homomorphism. Thus

$$F(\alpha(t_{j-1})^{-1}\alpha(s))F(\alpha(s)^{-1}\alpha(t_j)) = F(\alpha(t_{j-1})^{-1}\alpha(s) \cdot \alpha(s)^{-1}\alpha(t_j)) = F(\alpha(t_{j-1})^{-1}\alpha(t_j)),$$

and the expression on the right-hand-side of (14.12) is equal to $\Phi_{\alpha,P}(g)$. Thus $\Phi_{\alpha,P}$ is unchanged under refinements of P.

Well, given any two partitions P, P' adapted to (α, V), their common refinement P'' (the union of all partition points in both) therefore gives the same value for both: $\Phi_{\alpha,P}(g) = \Phi_{\alpha,P''}(g) = \Phi_{\alpha,P'}(g)$. Thus, $\Phi_{\alpha,P}(g)$ is independent of the choice of partition P adapted to (α, V). We now refer to it simply as $\Phi_\alpha = \Phi_{\alpha,P}$.

Step 3: independence of the path. Here is where we use the simple-connectedness of G. Let α_1 and α_2 be two continuous paths $[0,1] \to G$ each connecting e to g. By simple-connectedness, these paths are homotopic with endpoints fixed: there is a continuous map $A\colon [0,1] \times [0,1] \to G$ so that $A(s,t) = \alpha_s(t)$ for $s \in \{0,1\}$, and moreover $A(s,0) = e$ and $A(s,1) = g$ for all $s \in [0,1]$. This gives us a continuous family of curves $\alpha_s\colon [0,1] \to G$ each connecting e to g. We will use this homotopy to show that $\Phi_{\alpha_0} = \Phi_{\alpha_1}$ by showing, in fact, that Φ_{α_s} does not depend on s.

To begin, we extend Lemma 14.33 to the two variable setting, and find $\delta > 0$ so that

$$A(s',t')^{-1}A(s,t) \in V \text{ whenever } |s-s'| < \delta \text{ and } |t-t'| < \delta.$$

(The proof is essentially the same). Now we deform $\alpha_0 = A(0,\cdot)$ into $\alpha_1 = A(1,\cdot)$ "a little bit at a time". Fix $N \in \mathbb{N}$ with $\frac{2}{N} < \delta$, and for $k \in \{0,\dots,N-1\}$ and $\ell \in \{0,\dots,N\}$ we define a new curve $A_{k,\ell}\colon [0,1] \to G$ as follows:

$$A_{k,\ell}(s,t) = A(c_{k,\ell}(s,t))$$

where $c_{k,\ell}\colon [0,1] \times [0,1] \to [0,1] \times [0,1]$ is the piecewise linear map

$$c_{k,\ell}(t) = \begin{cases} (\frac{k+1}{N}, t), & 0 \le t \le \frac{\ell-1}{N} \\ \cdots, & \frac{\ell-1}{N} \le t \le \frac{\ell}{N} \\ (\frac{k}{N}, t), & \frac{\ell}{N} \le t \le 1 \end{cases}$$

where the $\cdots$ is determined by the piecewise-linear condition. So $A_{0,0} = A(0,\cdot) = \alpha_0$ and $A_{N,0} = A(1,\cdot) = \alpha_1$, and in general $A_{k,0}(t) = A(k/N, t)$.

We will now deform α_0 into α_1 in the following sequence:

$$A_{0,0} \to A_{0,1} \to \cdots \to A_{0,N} \to A_{1,0} \to \cdots A_{1,N} \to \cdots \to A_{N-1,0} \to \cdots \to A_{N-1,N} \to A_{N,0}.$$

The claim is that, for every step in this deformation process, the value of $\Phi_{A_{\cdot,\cdot}}$ is unchanged. Indeed, from the first step, we know that the value of Φ along any path is determined only by the values of the curve at the partition points of any adapted partition. So consider a step of the form $A_{k,\ell} \to A_{k,\ell+1}$; these two curves agree except on the interval $[\frac{\ell-1}{N}, \frac{\ell+1}{N}]$. Thus, for this step we choose the partition $\{0, \frac{1}{N}, \dots, \frac{\ell-1}{N}, \frac{\ell+1}{N}, \cdots, 1\}$. Since the maximum partition width is $\frac{2}{N} < \delta$, the partition is adapted, and so we may use it to calculate $\Phi_{A_{k,\ell}}$ and $\Phi_{A_{k,\ell+1}}$. As these two agree on this partition, it follows that $\Phi_{A_{k,\ell}} = \Phi_{A_{k,\ell+1}}$. Similarly, for a transition of the form $A_{k,N} \to A_{k+1,0}$, we note that these two differ only on the intervals $(0, \frac{1}{N}]$ and $[\frac{N-1}{N}, 1)$, so they agree on all the partition points in $\{0, \frac{2}{N}, \frac{3}{N}, \cdots, \frac{N-2}{N}, 1\}$. Again, this is an adapted partition since its maximum interval length is $\frac{2}{N} < \delta$, and so $\Phi_{A_{k,N}} = \Phi_{A_{k+1,0}}$. Thus, stringing them all together, we have $\Phi_{\alpha_0} = \Phi_{A_{0,0}} = \Phi_{A_{N,0}} = \Phi_{\alpha_1}$ as desired. Hence, Φ does not depend on the choice of α.

Step 4: Φ is a Lie group homomorphism that agrees with F on V. Showing that Φ is a homomorphism is left as an homework exercise. To see that it is smooth, fix $g \in G$, and let (U, φ) be a chart at g. Fix some point $g_0 \in U$ not equal to g, and let $U_0 \subset U$ be a coordinate ball centered at g (so $\varphi(U_0) = B_r(\varphi(g))$) such that $g_0 \in U_0$. Let $0 < \epsilon < r$ be such that $\varphi(g_0) \notin B_\epsilon(\varphi(g))$. Let $U_\epsilon = \varphi^{-1}(B_\epsilon(\varphi(g)))$ Now, choose any continuous curve α_0 connecting e to g_0, and then let $\ell_g = \varphi^{-1} \circ \widehat{\ell}_g$ where $\widehat{\ell}_g(t) = (1-t)\varphi(g_0) + t\varphi(g)$ is the straight line path from $\varphi(g_0)$ to $\varphi(g)$. Then the concatenation $\ell_g \alpha_{g_0}$ is a continuous path from e to g, and so we have $\Phi(g) = \Phi_{\ell_g \alpha_{g_0}}(g)$. Now, choose a δ small enough that all partitions of width $< \delta$ are adapted to all the curves $\ell_g \alpha_{g_0}$ for all $g \in U_\epsilon$ (this is possible by compactness of $\overline{U_\epsilon}$). Then we may choose a fixed partition P of

width $< \delta$ such that $g_0 = \alpha_{g_0}(\frac{1}{2})$ for definiteness, and then we have

$$\Phi(g) = \prod_{j\colon t_j \le \frac{1}{2}} F(\alpha_{g_0}(t_{j-1})^{-1}\alpha_{g_0}(t_j)) \cdot \prod_{t\colon t_j > \frac{1}{2}} F(\ell_g(t_{j-1})^{-1}\ell_g(t_j)).$$

The first product is a fixed element of G. The second product is a fixed, finite product of increments of a curve ℓ_g that varies smoothly with g, and thus are smooth functions of g; since F is smooth, it now follows that Φ is a smooth function of $g \in U_\epsilon$. So Φ is smooth on a neighborhood of each point in G, and thus is smooth.

So we are left to show that $\Phi|_V = F$. Here is where our assumption that V is connected (in Definition 14.38) comes into play. Fix any $g \in V$, and let $\alpha\colon [0,1] \to V$ be a continuous path connecting e to g. Fix a partition $P = \{t_0 < t_1 < \cdots < t_n\}$ adapted to (α, V); then we claim that, for all $t_j \in P$, $\Phi(\alpha(t_j)) = F(\alpha(t_j))$. To see this, note that for each j, the (reparametrized to $[0,1]$) path $\alpha|_{[0,t_j]}$ is a continuous path joining e to $\alpha(t_j)$, and $\{t_0 < \cdots < t_j\}$ is a partition adapted to this curve and V. Hence, by definition,

$$\Phi(\alpha(t_j)) = F(\alpha(t_1))F(\alpha(t_1)^{-1}\alpha(t_2)) \cdots F(\alpha(t_{j-1})^{-1}\alpha(t_j)).$$

In the case $j = 1$, this shows $\Phi(\alpha(t_1)) = F(\alpha(t_1))$. Proceeding by induction, suppose we have shown the desired equality for all $j \le k$. Then

$$\begin{aligned}
\Phi(\alpha(t_{k+1})) &= F(\alpha(t_1))F(\alpha(t_1)^{-1}\alpha(t_2)) \cdots F(\alpha(t_{k-1})^{-1}\alpha(t_k))F(\alpha(t_k)^{-1}\alpha(t_{k+1})) \\
&= \Phi(\alpha(t_k))F(\alpha(t_k)^{-1}\alpha(t_{k+1})) \\
&= F(\alpha(t_k))F(\alpha(t_k)^{-1}\alpha(t_{k+1})) \\
&= F(\alpha(t_k) \cdot \alpha(t_k)^{-1}\alpha(t_{k+1})) = F(\alpha(t_{k+1}))
\end{aligned}$$

where we have used the fact that $\alpha(t_k)$ and the increment $\alpha(t_k)^{-1}\alpha(t_{k+1})$ are in V, and F is a local homomorphism on V. Thus, by induction, $\Phi(\alpha(t_j)) = F(\alpha(t_j))$ for all j, and taking $j = n$ yields the desired conclusion $\Phi(g) = \Phi(\alpha(t_n)) = F(\alpha(t_n)) = F(g)$. Thus $\Phi = F$ on V, as desired. $\square$

Let us know summarily collect the elements of the proof of Theorem 14.35.

PROOF OF THEOREM 14.35. First, by Proposition 14.39, there exists a local homomorphism (F, V) such that $\exp$ is one-to-one on $\exp^{-1}(V)$, and $F(\exp X) = \exp \phi(X)$ for $X \in U \equiv \exp^{-1}(V)$. Since G is simply connected, by Proposition 14.40 F extends to a Lie group homomorphism $\Phi\colon G \to H$. Since V is a neighborhood of the identity in G and $\Phi|_V = F$, it follows that $d\Phi_e = dF_e = \phi$; by Lemma 14.36, Φ is therefore the unique Lie group homomorphism with $\Phi_* = \phi$. It then follows from Theorem 13.9(g) that $\Phi(\exp X) = \exp \phi(X)$ for all $X \in \mathfrak{g}$, concluding the proof. $\square$

CHAPTER 15

Quotients and Covering Groups

1. Homogeneous Spaces and Quotient Lie Groups

Let G be any group, and let H be a subgroup (not necessarily normal). The quotient G/H can be made sense of as a set, if not a group: it consists of all left cosets:

$$G/H = \{gH \colon g \in G\}.$$

If two left cosets g_1H and g_2H intersect, meaning there are $h_1, h_2 \in H$ with $g_1h_1 = g_2h_2$, then $g_1h_1h_2^{-1} = g_2$, and since H is a subgroup, $h_1h_2^{-1} \in H$, meaning that $g_2 \in g_1H$. A similar argument shows that $g_1 \in g_2H$, and so $g_1H = g_2H$. Thus, any two left cosets are either equal or disjoint.

Let $\pi \colon G \to G/H$ be the quotient map $\pi(g) = gH$. We will use the notation $\pi(g) = [g]$. Now, if G is a topological space in addition to being a group, then we can give G/H the quotient topology: $U \subseteq G/H$ is declared to be open iff $\pi^{-1}(U) = \{g \in G \colon [g] \in U\}$ is open. Our first result is that, if G is a Lie group and H is a closed subgroup, then G/H is a second-countable Hausdorff space.

LEMMA 15.1. *Let G be a Lie group, and let H be a closed subgroup. Then the quotient topology on G/H is Hausdorff and second-countable.*

The following elementary proof is due to Jonathan Conder. It actually applies in wide generality, to quotients of topological groups by closed subgroups.

PROOF. Let $\pi \colon G \to G/H$ denote the quotient map. First, note that π is an open map: if $U \subseteq G$ is any set, then $\pi^{-1}(\pi(U)) = UH = \bigcup_{h \in H} L_h(U)$. If U is open, so is $L_h(U)$ as L_h is a homeomorphism, so $\pi(U)$ is open in the quotient topology. This immediately gives us the second-countability of G/H: if $\{U_j \colon j \in \mathbb{N}\}$ is a countable base for the topology of G, then $\{\pi(U_j) \colon j \in \mathbb{N}\}$ is a countable base for the topology of G/H.

Now, let $\rho \colon G \times G \to G$ be the continuous function $\rho(x, y) = x^{-1}y$, so two elements x, y are in the same coset in G/H iff $\rho(x, y) \in H$. Since H is closed, $G \setminus H$ is open, and so $\rho^{-1}(G \setminus H)$ is open in $G \times G$. Now, if $\pi(x) \neq \pi(y)$, meaning $\rho(x, y) \notin H$, we have $(x, y) \in \rho^{-1}(G \setminus H)$. Choose a rectangular neighborhood $U \times V$ of (x, y) that is contained in $\rho^{-1}(G \setminus H)$. Thus, for all $u \in U$ and $v \in V$, $u^{-1}v = \rho(u, v) \in G \setminus H$, and so $\pi(u) \neq \pi(v)$. Thus $\pi(U)$ and $\pi(V)$ are disjoint sets in G/H, and they are open since π is an open map. Since $\pi(x) \in \pi(U)$ and $\pi(y) \in \pi(V)$, this proves that any distinct points in G/H have disjoint open neighborhoods, so G/H is Hausdorff. $\square$

Our goal is to show that, in fact, G/H is a smooth manifold (or more precisely that there is a unique smooth structure on G/H such that the quotient map π is smooth). To accomplish this, we prove the following so-called *slice lemma*.

LEMMA 15.2 (Slice Lemma). *Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let H be an embedded (i.e. closed) Lie subgroup of G with Lie algebra $\mathfrak{h} \subseteq \mathfrak{g}$. Let $\mathfrak{f} = \mathfrak{g} \ominus \mathfrak{h}$ (as a vector space), and define a map $\Lambda \colon \mathfrak{f} \times H \to G$ by*

$$\Lambda(X, h) = \exp(X)h.$$

There is a neighborhood U of 0 in $\mathfrak{f}$ such that Λ maps $U \times H$ diffeomorphically onto an open subset of G. In particular, if X, X' are distinct elements of U, then $\exp X$ and $\exp X'$ belong to distinct cosets of G/H.

PROOF. As usual, we may identify the tangent space $T_{(0,e)}\mathfrak{f} \times H$ with $\mathfrak{f} \oplus \mathfrak{h}$, and we also have $T_eG = \mathfrak{g} = \mathfrak{f} \oplus \mathfrak{h}$. Fix $X + Y \in \mathfrak{f} \oplus \mathfrak{h}$, and let $\alpha\colon (-\epsilon, \epsilon) \to \mathfrak{f} \times H$ be a smooth curve with $\alpha(0) = (0, e)$ and $\dot\alpha(0) = X + Y$; then we may write $\alpha(t) = (\alpha_1(t), \alpha_2(t))$ where $\dot\alpha_1(0) = X$ and $\dot\alpha_2(0) = Y$. Then

$$\begin{aligned} d\Lambda_{(0,e)}(X+Y) = \frac{d}{dt}\bigg|_{t=0} \Lambda(\alpha_1(t), \alpha_2(t)) &= \frac{d}{dt}\bigg|_{t=0} \Lambda(\alpha_1(t), e) + \frac{d}{dt}\bigg|_{t=0} \Lambda(0, \alpha_2(t)) \\ &= \frac{d}{dt}\bigg|_{t=0} \exp \alpha_1(t) + \frac{d}{dt}\bigg|_{t=0} \alpha_2(t) = X + Y. \end{aligned}$$

That is: the smooth map Λ has $d\Lambda_{(0,e)} = \mathrm{Id}$, and so by the inverse function theorem there is a neighborhood V of $(0, e)$ in $\mathfrak{f} \times H$ on which Λ is a diffeomorphism onto its image in G. We can choose neighborhoods U_1 of 0 in $\mathfrak{f}$ and U_2 of e in H with $U_1 \times U_2 \subseteq V$, and so in particular $d\Lambda_{(X,e)}$ is invertible for all $X \in U_1$.

Now, note that, for all $h, \in H$ and $X \in \mathfrak{f}$, $R_h\Lambda(X, e) = \Lambda(X, e)h = \Lambda(X, h)$. It then follows that $dR_h \circ d\Lambda_{(X,e)} = d\Lambda_{(X,h)}$, and so since R_h is a diffeomorphism, invertibility of $d\Lambda_{(X,e)}$ implies invertibility of $d\Lambda(X, h)$. Thus $d\Lambda_{(X,h)}$ is invertible for all $(X, h) \in U_1 \times H$. By the inverse function theorem, each point in $U_1 \times H$ has a neighborhood that is mapped by Λ diffeomorphically onto its image. It follows that $\Lambda(U_1 \times H)$ is open in G.

Now, let U_1, U_2 be as above, so Λ maps $U_1 \times U_2$ diffeomorphically onto its image in G. Let V' be a neighborhood of e in G such that $V' \cap H = U_2$; such a neighborhood exists because H is embedded in G. Let $U \subseteq U_1$ be small enough that, for all $X, X' \in U$, $\exp(-X') \exp X \in V'$; this is possible by continuity of $\exp$. Thus, if $X, X' \in U$ are such that $\exp(-X') \exp X \in H$, then in fact $\exp(-X') \exp X \in V' \cap H = U_2$. It now follows that Λ is injective on $U \times H$. Indeed, suppose $(X, h), (X', h') \in U \times H$ and $\Lambda(X, h) = \Lambda(X', h')$. That is: $\exp X \cdot h = \exp X' \cdot h'$, and so

$$h'h^{-1} = \exp(-X') \exp X \in U_2$$

by our choice of U. Thus $\Lambda(X, e) = \exp X = \exp X' \cdot (h'h^{-1}) = \Lambda(X', h'h^{-1})$. Since both (X, e) and $(X', h'h^{-1})$ are in $U_1 \times U_2 \subset V$ where Λ is injective, it follows that $X = X'$ and $e = h'h^{-1}$, which shows that $(X, h) = (X', h')$ as desired.

Thus, Λ is a local diffeomorphism on $U_1 \times H$ and is injective on $U \times H$; thus it is a diffeomorphism onto its (open) image $\Lambda(U \times H)$. In particular, this means that if $X, X' \in U$ are such that the cosets $\exp X \cdot H$ and $\exp X' \cdot H$ are not distinct, there are $h, h' \in H$ with $\Lambda(X, h) = \exp X \cdot h = \exp X' \cdot h' = \Lambda(X', h')$, and hence by injectivity $X = X'$ (and $h = h'$). Thus distinct elements of U are in distinct cosets of G/H. □

REMARK 15.3. Without the assumption that H is a closed subgroup, the slice lemma can be false: it can happen, for example, that $\exp X \in H$ even for arbitrarily small $X \in \mathfrak{f}$. The subtlety is the proof of injectivity of Λ: if H is not embedded, then it is not generally possible to choose a neighborhood U so the $\exp(-X') \exp X \in U_2$ whenever $X, X' \in U$ satisfy $\exp(-X') \exp X \in H$. It is, of course, possible to make $\exp(-X') \exp X$ fall in some arbitrarily small neighborhood V' of e in G, but if H is not a closed (i.e. embedded) subgroup, any such small neighborhood V' need not intersect H in a nice way. (Think, for example, of the case that H is dense in G.)

The Slice Lemma allows us to define a smooth manifold structure on G/H.

THEOREM 15.4. *Let G be a Lie group, and let $H \subseteq G$ be a closed subgroup; let $\mathfrak{g}$ and $\mathfrak{h} \subseteq \mathfrak{g}$ denote their Lie algebras. Then there is a unique smooth structure on G/H such that the quotient map $\pi\colon G \to G/H$ is a smooth submersion. The smooth manifold G/H satisfies the following properties:*

(1) $\dim(G/H) = \dim(G) - \dim(H)$,
(2) $d\pi_e\colon \mathfrak{g} \to T_{[e]}G/H$ *has kernel* $\mathfrak{h}$, *and*
(3) *the left action of G on G/H ($g \cdot [x] = [gx]$) is smooth and transitive.*

Before proceeding to the proof, we need the following general property of submersions in order to prove uniqueness.

LEMMA 15.5. *Suppose M, N are smooth manifolds, and $\pi\colon M^m \to N^n$ is a surjective submersion. If P as another smooth manifold, then a map $F\colon N \to P$ is smooth if and only if $F \circ \pi$ is smooth:*

$$\begin{tikzcd} M \arrow[d, "\pi"'] \arrow[dr, "F\circ\pi"] & \\ N \arrow[r, "F"'] & P \end{tikzcd}$$

PROOF. If F is smooth then $F \circ \pi$ is a composition of smooth maps, hence is smooth. Conversely, suppose $F \circ \pi$ is smooth, and let $q \in N$. As π is surjective, there is some $p \in \pi^{-1}(q)$. By the Rank Theorem, we can choose a chart $(U, \varphi = \{x^j\}_{j=1}^m)$ at p and a chart $(V, \psi = \{y^k\}_{k=1}^n)$ at q such that $\widehat{\pi} = \psi \circ \pi \circ \varphi^{-1}$ has the form $\pi(x^1, \ldots, x^n, x^{n+1}, \ldots, x^m) = (y^1, \ldots, y^m)$. Now, let $\epsilon > 0$ be small enough that the coordinate cube $C_\epsilon^m = \{x\colon |x^j| < \epsilon, 1 \le j \le m\}$ is contained in U; then $\widehat{\pi}(C_\epsilon^m)$ is the coordinate cube $\widehat{\pi}(C_\epsilon^m) = C_\epsilon^n = \{y\colon |y^k| < \epsilon, 1 \le j \le n\}$. Then we may define $\widehat{\sigma}\colon C_\epsilon^n \to C_\epsilon^m$ by $\widehat{\sigma}(x^1, \ldots, x^n) = (x^1, \ldots, x^n, 0, \ldots, 0)$, and let $\sigma = \varphi^{-1} \circ \widehat{\sigma} \circ \psi|_{\psi^{-1}(C_\epsilon^n)}$. Note that $\sigma(q) = p$, and σ is a local section for π: $\pi \circ \sigma = \mathrm{Id}$ on the neighborhood $V' = \psi^{-1}(C_\epsilon^n)$ of q where σ is defined.

Now, restricting F to V', we have

$$F|_{V'} = F|_{V'} \circ \mathrm{Id}_{V'} = F|_{V'} \circ (\pi \circ \sigma) = (F \circ \pi) \circ \sigma$$

and this is a composition of smooth maps. Thus F is smooth on a neighborhood V' of q. As this holds for each $q \in N$, F is smooth. □

REMARK 15.6. Embedded in the above proof is (the harder half of) the *local section theorem* for submersions: every submersion has a local section at every point in its image. The easy direction is that this characterizes submersions: if $\pi \circ \sigma = \mathrm{Id}_{V'}$ on some neighborhood of a point $q = \pi(p)$, then $d\pi_p \circ d\sigma_q = \mathrm{Id}_{T_qN}$, which shows that $d\pi_p$ is surjective.

PROOF OF THEOREM 15.4. First, uniqueness: suppose G/H has two smooth structures (call them N_1 and N_2) such that $\pi\colon G \to N_j$ is a smooth submersion for $j = 1, 2$. Then consider the identity map $\mathrm{Id}\colon N_1 \to N_2$; by Lemma 15.5, this map is smooth since $\mathrm{Id} \circ \pi = \pi\colon G \to N_2$ is assumed to be smooth. Similarly, $\mathrm{Id}\colon N_2 \to N_1$ is smooth. Thus Id is a diffeomorphism, and so the two smooth structures are actually the same.

Now, we proceed to show that a smooth structure exists (and satisfies properties (1)-(3)); this is where the Slice Lemma 15.2 comes in. First we shift the map Λ in that Lemma by defining, for each $g \in G$, $\Lambda_g = L_g \circ \Lambda\colon U \times H \to G$:

$$\Lambda_g(X, h) = g \exp(X) h$$

where $\mathfrak{f} = \mathfrak{g} \ominus \mathfrak{h}$ and $U \subset \mathfrak{f}$ is a neighborhood of 0 in $\mathfrak{f}$. This is a composition of diffeomorphisms, so Λ_g is a diffeomorphism from $U \times H$ onto its image, and $\Lambda_g(X, h)$ and $\Lambda_g(X', h')$ are in distinct cosets of H whenever $X \neq X'$ (for any $g \in G$ and $h, h' \in H$). Let $W_g = \Lambda_g(U \times H)$ and let $V_g = \pi(W_g)$:

$$V_g = \{[g \exp X] \in G/H \colon X \in U\}.$$

Since W_g is open and G/H has the quotient topology, V_g is open in G/H. Note also that $\pi^{-1}(V_g) = W_g$. The argument above shows that the map $X \mapsto [g \exp X]$ is injective from U to V_g.

We use the maps $X \mapsto [g \exp X]$ is (inverse) local coordinates. Note that $U \subseteq \mathfrak{f} \cong \mathbb{R}^k$ where $k = \dim\mathfrak{f} = \dim\mathfrak{g} - \dim\mathfrak{h}$. The map $X \mapsto [g \exp X]$ is an injective map from U onto V_g, and it is continuous (it is the composition of the quotient map π with the a continuous map). Let $\varphi_g \colon V_g \to U$ denote its inverse. We consider the composition $\varphi_g \circ \pi \colon W_g \to U$. Note that, for any point $\Lambda_g(X, h) \in W_g$,

$$(\varphi_g \circ \pi)(\Lambda_g(X, h)) = \varphi_g([g \exp Xh]) = \varphi_g[g \exp X] = X.$$

Thus, $\varphi \circ \pi = \pi_1 \circ \Lambda_g^{-1}$ where $\pi_1 \colon U \times H \to U$ is the projection $\pi_1(X, h) = X$. Since Λ_g is a diffeomorphism $U \times H \to W_g$, its inverse $\Lambda_g^{-1} \colon W_g \to U \times H$ is continuous, and so too is the projection map. Thus $\varphi_g \circ \pi$ is continuous. It follows that φ_g is continuous: if $B \subseteq U$ is open, then $\pi^{-1}(\varphi^{-1}(B)) = (\varphi \circ \pi)^{-1}(B)$ is open in W_g, and thus by definition of the quotient topology $\varphi^{-1}(B)$ is open in $V_g = \pi(W_g)$.

Thus, for each $g \in G$, there is a neighborhood V_g of $[g]$ in G/H and a homeomorphism $\varphi_g \colon V_g \to U \subseteq \mathfrak{f} \cong \mathbb{R}^k$. (Note: it is not necessarily true that (V_g, φ_g) and $(V_{g'}, \varphi_{g'})$ are the same if $[g] = [g']$; rather, we simply select some representative element from each equivalence class $[g]$ and use it to define the chart.) Since G/H is also Hausdorff and second countable (by Lemma 15.1), we have thus shown that G/H is a topological manifold, of dimension $k = \dim\mathfrak{g} - \dim\mathfrak{h} = \dim(G) - \dim(H)$, verifying (1).

Now let us compute the transition map $\varphi_{g'} \circ \varphi_g^{-1} \colon \varphi_g(V_g \cap V_{g'}) \to \varphi_{g'}(V_g \cap V_{g'})$. by definition $\varphi_g^{-1}(X) = [g \exp X]$. Hence the transition map can be described as $X \mapsto X'$ where $[g \exp X] = [g' \exp X']$. This may be accomplished as follows: first map $X \mapsto g \exp X$. Since $X \in \varphi_g(V_{g'})$, this means $[g \exp X] = \pi(g \exp X) \in V_{g'} = \pi(W_{g'})$, and since $\pi^{-1}(V_{g'}) = W_{g'}$, it follows that $g \exp X \in W_{g'}$. Now $\Lambda_{g'}^{-1} \colon W_{g'} \to U \times H$ is a diffeomorphism, and so $\Lambda_{g'}^{-1}(g \exp X) = (X', h')$ for some unique $X' \in U$ and $h' \in H$, and by definition $\Lambda_{g'}(X', h') = g' \exp X' h'$. That is: $g \exp X = g' \exp X' h'$, and so $[g \exp X] = [g' \exp X']$ as required. Thus, the transition map $X \mapsto X'$ can be computed as

$$X \mapsto g \exp X \xrightarrow{\Lambda_{g'}^{-1}} (X', h') \xrightarrow{\pi_1} X'$$

i.e. $\varphi_{g'} \circ \varphi_g^{-1} = \pi_1 \circ \Lambda_{g'}^{-1} \circ L_g \circ \exp$ is a composition of smooth maps, hence is smooth. We have thus shown that the collection $\{(V_g, \varphi_g)\}$ is a smooth atlas, and so it makes G/H into a smooth manifold.

Now let us show that π is a submersion. For any given $g \in G$, on the neighborhood W_g of g, we have $\varphi_g \circ \pi = \pi_1 \circ \Lambda_g^{-1}$ as computed above, and this is smooth. Thus, relative to the smooth structure given by the charts (V_g, φ_g), π is a smooth map. Moreover, at a point g' near g, we can apply the chain rule to $\varphi_g \circ \pi = \pi_1 \circ \Lambda_g^{-1}$ to yield

$$(d\varphi_g)_{\pi(g')} \circ d\pi_{g'} = d(\varphi_g \circ \pi)_{g'} = (d\pi_1 \circ \Lambda_g^{-1})_{g'} = (d\pi_1)_{\Lambda_g^{-1}(g')} \circ d(\Lambda_g^{-1})_{g'}. \tag{15.1}$$

Since φ_g and Λ_g^{-1} are diffeomorphisms, it follows that $d\pi_{g'}$ has the same rank as $(d\pi_1)_{\Lambda_g^{-1}(g')}$, which is constant: at any point $(X,h) \in \mathfrak{f} \times H$, $\pi_1(X,h) = X$ and so $d\pi_1|_{(X,h)}$ is a projection of rank $\dim\mathfrak{f} = \dim(G/H)$. This shows that π is indeed a submersion, and since it is surjective, by the uniqueness argument at the beginning, the smooth structure we defined here is the unique smooth structure on G/H.

It remains to verify (2) and (3). Take $g = g' = e$ in (15.1), and note that $\Lambda_e^{-1}(e) = (0,e)$, and that $d(\Lambda_e^{-1})_e = (d\Lambda_e|_{(0,e)})^{-1} = \mathrm{Id}_{\mathfrak{f}\oplus\mathfrak{h}}$ as proved in the Slice Lemma 15.2. Thus (15.1) becomes

$$(d\varphi_e)_{[e]} \circ d\pi_e = (d\pi_1)_{(0,e)}.$$

As $\mathfrak{h} = \ker(d\pi_1)_{(0,e)}$ and $(d\varphi_e)_{[e]}$ is an isomorphism, we conclude that $\ker(d\pi_e) = \mathfrak{h}$, establishing (2). Finally, let $\theta\colon G \times G/H \to G/H$ denote the action $\theta(g,[x]) = g\cdot[x] = [gx] = \pi(gx)$. Fix a point $g_0 \in G$ and look at θ in the chart (V_{g_0}, φ_{g_0}); i.e. consider the map $\widehat{\theta}\colon G\times U \to G/H$ given by $\widehat{\theta}(g,X) = \theta(g, \varphi_{g_0}^{-1}(X))$. Then θ is smooth near g_0 iff $\widehat{\theta}$ is smooth. Since $\varphi_{g_0}^{-1}(X) = [g_0 \exp X]$, we have

$$\widehat{\theta}(g,X) = g\cdot[g_0\exp X] = [gg_0\exp X] = \pi(gg_0\exp X)$$

which is the composition $\pi\circ \mathrm{m}\circ(R_{g_0},\exp)$, a smooth map. Thus, the left action of G on G/H is smooth. It is also transitive for the same reason the left action of G on itself is: if $g_0, g_1 \in G$, then $g_1 g_0^{-1}\cdot[g_0] = [g_1]$. This verifies (3), and concludes the proof. □

The smooth manifold G/H is called a **homogeneous space**. Homogeneous spaces are often defined alternatively in terms of group actions: a smooth manifold M is called a **homomgeneous G-space** if G acts smoothly and transitively on M. If we then choose a given basepoint $p \in M$ and let $H = G_p$ (the stabilizer / isotropy subgroup of p), then it turns out that G/H is naturally diffeomorphic to M, as the following result shows.

PROPOSITION 15.7 (Homogeneous Space Characterization Theorem). *Let G be a Lie group, and let M be a homogeneous G-space. Fix a point $p \in M$. Then the isotropy subgroup $H = G_p$ is a closed subgroup of G, and the map $F\colon G/H \to M$ defined by $F([g]) = g\cdot p$ is a (well-defined) equivariant diffeomorphism.*

PROOF. As usual, denote the action by $g\cdot p = \theta(g,p) = \theta^{(p)}(g)$. Then the stabilizer subgroup $H = G_p$ is just $H = (\theta^{(p)})^{-1}(p)$, and by continuity this is a closed subgroup of G. Thus G/H is a smooth manifold, as per Theorem 15.4, and the quotient map $g \mapsto [g]$ is a submersion.

Now, define $F([g]) = g\cdot p$. First we must see this is well-defined. So suppose $g_1, g_2 \in G$ satisfy $[g_1] = [g_2]$. Thus $g_2 \in [g_1]$, so there is some $h \in H$ with $g_2 = g_1 h$. It follows that

$$g_2\cdot p = g_1 h\cdot p = g_1\cdot(h\cdot p) = g_1\cdot p$$

since H fixes p. Thus $F([g_1]) = F([g_2])$, and F is well-defined. Moreover, it is equivariant, since for any $g, g' \in G$,

$$F(g'\cdot[g]) = F([g'g]) = (g'g)\cdot p = g'\cdot(g\cdot p) = g'\cdot F([g]).$$

Now, note that $F\circ\pi\colon G \to M$ is the map $F\circ\pi(g) = g\cdot p$, which is smooth; hence, by Lemma 15.5, F is smooth. What's more, it is injective: if $g, g' \in G$ and $F([g]) = F([g'])$ then $g\cdot p = g'\cdot p$ whic hshows that $g^{-1}g'\cdot p = p$. Thus $g^{-1}g' \in G_p = H$, and so $[g] = gH = g'H = [g']$.

So far, everything holds true for a generic smooth action (so in that level of generality F is an equivariant smooth injection). Now, as we assume the action is transitive, for any point $q \in M$ there is some $g \in G$ with $F([g]) = g\cdot p = q$; thus F is surjective. Thus, F is an equivariant smooth bijections. By the Equivariant Rank Theorem 11.21 and the transitivity of the action, F

has constant rank. Thus, F is a constant rank smooth bijection, and so by the Global Rank Theorem 9.10, it is a diffeomorphism. $\square$

REMARK 15.8. Note that a different basepoint may have a different isotropy subgroup, so this identification depends on the chosen basepoint (although the subgroups G_p and G_q are related by an inner automorphism $G_q = gG_pg^{-1}$ for any g with $g \cdot p = q$). In fact, one approach to defining the smooth manifold structure on G/H is to go this route and prove homogeneous spaces, defined by group actions, are smooth manifolds.

EXAMPLE 15.9. Consider the Lie group $G = \mathrm{O}(n)$, viewed in the usual way as orthogonal transformations of $\mathbb{R}^n$. This gives it a smooth action on $\mathbb{R}^n$, but this action is not transitive: if $x, y \in \mathbb{R}^n$ are in the same orbit of G, then $|x| = |y|$. It is also easy to see that this is iff: any two vectors of the same length can be interchanged by an orthogonal transformation. Thus, the restriction of the action to $\mathbb{S}^{n-1}$ is transitive (and is smooth since $\mathbb{S}^{n-1}$ is an embedded submanifold of $\mathbb{R}^n$).

Consider the subgroup H of orthogonal transformations that fix the north pole $e_n \in \mathbb{S}^{n-1}$. As matrices, these have the form

$$H = \left\{ \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} : Q \in \mathrm{O}(n-1) \right\}$$

as can be easily verified. This shows H is isomorphic as a Lie group to $\mathrm{O}(n-1)$. Hence, by Proposition 15.7, the map $F\colon G/H \to \mathbb{S}^{n-1}$ given by $F([A]) = Ae_n$ is an equivariant diffeomorphism, and we have $\mathrm{O}(n)/\mathrm{O}(n-1) \cong \mathbb{S}^{n-1}$.

It is instructive to see how this works explicitly in this example. First, it is clear from its form that H is a closed subgroup isomorphic to $\mathbb{O}(n-1)$. By Theorem 15.4, $G/H = \mathrm{O}(n)/\mathrm{O}(n-1)$ has a unique smooth structure with respect to which the quotient map is a submersion, making it into a manifold of dimension

$$\dim \mathrm{O}(n) - \dim \mathrm{O}(n-1) = \frac{n(n-1)}{2} - \frac{(n-1)(n-2)}{2} = n-1.$$

We define a map $F\colon \mathrm{O}(n)/\mathrm{O}(n-1) \to \mathbb{S}^{n-1}$ as follows: note that, for any $A \in \mathrm{O}(n)$, the coset $A \cdot \mathrm{O}(n-1)$ consists of all matrices of the form

$$[A] = A \cdot \mathrm{O}(n-1) = \left\{ A \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} A \begin{bmatrix} Q \\ 0 \end{bmatrix} & Ae_n \end{bmatrix} : Q \in \mathrm{O}(n-1) \right\}.$$

Thus every matrix in $[A]$ has the same final column: Ae_n. We thus define the map as $F([A]) = Ae_n$, which is in $\mathbb{S}^{n-1}$ since all columns of the orthogonal matrix $\mathrm{O}(n)$ are orthonormal vectors. Letting $\pi\colon \mathrm{O}(n) \to \mathrm{O}(n)/\mathrm{O}(n-1)$ denote that quotient map, we see that the map $F \circ \pi\colon \mathrm{O}(n) \to \mathrm{O}(n)/\mathrm{O}(n-1)$ is $(F \circ \pi)(A) = Ae_n$, and this is smooth; thus, since π is a surjective submersion, F is smooth by Lemma 15.5. It is surjective since any vector $v \in \mathbb{S}^{n-1}$ can be completed to an orthonormal basis of $\mathbb{R}^n$, and the matrix A with this basis as columns (v last) is orthogonal satisfying $Ae_n = v$. F is also injective: if $F([A]) = F([B])$, then $Ae_n = Be_n$, and so $B^{-1}A$ is an orthogonal matrix that fixes e_n, meaning that it is in H. Thus $A \in BH$, and so $[A] = [B]$.

Thus, we have defined a smooth bijection $F\colon \mathrm{O}(n)/\mathrm{O}(n-1) \to \mathbb{S}^{n-1}$. Finally, note that for any $A \in \mathrm{O}(n)$, $d(F \circ \pi)_A = dF_{[A]} \circ d\pi_A$, and since $F \circ \pi$ is the restriction to $O(n)$ of the linear map $A \mapsto Ae_n$, it is constant rank. As π is a submersion, it is constant rank, and so it follows that F has constant rank. A smooth bijection of constant rank is a diffeomorphism by the Global Rank Theorem 9.10, and so we finally conclude that

$$\mathrm{O}(n)/\mathrm{O}(n-1) \cong \mathbb{S}^{n-1}.$$

EXAMPLE 15.10. A similar argument to the one given in Example 15.9 shows that

$$\mathrm{SO}(n)/\mathrm{SO}(n-1) \cong \mathbb{S}^{n-1}$$

(with the same realization of the subgroup). The only slight subtlety is keeping track of the orientation of the orthonormal basis in the proof that the map F is surjective.

Now, in Theorem 15.4, if H happens to be a normal subgroup, then we get the following.

COROLLARY 15.11. *Let G be a Lie group, and let $H \subseteq G$ be a closed normal subgroup; let $\mathfrak{g}$ and $\mathfrak{h} \subseteq \mathfrak{g}$ denote their Lie algebras. Then the homogeneous space G/H is a Lie group of dimension $\dim(G) - \dim(H)$, and the quotient map $\pi\colon G \to G/H$ is a Lie group homomorphism whose induced Lie algebra homomorphism $\pi_*\colon \mathfrak{g} \to \mathrm{Lie}(G/H)$ has kernel $\ker\pi_* = \mathfrak{h}$.*

PROOF. Since H is a normal subgroup, the coset space G/H is a group under the product $[g] \cdot [g'] \equiv [gg']$. Thus, the multiplication map $\mathrm{m}_{G/H}$ is satisfies $\mathrm{m}_{G/H} \circ (\pi \times \pi) = \pi \circ \mathrm{m}_G$ which is smooth, and so by Lemma 15.5 applied to the surjective submersion $\pi \times \pi$, $\mathrm{m}_{G/H}$ is smooth. Thus G/H is a Lie group. The quotient map is automatically a group homomorphism, and it is smooth, hence it is a Lie group homomorphism. The final statement (regarding the kernel of π_*) is precisely Theorem 15.4(2). □

EXAMPLE 15.12. In Examples 15.9 and 15.10, we see that the given identification of $\mathrm{O}(n-1)$ or $\mathrm{SO}(n-1)$ as a subgroup of $\mathrm{O}(n)$ or $\mathrm{SO}(n)$ cannot be a normal subgroup if $n-1 \notin \{1,3,7\}$. This is because $\mathbb{S}^{n-1}$ is not parallelizable unless $n \in \{1,3,7\}$ (cf. the discussion following Proposition 4.10), and Lie groups are always parallelizable (cf. Proposition 12.4). In fact, it is a deeper fact that $\mathbb{S}^7$ is not a Lie group, even though it is parallelizable. All that being said, it is easy to see that $\mathrm{O}(n-1)$ is never a normal subgroup of $\mathrm{O}(n)$ with the given identifications.

Perhaps the most useful special case of Corollary 15.11 is the case when the closed subgroup H is a discrete normal subgroup.

COROLLARY 15.13. *Let G be a Lie group, and let $H \subseteq G$ be a discrete normal subgroup (i.e. a 0-dimensional Lie group). Then the quotient map $\pi\colon G \to G/H$ is a surjective local diffeomorphism, and $\pi_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G/H)$ is a Lie algebra isomorphism.*

PROOF. Since $\mathrm{Lie}(H) = \{0\}$, $\ker(d\pi_e) = \{0\}$, and so π_* is a Lie algebra isomorphism. Note that $\dim(G/H) = \dim(G) = \mathrm{rank}(d\pi_e)$. As π is a Lie group homomorphism, it has constant rank, and so $\mathrm{rank}(d\pi_g) = \mathrm{rank}(d\pi_e) = \dim(G) = \dim(G/H)$ at all points $g \in G$, which shows that π is a local diffeomorphism. □

EXAMPLE 15.14. Consider the Lie group homomorphism $\varepsilon\colon \mathbb{R} \to \mathbb{S}^1$ given by $\varepsilon(x) = e^{2\pi i x}$; it is surjective, and the kernel is the normal subgroup $\mathbb{Z} \subset \mathbb{R}$. Thus, by the first isomorphism theorem for groups, $\mathbb{S}^1$ is isomorphic (as a group) to $\mathbb{R}/\mathbb{Z}$ via the isomorphism $\Phi\colon \mathbb{R}/\mathbb{Z} \to \mathbb{S}^1\colon [x] \mapsto \varepsilon(x)$. That is: $\Phi \circ \pi = \varepsilon$, which is smooth by Lemma 15.5, and hence Φ is smooth, so it is a Lie group isomorphism. Of course, both of these are abeliean groups, with Lie algebra isomorphic to $\mathbb{R}$.

The next example is so important that it gets its own section.

2. Interlude: $\mathrm{SU}(2)$ and $\mathrm{SO}(3)$

As a perfect illustration of the structure of quotient Lie groups (by discrete subgroups), we consider now a quotient of $\mathrm{SU}(2)$. First, let us note that this is a connected group; in fact, $\mathrm{U}(n)$ and $\mathrm{SU}(n)$ are connected for all n.

LEMMA 15.15. *For all $n \in \mathbb{N}$, the Lie groups $\mathrm{U}(n)$ and $\mathrm{SU}(n)$ are connected.*

PROOF. Let U be a unitary matrix in $\mathrm{U}(n)$ (resp. $\mathrm{SU}(n)$). By the spectral theorem, it has a decomposition $U = VDV^{-1}$ for some $V \in U(n)$, and diagonal D containing the eigenvalues of U, which are all in $\mathbb{S}^1$: $D = \mathrm{diag}[e^{i\theta_1}, \dots, e^{i\theta_n}]$ for some $\theta_1, \dots, \theta_n \in \mathbb{R}$. Accordingly, we define a continuous curve $U(t)_{0 \le t \le 1}$ as follows:

$$U(t) = V \begin{bmatrix} e^{it\theta_1} & 0 & \cdots & 0 \\ 0 & e^{it\theta_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{it\theta_n} \end{bmatrix} V^{-1}, \quad 0 \le t \le 1.$$

This is a continus curve, and $U(t)$ is unitary (being a product of unitary matrices). Since $U(0) = I$ and $U(1) = U$, we see that U is in the same path component as I in $\mathrm{U}(n)$, which shows $\mathrm{U}(n)$ is connected, as claimed. If we additionally have $U \in \mathrm{SU}(n)$, we compute the determinant of $U(t)$ thus:

$$\det U(t) = e^{it\theta_1} e^{it\theta_2} \cdots e^{it\theta_n} = e^{i(1-t)(\theta_1+\theta_2+\cdots+\theta_n)} = (e^{i\theta_1} e^{i\theta_2} \cdots e^{i\theta_n})^t = (\det U)^t = 1$$

so $U(t)$ is a curve in $\mathrm{SU}(2)$ connecting I to U, yielding the same result. □

It will be convenient to have a "Euclidean" inner product on the space $\mathbb{M}_n(\mathbb{C})$ of matrices in which $\mathrm{SU}(n)$ lives. The *Hilbert-Schmidt inner product* is given by

$$\langle A, B \rangle \equiv \frac{1}{n} \mathrm{Tr}(B^* A).$$

This is clearly sesquilinear, and we have $\langle A, A \rangle = \mathrm{Tr}(A^* A) = \sum_{i,j} |A_{ij}|^2 = 0$ iff $A = 0$, so it is an inner product. (This is the *normalized* Hilbert-Schmidt inner product, so that $\langle I, I \rangle = 1$; many authors choose not to normalize in the definition.) It therefore descends to an inner product on any subspace of $\mathbb{M}_n(\mathbb{C})$. The Hilbert space topology given by this inner product is (of course) the usual topology on $\mathbb{M}_n(\mathbb{C})$, and thence also on any subspace.

One use for this is to give a trivial proof that the (special) unitary and orthogonal groups are compact.

LEMMA 15.16. *For all $n \ge 1$, the groups $\mathrm{U}(n)$, $\mathrm{SU}(n)$, $\mathrm{O}(n)$, and $\mathrm{SO}(n)$ are compact.*

PROOF. The group $\mathrm{U}(n)$ is closed in $\mathbb{M}_n(\mathbb{C})$: it is a level set of the continuous function $U \mapsto U^* U$. Now we compute that, for $U \in \mathrm{U}(n)$,

$$\langle U, U \rangle = \frac{1}{n} \mathrm{Tr}(U^* U) = \frac{1}{n} \mathrm{Tr}(I) = 1$$

so $\mathrm{U}(n)$ is contained in the Hilbert-Schmidt unit sphere in $\mathbb{M}_n(\mathbb{C})$, which is compact. Thus $\mathrm{U}(n)$ is compact, and $\mathrm{SU}(n)$ is a closed subset of $\mathrm{U}(n)$ (given as a level set of the continuous function $\det$), so it is also compact. Analogous arguments apply to $\mathrm{O}(n)$ and $\mathrm{SO}(n)$. □

Now let us focus on the group $\mathrm{SU}(2)$. The Lie algebra of $\mathrm{SU}(2)$, identified as the tangent space at the identity $\mathfrak{su}(2)$, consists of those 2×2 complex matrices A that are skew Hermitian $A^* = -A$ and have trace 0. This is a Lie algebra of (real) dimension 3: a convenient basis is given by

$$\beta = \left\{ X = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, Z = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \right\}.$$

It is straightforward to compute the brackets

$$[X,Y] = 2Z, \quad [X,Z] = -2Y, \quad [Y,Z] = -2X. \tag{15.2}$$

Now, the Hilbert-Schmidt inner product restricted to $\mathfrak{su}(2)$ simplifies due to the condition $A^* = -A$, giving

$$\langle A, B\rangle = -\frac{1}{2}\mathrm{Tr}(BA) = -\frac{1}{2}\mathrm{Tr}(AB), \qquad A, B \in \mathfrak{su}(2),$$

and in fact is a real inner product. It is easy to check by direct computation that β is an orthonormal basis for $\mathfrak{su}(2)$ with respect to this inner product (since $X^2 = Y^2 = Z^2 = -I$ and $XY = Z$, $XZ = -Y$, and $YZ = X$).

Now, consider the Adjoint homomorphism $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{GL}(\mathfrak{su}(2))$. In this case, since $\mathrm{SU}(2)$ is a matrix Lie group, this is simply the conjugation map

$$\mathrm{Ad}(U)A = UAU^{-1} = UAU^*, \qquad U \in \mathrm{SU}(2),\ A \in \mathfrak{su}(2).$$

We consider the image of the homomorphism Ad. We can compute that, for any $A, B \in \mathfrak{su}(2)$ and $U \in \mathrm{SU}(2)$,

$$\begin{aligned}\langle \mathrm{Ad}(U)A, \mathrm{Ad}(U)B\rangle = -\frac{1}{2}\mathrm{Tr}(\mathrm{Ad}(U)A \cdot \mathrm{Ad}(U)B) &= -\frac{1}{2}\mathrm{Tr}(UAU^{-1}UBU^{-1}) \\ &= -\frac{1}{2}\mathrm{Tr}(AB) = \langle A, B\rangle\end{aligned}$$

where we have canceled the $U^{-1}U$ and also used the cyclic trace property to yield $\mathrm{Tr}(UABU^{-1}) = \mathrm{Tr}(U^{-1}UAB) = \mathrm{Tr}(AB)$. This shows that $\mathrm{Ad}(U)$ is actually an orthogonal transormation of the Hilbert space $\mathfrak{su}(2)$, and so $\mathrm{Ad}(\mathrm{SU}(2)) \subseteq \mathrm{O}(\mathfrak{su}(2))$. What's more, since $\mathrm{SU}(2)$ is connected, the image of Ad in $\mathrm{O}(\mathfrak{su}(2))$ (which contains I) must be contained in the component of the group containing I, which is $\mathrm{SO}(\mathfrak{su}(2))$. Thus, Ad is actually a homomorphism

$$\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{SO}(\mathfrak{su}(2)).$$

We will see that, in fact, this homomorphism is surjective; for the time being, we simply denote the image as $\mathrm{Ad}(\mathrm{SU}(2)) \subseteq \mathrm{SO}(\mathfrak{su}(2))$. (This is called the *Adjoint group* of $\mathrm{SU}(2)$.) It is the image of a homomorphism, so it is a subgroup of $\mathrm{SO}(\mathfrak{su}(2))$. Since Ad is continuous and $\mathrm{SU}(2)$ is compact, the image $\mathrm{Ad}(\mathrm{SU}(2))$ is compact. In particular, it is a closed subgroup of $\mathrm{SO}(\mathfrak{su}(2))$, and so by the Closed Subgroup Theorem 14.5, $\mathrm{Ad}(\mathrm{SU}(2))$ is an embedded Lie subgroup of $\mathrm{SO}(\mathfrak{su}(2))$.

Thus, we have a surjective Lie group homomorphism $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{Ad}(\mathrm{SU}(2))$. Let us now compute its kernel. A matrix $U \in \mathrm{SU}(2)$ is in $\ker(\mathrm{Ad})$ iff $\mathrm{Ad}(U) = \mathrm{Id}_{\mathfrak{su}(2)}$, meaning that for all $A \in \mathfrak{su}(2)$ $A = \mathrm{Ad}(U)A = UAU^{-1}$. That is to say: if U is in the kernel of Ad, then $AU = UA$ for all $A \in \mathfrak{su}(2)$. This forces U to be $\pm I$.

LEMMA 15.17. *If U is any matrix in $\mathbb{M}_2(\mathbb{C})$ that commutes with all matrices in $\mathfrak{su}(2)$, then $U = \lambda I$ for some $\lambda \in \mathbb{C}$. Thus, if $U \in \mathrm{SU}(2)$, $U = \pm I$.*

PROOF. Take the commutation relation $UA = AU$ over the three matrices $A \in \beta$. This gives a collection of linear equations for the entries of U. The condition $UX = XU$ yields the two independent equations $u_{21} = -u_{12}$ and $u_{11} = u_{22}$; letting $\lambda = u_{11}$ and $\mu = u_{12}$, this shows U has the form

$$U = \begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}.$$

Now the commutation relation $UY = YU$ gives the new equation $i\mu = -i\mu$, so $\mu = 0$, and we conclude that $U = \lambda I$ as claimed. Now, if $U \in \mathrm{SU}(2)$, we have in particular $1 = \det U = \det(\lambda I) = \lambda^2$, and so $\lambda = \pm 1$. $\square$

Thus, $\ker \mathrm{Ad} = \{I, -I\} \subset \mathrm{SU}(2)$, which is easily seen to be a normal subgroup. Hence, we have the chain

$$\mathbb{Z}_2 \cong \{I, -I\} \hookrightarrow \mathrm{SU}(2) \xrightarrow{\mathrm{Ad}} \mathrm{Ad}(\mathrm{SU}(2))$$

and so by the first isomorphism theorem, $\mathrm{Ad}(\mathrm{SU}(2))$ is isomorphic to $\mathrm{SU}(2)/\mathbb{Z}_2$, via the isomorphism $\Phi\colon \mathrm{SU}(2)/\mathbb{Z}_2 \to \mathrm{Ad}(\mathrm{SU}(2))\colon [U] \mapsto \mathrm{Ad}U$. This map is defined so that $\Phi \circ \pi = \mathrm{Ad}$ where $\pi\colon \mathrm{SU}(2) \to \mathrm{SU}(2)/\mathbb{Z}_2$ is the quotient map. Since $\mathbb{Z}_2 \cong \{I, -I\}$ is a (0-dimensional) closed Lie subgroup of $\mathrm{SU}(2)$, the quotient map π is a surjective submersion, and hence by Lemma 15.5, Φ is smooth, hence is a Lie group isomorphism.

Now, by Corollary 15.13, π is actually a local diffeomorphism. In particular, this means that $\dim(\mathrm{Ad}(\mathrm{SU}(2))) = \dim(\mathrm{SU}(2)/\mathbb{Z}_2) = \dim(\mathrm{SU}(2)) = 3$. Thus $\mathrm{Ad}(\mathrm{SU}(2))$ is an embedded Lie subgroup of $\mathrm{SO}(\mathfrak{su}(2))$ of dimension 3, which is also the dimension of $\mathrm{SO}(\mathfrak{su}(2)) \cong \mathrm{SO}(3)$. It follows that $\mathrm{Ad}(\mathrm{SU}(2))$ is actually an open subgroup of $\mathrm{SO}(\mathfrak{su}(2))$, and since the latter is connected, we finally conclude by Lemma 11.11 that $\mathrm{Ad}(\mathrm{SU}(2)) = \mathrm{SO}(\mathfrak{su}(2))$. Summarizing:

PROPOSITION 15.18. *The Adjoint homomorphism* $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{SO}(\mathfrak{su}(2)) \cong \mathrm{SO}(3)$ *is a surjective submersion with kernel isomorphic to* $\mathbb{Z}_2$*, and descends to a Lie group homomorphism*

$$\mathrm{SU}(2)/\mathbb{Z}_2 \cong \mathrm{SO}(3).$$

In particular, the Lie algebras $\mathfrak{su}(2)$ *and* $\mathfrak{so}(3) = \mathfrak{o}(3)$ *are isomorphic.*

REMARK 15.19. It is worth noting that the surjectivity of $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{SO}(\mathfrak{su}(2))$ can be shown directly, by explicitly constructing, for any rotation Q on $\mathbb{R}^3$, a special unitary U which implements Q by conjugation on $\mathfrak{su}(2) \cong \mathbb{R}^3$. The key is to realize Q as a counterclockwise rotation of some angle θ in some plane (perpendicular to a given vector in $\mathbb{R}^3$), and then build U with eigenvalues $e^{\pm i\theta/2}$ with eigenvectors determined by the plane of rotation. This is slightly painful and left to the interested reader.

Note that $\mathfrak{o}(3)$ consists of antisymmetric 3×3 real matrices, which is spanned by the three matrices $\widehat{E}_{jk} = E_{jk} - E_{kj}$ for $(j,k) \in \{(1,2),(1,3),(2,3)\}$, where E_{jk} is the 3×3 matrix unit with a 1 in the jk position and 0s elsewhere. One can readily compute that

$$[\widehat{E}_{12}, \widehat{E}_{23}] = \widehat{E}_{13}, \quad [\widehat{E}_{12}, \widehat{E}_{13}] = -\widehat{E}_{23}, \quad [\widehat{E}_{13}, \widehat{E}_{23}] = -\widehat{E}_{12}.$$

Comparing this to (15.2), we see that the map $\mathfrak{su}(2) \to \mathfrak{o}(3)$ defined by

$$X \mapsto 2\widehat{E}_{12}, \quad Y \mapsto 2\widehat{E}_{23}, \quad Z \mapsto 2\widehat{E}_{13}$$

defines a Lie algebra isomorphism. So the fact that these two Lie algebras are isomorphic is not so surprising (there are not many 3-dimensional Lie algebras). That the isomorphism is implemented by such a discrete quotient map is illustrative of the topic of the next section (covering maps), which quantifies precisely the degree to which the Lie correspondence fails when the groups need not be simply connected.

3. Universal Covering Groups

As we noted in Corollary 15.13, if G is a Lie group and H is a discrete normal subgroup, then G and G/H have the same Lie algebra. It is then natural to ask whether this is the only way that the Lie algebra can fail to be a complete isomorphism invariant. The answer is no. But if we restrict our attention to connected Lie groups, this is in the right direction for a complete answer.

Let us say two Lie groups are *Lie algebra equivalent* if their Lie algebras are isomorphic; this is clearly an equivalence relation on Lie groups. If H is a discrete normal subgroup of G, then G and G/H are Lie algebra equivalent. But there are other ways this can happen in the same fashion. For example, suppose H_1 and H_2 are two discrete normal subgroups of G, neither contained in the other. Then G/H_1 and G/H_2 are both Lie algebra equivalent to G, and thus they are Lie algebra equivalent. It will turn out that *this* is generic: if two connected Lie groups are Lie algebra equivalent, then they are both isomorphic to quotients of some Lie group by discrete normal subgroups. At the moment, it is not even clear that this is an equivalence relation: it is easily seen to be reflexive and symmetric, but transitivity is unclear. This leads us to a discussion of *smooth covering spaces*.

DEFINITION 15.20. *Let E and M be connected smooth manifolds. A continuous map $\pi\colon E \to M$ is called a* **smooth covering map** *if it is smooth, surjective, and if every point in M has a neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of connected components, each of which is mapped onto U diffeomorphically by π. (We say that U is smoothly evenly covered by π.) We call E a* **smooth covering space** *for M.*

EXAMPLE 15.21. Let $\varepsilon\colon \mathbb{R} \to \mathbb{S}^1 \subset \mathbb{C}$ be the usual Lie group homomorphism $\varepsilon(x) = e^{2\pi i x}$. It is surjective and smooth, and for any point $e^{2\pi i\theta}$ with $\theta \in [0, 2\pi)$, we can take for example the open neighborhood $U_\theta = \{e^{2\pi i\phi}\colon |\phi - \theta| < \frac{\pi}{2}\}$. Then

$$\varepsilon^{-1}(U_\theta) = \bigsqcup_{n\in\mathbb{Z}} (\theta + n - \tfrac{1}{4}, \theta + n + \tfrac{1}{4}).$$

Each interval in the preimage is mapped onto U_θ smoothly and bijectively, and the inverse is a branch of the complex logarithm which is smooth. Thus ε is a smooth covering map, and $\mathbb{R}$ is a smooth covering space for $\mathbb{S}^1$.

Notice that the smooth covering map ε of Example 15.21 is actually a Lie group homomorphism $\mathbb{R} \to \mathbb{S}^1$, whose kernel is the discrete subgroup $\mathbb{Z}$ of $\mathbb{R}$. All such Lie group homomorphisms are smooth covering maps.

PROPOSITION 15.22. *Let G be a connected Lie group, and let H be a discrete normal subgroup of G. Then the quotient map $\pi\colon G \to G/H$ is a smooth covering map.*

PROOF. The quotient map π is a surjective submersion, so we need only show that it evenly covers the quotient. Let U_0 be a connected open neighborhood of $e \in G$ that contains no point of $H \setminus \{e\}$ (which is possible since H is discrete). Let $U \subseteq U_0$ be a connected neighborhood with the property that for any $u, v \in U$, $v^{-1}u \in U_0$ (the existence of which was proved in a homework problem).

Now, let $g \in G$, and consider the open neighborhood gU of g. Let $h \neq h'$ be two elements of H. Then $gUh \cap gUh' = \varnothing$, for otherwise there would exists $u, v \in U$ with $guh = gvh'$, meaning $uh = vh'$, and so $v^{-1}u = h'h^{-1} \in H$. But we also have $v^{-1}u \in U_0$ by definition of U, and since $U \cap H = \{e\}$, it follows that $h = h'$, contradicting the choice of $h \neq h'$.

Since π is an open map, $V = \pi(gU)$ is an open neighborhood of $[g]$. The preimage $\pi^{-1}(V)$ is equal to $gU \cdot H$, which we just proved is the disjoint union of the open connected sets $\{gUh\colon h \in H\}$. For any two points $u \neq v \in U$ and given $h \in H$, guh and gvh are in distinct cosets (this follows from an argument similar to the one above); thus $\pi\colon gUh \to V$ is a smooth bijection, and since π is a local diffeomorphism, it is a diffeomorphism. $\square$

Here are a few important properties of smooth covering maps.

PROPOSITION 15.23. *Let E and M be smooth manifolds, and let $\pi\colon E \to M$ be a smooth covering map.*

(1) *π is a local diffeomorphism.*
(2) *π is a smooth submersion.*
(3) *π is an open map.*
(4) *If π is injective, then it is a diffeomorphism.*

The proof of this proposition is left as a(n easy) homework exercise.

The theory of covering spaces is one of the earliest topics in algebraic topology. If M is a sufficiently nice topological space (precisely: connected, locally path connected, and semilocally simply connected) and if Γ is *any* subgroup of the homotopy group $\pi_1(M)$, then there is a covering space E_Γ for M with $\pi_1(E_\Gamma) \cong \Gamma$. We will only need this fact for the trivial subgroup, in which case the covering space is unique and natural.

THEOREM 15.24. *Let M be a connected smooth manifold. There exists a simply connected smooth manifold $\widetilde{M}$, called the* **universal covering manifold of** *M, and a smooth covering map $\pi\colon \widetilde{M} \to M$. It is, moreover, unique in the following strong sense: if $\pi'\colon \widetilde{M}' \to M$ is a smooth covering map and $\widetilde{M}'$ is simply connected, then there is a unique diffeomorphism $\Phi\colon \widetilde{M} \to \widetilde{M}'$ such that $\pi' \circ \Phi = \pi$.*

We will not prove Theorem 15.24. Let us just briefly outline how it works. Most of the work is at the topological level: for any connected, locally path connected, semilocally simply connected topological space B, there is a universal covering space E in the sense that E is simply connected and there is a continuous map $\pi\colon E \to B$ which is a topological covering map: π is surjective and each point in B has a neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of connected components each of which is mapped *homeomorphically* onto U by π. The proof of the existence of (B, π) can be found, for example, in [**5**, §82]. Fixing a basepoint $b_0 \in B$, E is defined to be the set of (endpoints-fixed) homotopy equivalence classes of continuous curves $\alpha\colon [0,1] \to B$ with $\alpha(0) = b_0$. The covering map is defined by $\pi([\alpha]) = \alpha(1)$. The space is then topologized by defining an open neighborhood of $[\alpha]$ to be the collection of (equivalence classes of) concatenated paths $\alpha * \delta$ with $\delta(0) = \alpha(1)$. It can then be shown that E is simply connected (as it is cooked up to be) and π is a topological covering map.

Now, once we know $\pi\colon E \to M$ is a topological covering map, note that E is immediately seen to be locally Euclidean: if U is a chart in M then $\pi^{-1}(U)$ is a stack of homeomorphic copies of U. It is also not hard to see that E is Hausdorff and second countable, so E is a topological manifold. To define an atlas on it, for any $p \in E$, fix a chart in (U, φ) in M centered at $\pi(p)$ that is evenly covered by π, and let $\widetilde{U}$ be the unique component of $\pi^{-1}(U)$ that contains p. Define $\widetilde{\varphi}\colon \widetilde{U} \to \mathbb{R}^n$ be the lift $\widetilde{\varphi} = \varphi \circ \pi|_{\widetilde{U}}$. These charts are smoothly compatible since $\widetilde{\psi} \circ \widetilde{\varphi}^{-1} = \psi \circ \varphi^{-1}$. This gives E a smooth manifold structure, and it is clear (from the fact that local coordinates on E are given by local coordinates on M) that π is a smooth covering map. Thus, we take $(\widetilde{M}, \pi) = (E, \pi)$ with this smooth structure. We leave the natural uniqueness statement as an exercise for the reader.

EXAMPLE 15.25. Since $\mathbb{R}$ is simply connected and $\varepsilon\colon \mathbb{R} \to \mathbb{S}^1$ of Example 15.21 is a smooth covering map, it follows that the universal covering manifold of $\mathbb{S}^1$ is $\mathbb{R}$.

EXAMPLE 15.26. In Section 2, we found a Lie group isomorphism $\mathrm{SO}(3) \cong \mathrm{SU}(2)/\{I, -I\}$, a quotient of the connected Lie group $\mathrm{SU}(2)$. Hence, by Proposition 15.22, the quotient map $\pi\colon \mathrm{SU}(2) \to \mathrm{SU}(2)/\{I, -I\} \cong \mathrm{SO}(3)$ is a smooth covering map. Now, as a manifold, $\mathrm{SU}(2)$ is diffeomorphic to $\mathbb{S}^3$ (this is a homework exercise), and is hence simply connected. Thus, the manifold $\mathbb{S}^3 \cong \mathrm{SU}(2)$ is the universal covering space for $\mathrm{SO}(3)$.

It is relatively easy to construct new universal covering manifolds from old ones. For example:

LEMMA 15.27. *Let $E_1, \dots, E_k$ and $M_1, \dots, M_k$ be smooth manifolds, and suppose $\pi_j\colon E_j \to M_j$ are smooth covering maps. Then $\pi_1 \times \cdots \times \pi_k\colon E_1 \times \cdots \times E_k \to M_1 \times \cdots \times E_k$ is a smooth covering map. Hence, if $\widetilde{M}_j$ is the universal covering manifold of M_j, then $\widetilde{M}_1 \times \cdots \times \widetilde{M}_k$ is the universal covering manifold of $M_1 \times \cdots \times M_k$.*

The proof is immediate: it is quick to verify that $\pi_1 \times \cdots \times \pi_k$ is a smooth covering map, and since products of simply connected spaces are simply connected, the uniqueness part of Theorem 15.24 shows that the product of universal covers is the universal cover of the product.

One important property of covering manifolds is the existence of lifts of smooth maps (from simply connected manifolds).

PROPOSITION 15.28 (Lifting Criterion). *Let $\pi\colon E \to M$ be a smooth covering map between connected manifolds. Let N be a simply connected manifold, and let $F\colon N \to M$ be a smooth map. Fix $p_0 \in N$ and $q_0 \in \pi^{-1}(F(p))$. Then there exists a unique smooth map $\widetilde{F}\colon N \to E$ satisfying $\pi \circ \widetilde{F} = F$ (called a* lift *of F) such that $\widetilde{F}(p_0) = q_0$.*

$$\begin{tikzcd} & E \arrow[d, "\pi"] \\ N \arrow[r, "F"'] \arrow[ur, dashed, "\widetilde{F}"] & M \end{tikzcd}$$

Here is a brief outline of the proof. For $p \in N$, let α_p be any continuous path with $\alpha(0) = p_0$ and $\alpha(1) = p$. The curve $F \circ \alpha_p$ has a unique lift to a continuous path β_p in E (i.e. $\pi \circ \beta_p = F \circ \alpha_p$) satisfying $\beta_p(0) = q_0$; this is the path-lifting property of covering maps. Define $\widetilde{F}(p) = \beta_p(1)$. This is well-defined: since N is simply connected, if α_p and α'_p are two paths connecting p_0 to p, they are homotopic, and hence $F \circ \alpha_p$ and $F \circ \alpha'_p$ are homotopic. It follows from the Monodromy Theorem that the lifts β_p and β'_p are homotopic and $\beta_p(1) = \beta'_p(1)$. Now, to prove $\widetilde{F}$ is smooth near each point p, fix a chart V in M at $F(p)$; then V lifts to a chart $\widetilde{V}$ in E at $\widetilde{F}(p)$, and π is the identity map in the $\widetilde{V}, V$ local coordinates. Thus, for points in $\widetilde{V}$, F and $\widetilde{F}$ have the same local coordinate representation, and so since F is smooth, so is $\widetilde{F}$. Uniqueness follows from the uniqueness in the path lifting property.

Now, suppose G is a connected Lie group. Then, by Theorem 15.24, it possesses a (naturally unique) universal covering manifold $(\widetilde{G}, \pi)$ that is simply connected. In fact, $\widetilde{G}$ possesses a unique group structure with respect to which $\pi\colon \widetilde{G} \to G$ is a Lie group homomorphism.

THEOREM 15.29. *Let G be a connected Lie group. There exists a simply connected Lie group $\widetilde{G}$ (called the* **universal covering group of** *G), and a smooth covering map $\pi\colon \widetilde{G} \to G$, such that π*

is a Lie group homomorphism. It is, moreover, unique in the following strong sense: if $\pi'\colon \widetilde{G}' \to G$ is a smooth covering map and Lie group homomorphism from another simply connected Lie group $\widetilde{G}'$, then there exists a unique Lie group isomorphism $\Phi\colon \widetilde{G} \to \widetilde{G}'$ such that $\pi' \circ \Phi = \pi$.

PROOF. Let $(\widetilde{G}, \pi)$ be the universal covering manifold of G, cf. Theorem 15.24. Then by Lemma 15.27, $\pi \times \pi\colon \widetilde{G} \times \widetilde{G} \to G \times G$ is a smooth covering map. Let $\mathrm{m}\colon G \times G \to G$ denote the multiplication map as usual, and fix some element $\widetilde{e} \in \pi^{-1}(e)$. Since $\widetilde{G}$ is simply connected, so is $\widetilde{G} \times \widetilde{G}$, and so by Proposition 15.28, the smooth map $\mathrm{m} \circ \pi \times \pi\colon \widetilde{G} \times \widetilde{G} \to G$ has a unique smooth lift to a map $\widetilde{\mathrm{m}}\colon \widetilde{G} \times \widetilde{G} \to \widetilde{G}$ satisfying $\widetilde{\mathrm{m}}(\widetilde{e}, \widetilde{e}) = \widetilde{e}$. We define the product on $\widetilde{G}$ as $\widetilde{\mathrm{m}}$: $\widetilde{g}\widetilde{h} \equiv \widetilde{\mathrm{m}}(\widetilde{g}, \widetilde{h})$. We must show this makes $\widetilde{G}$ into a Lie group (with identity $\widetilde{e}$). Before we do so, note that the definition of $\widetilde{m}$ gives $\mathrm{m} \circ \pi \times \pi = \pi \circ \widetilde{\mathrm{m}}$, which says that

$$\pi(\widetilde{g}\widetilde{h}) = \pi(\widetilde{\mathrm{m}}(\widetilde{g}, \widetilde{h})) = \mathrm{m}(\pi(\widetilde{g}), \pi(\widetilde{h})) = \pi(\widetilde{g})\pi(\widetilde{h}).$$

This, together with the fact that $\pi(\widetilde{e}) = e$, shows that π is a homomorphism.

To see that $\widetilde{e}$ is the identity element, consider the function $F\colon \widetilde{G} \to \widetilde{G}$ given by $F(\widetilde{g}) = \widetilde{e}\widetilde{g}$. Since π is a homomorphism and $\pi(\widetilde{e}) = e$, $\pi \circ F(\widetilde{g}) = e\pi(\widetilde{g}) = \pi(\widetilde{g})$ – i.e. $\pi \circ F = \pi$, so F is a lift of $\pi\colon \widetilde{G} \to G$ satisfying $F(\widetilde{e}) = \widetilde{m}(\widetilde{e}, \widetilde{e}) = \widetilde{e}$. But $\mathrm{Id}\colon \widetilde{G} \to \widetilde{G}$ is another lift, sending $\widetilde{e}$ to itself, and since $\widetilde{G}$ is simply connected, the uniqueness part of Proposition 15.28 shows that $F = \widetilde{Id}$. That is: $\widetilde{e}\widetilde{g} = \widetilde{g}$ for all $\widetilde{g} \in \widetilde{G}$, showing that $\widetilde{e}$ is an identity element.

Next, we show every element in $\widetilde{G}$ has an inverse. Let $\mathrm{i}\colon G \to G$ denote the inversion map $\mathrm{i}(g) = g^{-1}$. The smooth map $\mathrm{i} \circ \pi\colon \widetilde{G} \to G$ has a unique smooth lift $\widetilde{\mathrm{i}}\colon \widetilde{G} \to \widetilde{G}$ with $\widetilde{\mathrm{i}}(\widetilde{e}) = \widetilde{e}$ by Proposition 15.28. Now define a function $H\colon \widetilde{G} \to \widetilde{G}$ by $H(\widetilde{g}) = \widetilde{g} \cdot \widetilde{\mathrm{i}}(\widetilde{g})$. Then, since π is a homomorphism, we have

$$\pi \circ H(\widetilde{g}) = \pi(\widetilde{g}) \cdot \pi \circ \widetilde{\mathrm{i}}(\widetilde{g}) = \pi(\widetilde{g}) \cdot \mathrm{i} \circ \pi(\widetilde{g}) = e.$$

So H is a lift of the constant function e taking value $\widetilde{e}$ at $\widetilde{e}$; so is the constant function $\widetilde{e}$, and so by uniqueness $H(\widetilde{g}) = \widetilde{e}$ for all $\widetilde{g} \in \widetilde{G}$. An analogous argument shows that $\widetilde{i}(\widetilde{g})$ is a left inverse for $\widetilde{g}$ as well.

Finally, we must show that the multiplication on $\widetilde{G}$ is associative. Define the two bracket maps $b_L, b_R\colon \widetilde{G} \times \widetilde{G} \times \widetilde{G} \to \widetilde{G}$ by

$$b_L(\widetilde{g}, \widetilde{h}, \widetilde{k}) = (\widetilde{g}\widetilde{h})\widetilde{k}, \quad b_R(\widetilde{g}, \widetilde{h}, \widetilde{k}) = \widetilde{g}(\widetilde{h}\widetilde{k}).$$

Applying π to both sides, and repeatedly using the homomorphism property, we have

$$\pi \circ b_L(\widetilde{g}, \widetilde{h}, \widetilde{k}) = \pi(\widetilde{g}\widetilde{h})\pi(\widetilde{k}) = \pi(\widetilde{g})\pi(\widetilde{h})\pi(\widetilde{k}) = \pi(\widetilde{g})\pi(\widetilde{h}\widetilde{k}) = \pi \circ b_R(\widetilde{g}, \widetilde{h}, \widetilde{k}).$$

Thus, b_L and b_R are both lifts of the same map $\widetilde{G} \times \widetilde{G} \times \widetilde{G} \to G$ given by $(\widetilde{g}, \widetilde{h}, \widetilde{k}) \mapsto \pi(\widetilde{g})\pi(\widetilde{h})\pi(\widetilde{k})$. Since both take the same value at $(\mathrm{e}, \mathrm{e}, \mathrm{e})$, they are equal, and so the product is associative.

Thus, $\widetilde{G}$ is a group under this product. As the product is also smooth, $\widetilde{G}$ is a Lie group. The covering map $\pi\colon \widetilde{G} \to G$ is a smooth homomorphism, hence it is a Lie group homomorphism. The uniqueness follows immediately from the uniqueness in Theorem 15.24. □

REMARK 15.30. Note: we were able to choose any element of the fibre over e as the identity of $\widetilde{G}$; this, of course, determined the choice of product. A different choice of identity element results in an isomorphic Lie group, as the uniqueness implies

EXAMPLE 15.31. As we saw in Example 15.26, the universal covering manifold of $\mathrm{SO}(3)$ is the Lie group $\mathrm{SU}(2)$, a Lie group, where the covering map is the adjoint homomorphism $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{SO}(\mathfrak{su}(2))$. Note that the fibre over I is $\{\pm I\}$, and so we could have taken $-I$ as the identity of the covering group. Note that the map $U \mapsto -U$ is actually a self-diffeomorphism of $\mathrm{SU}(2)$. It is not a group homomorphism per se, but it is if we define the product one of the copies of $\mathrm{SU}(2)$ as it would be in the above construction:

$$U \boxtimes V \equiv -UV.$$

The identity element for this product is $-I$. The inverse of any element is the usual matrix inverse, since $U \boxtimes U^{-1} = -UU^{-1} = -I$. It is easy to verify $\boxtimes$ is associative: $(U \boxtimes V) \boxtimes W = UVW = U \boxtimes (V \boxtimes W)$. Since it is clearly smooth, this gives a new Lie group structure to $\mathrm{SU}(2)$, but it is not new since the diffeomorphism $U \mapsto -U$ is a Lie group isomorphism.

We are now in a position to see that there are Lie groups that are *not* matrix groups: the universal covering group $\widetilde{\mathrm{SL}(2)}$ is not a matrix group (where $\mathrm{SL}(2)$ denotes $\mathrm{SL}(2,\mathbb{R})$ to be clear). To prove this, we need a few lemmas.

LEMMA 15.32. *The Lie group* $\mathrm{SL}(2,\mathbb{C})$ *is simply connected, but the Lie group* $\mathrm{SL}(2,\mathbb{R})$ *is not simply connected.*

PROOF. We utilize the polar decomposition: any matrix A can be written in the form $A = UP$ where U is unitary (orthogonal in the real case) and P is positive semidefinite: in fact $P = \sqrt{A^*A}$. If A is invertible then P is positive definite, and in this case $U = A(A^*A)^{-1/2}$ is uniquely determined. If $\det A = 1$, then $\det P \det U = 1$, and since $\det P > 0$ and $\det U \in \mathbb{S}^1$, it follows that $\det P = \det U = 1$. Thus, every $A \in \mathrm{SL}(2,\mathbb{C})$ has a unique decomposition of the form UP with $U \in \mathrm{SU}(2)$ and $P > 0$ with $\det P = 1$ (denoted $\mathrm{SL}_{>2}(2,\mathbb{C})$), and clearly every such pair of matrices produces an element of $\mathrm{SL}(2,\mathbb{C})$; so we have a bijection $\mathrm{SL}(2,\mathbb{C}) \leftrightarrow \mathrm{SU}(2) \times \mathrm{SL}_{>2}(2,\mathbb{C})$. Now, the space $\mathrm{SL}_{>0}(2,\mathbb{C})$ is contractible: any P has two positive eigenvalues with product 1, which means they can be written in the form $\lambda, 1/\lambda$ for some positive $\lambda \leq 1$. Utilizing the spectral decomposition, write any P uniquely in the form

$$P = V_P \begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix} V_P^{-1}$$

with $0 < \lambda \leq 1$ and $V_P \in \mathrm{U}(2)$. Define a homotopy $H\colon [0,1] \times P \to P$ by

$$H(t,P) = V_P \begin{bmatrix} (1-t)\lambda + t & 0 \\ 0 & \frac{1}{(1-t)\lambda+t} \end{bmatrix} V_P^{-1}.$$

Then $H(0,P) = P$, $H(1,P) = I$, H is clearly continuous, and $H(t,P) \in \mathrm{SL}_{>0}(2)$ for all $t \in [0,1]$ (because we declared the first eigenvalue to be < 1, $(1-t)\lambda + t > 0$ for all $t \in [0,1]$). Thus $\mathrm{SL}_{>0}(2,\mathbb{C})$ is contractible to the point I. Combining this with the polar decomposition bijection, this gives a contraction from $\mathrm{SL}(2,\mathbb{C})$ onto $\mathrm{SU}(2)$, which is simply connected. Thus, $\mathrm{SL}(2,\mathbb{C})$ is simply connected.

A completely analogous argument shows that $\mathrm{SL}(2,\mathbb{R})$ contracts onto $\mathrm{SO}(2,\mathbb{R}) = \mathbb{S}^1$, which shows that $\mathrm{SL}(2,\mathbb{R})$ is not simply connected. $\square$

Despite the fact that $\mathrm{SL}(2,\mathbb{R})$ is not simply connected, the nice way it sits inside $\mathrm{SL}(2,\mathbb{C})$ makes it eligible for the conclusion of the Lie correspondence.

LEMMA 15.33. *Let $n \in \mathbb{N}$, and let $\phi\colon \mathfrak{sl}(2,\mathbb{R}) \to \mathrm{gl}(n,\mathbb{C})$ be a Lie algebra homomorphism. Then there exists a Lie group homomorphism $\Phi\colon \mathrm{SL}(2,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{C})$ such that $\Phi(e^X) = e^{\phi(X)}$ for all $X \in \mathrm{sl}(2,\mathbb{R})$.*

PROOF. The Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ consists of all 2×2 real matrices with trace 0. Similarly, the Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ consists of all 2×2 complex matrices with trace 0; this decomposes as

$$\begin{aligned}\mathfrak{sl}(2,\mathbb{C}) &= \left\{ \begin{bmatrix} a+ia' & b+ib' \\ c+ic' & -a-ia' \end{bmatrix} : a,a',b,b',c,c' \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} + i \begin{bmatrix} a' & b' \\ c' & -a' \end{bmatrix} : a,a',b,b',c,c' \in \mathbb{R} \right\} \\ &= \mathfrak{sl}(2,\mathbb{R}) + i\mathfrak{sl}(2,\mathbb{R}).\end{aligned}$$

We can therefore extend ϕ to a linear map $\phi_{\mathbb{C}}\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{gl}(n,\mathbb{C})$ in the obvious way: $\phi_{\mathbb{C}}(X+iY) = \phi(X) + i\phi(Y)$ for any $X, Y \in \mathrm{sl}(2,\mathbb{R})$. This defines a linear map, and we also have

$$\begin{aligned}[\phi_{\mathbb{C}}(X+iY), \phi_{\mathbb{C}}(X'+iY')] &= [\phi(X)+i\phi(Y), \phi(X')+i\phi(Y')] \\ &= [\phi(X),\phi(X')] + i[\phi(X),\phi(Y')] + i[\phi(Y),\phi(X')] - [\phi(Y),\phi(Y')] \\ &= [X,X'] + i[X,Y'] + i[Y,X'] - [Y,Y'] = [X+iY, X'+iY']\end{aligned}$$

where the third equality is the fact that ϕ is a Lie algebra homomorphism. Thus $\phi_{\mathbb{C}}\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathfrak{gl}(n,\mathbb{C})$ is a Lie algebra homomorphism.

Since $\mathrm{SL}(2,\mathbb{C})$ is simply connected by Lemma 15.32, by Theorem 14.35, there is a unique Lie group homomorphism $\Phi_{\mathbb{C}}\colon \mathrm{SL}(2,\mathbb{C}) \to \mathrm{GL}(n,\mathbb{C})$ such that $\Phi_{\mathbb{C}}(e^Z) = e^{\phi_{\mathbb{C}}(Z)}$ for all $Z \in \mathfrak{sl}(2,\mathbb{C})$. Taking $\Phi = \Phi_{\mathbb{C}}|_{\mathrm{SL}(2,\mathbb{R})}$ provides the desired homomorphism. □

REMARK 15.34. One might think that a similar trick should work for $\mathrm{SO}(2)$, but one needs to be careful: the Lie algebra $\mathfrak{so}(2) = \mathfrak{o}(2)$ consists of 2×2 real antisymmetric matrices, which is 1-dimensional over $\mathbb{R}$. On the other hand, $\mathfrak{su}(2)$ is 3-dimensional, so there is no way we could decompose it as $\mathfrak{o}(2) + i\mathfrak{o}(2)$. In fact, the complexification of $\mathfrak{o}(2,\mathbb{R})$ is $\mathfrak{o}(2,\mathbb{C})$, and the Lie group $\mathrm{SO}(2,\mathbb{C})$ is not simply connected.

COROLLARY 15.35. *Let $G \subseteq \mathrm{GL}(n,\mathbb{C})$ be a connected matrix Lie group with Lie algebra $\mathfrak{g}$, and suppose $\Psi\colon G \to \mathrm{SL}(2,\mathbb{R})$ is a Lie group homomorphism whose induced Lie algebra homomorphism $\Psi_*\colon \mathfrak{g} \to \mathfrak{sl}(2,\mathbb{C})$ is a Lie algebra isomorphism. Then Ψ is a Lie group isomorphism.*

PROOF. By assumption the inverse $\phi = \Psi_*^{-1}\colon \mathfrak{sl}(2,\mathbb{R}) \to \mathfrak{g} \subseteq \mathfrak{gl}(n,\mathbb{C})$ is a Lie algebra isomorphism (ergo homomorphism). By Lemma 15.33, there is a Lie group homomorphism $\Phi\colon \mathrm{SL}(2,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{C})$ such that $\Phi(e^X) = e^{\phi(X)}$ for all $X \in \mathfrak{g}$; since $\phi(X) \in \mathfrak{g}$, it follows that Φ maps into G. Now, since ϕ and ψ are inverses of each other, and since G is connected, it follows from Lemma 14.36 that Φ and Ψ are inverses of each other. □

COROLLARY 15.36. *The universal covering group $\widetilde{\mathrm{SL}(2,\mathbb{R})}$ is not a matrix Lie group.*

PROOF. The covering map $\pi\colon \widetilde{\mathrm{SL}(2,\mathbb{R})}$ is a Lie group homomorphism by Theorem 15.29. It is also a local diffeomorphism by Proposition 15.23(1); hence, π_* is invertible, so it is a Lie algebra isomorphism. By Corollary 15.35, if $\widetilde{\mathrm{SL}(2,\mathbb{R})} \subseteq \mathrm{GL}(n,\mathbb{C})$ for any $n \in \mathbb{N}$, then π is a Lie group isomorphism; but this is impossible since $\widetilde{\mathrm{SL}(2,\mathbb{R})}$ is simply connected, while $\mathrm{SL}(2,\mathbb{R})$ is not simply connected (cf. Lemma 15.32). □

4. The Lie-Cartan Theorem

We are nearly ready to classify the set of all Lie groups sharing a common Lie algebra. The first question is: is this set non-empty for some given Lie algebra? The answer is yes, and this is the content of the present section.

To motivate the proof, recall Lemma 13.24, which asserts that if H is a Lie subgroup of G, then $\mathrm{Lie}(H)$ can be identified as

$$\mathrm{Lie}(H) = \{X \in \mathrm{Lie}(G) \colon \exp tX \in H \text{ for all } t \in \mathbb{R}\}.$$

We now turn this around. Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let $H \subseteq G$ be a subgroup that is not a priori known to be a Lie subgroup. The only candidate for its Lie algebra is the set $\{X \in \mathfrak{g} \colon \exp tX \in H \text{ for all } t \in \mathbb{R}\}$. This is no reason that this set should be a Lie subalgebra (or even a linear subspace) if H is a generic subgroup; it turns out that the set is *always* a Lie subalgebra (though we will not prove this). If the subgroup in question is also generated by the image of the exponential map on this Lie subalgebra, it is called *analytic*.

DEFINITION 15.37. *Let G be a Lie group with Lie algebra $\mathfrak{g}$. An* **analytic subgroup** *H of G is a subgroup with the property that*

$$\mathfrak{h} \equiv \{X \in \mathfrak{g} \colon \exp tX \in H \textit{ for all } t \in \mathbb{R}\}$$

is a Lie subalgebra of $\mathfrak{g}$, and such that $\exp \mathfrak{h}$ generates H: i.e.

$$H = \{\exp X_1 \exp X_2 \cdots \exp X_n \colon n \in \mathbb{N}, X_1, \ldots, X_n \in \mathfrak{h}\}.$$

Note that a analytic subgroup is a connected subset of G: for any element $h = \exp X_1 \cdots \exp X_n \in H$, the curve $h(t) = \exp tX_1 \cdots \exp tX_n$ is continuous and connects $h(0) = I$ to $h(1) = h$. We will see that analytic subgroups are indeed connected Lie subgroups. It is worth keeping in mind that they *need not be closed*. For example: let $G = \mathbb{T}^2$, the 2-torus. It's Lie algebra is $\mathbb{R}^2$ with trivial bracket. If $X = (x, y) \in \mathbb{R}^2$ is any element with $x \neq 0$ and $y/x \in \mathbb{R} \setminus \mathbb{Q}$, then $\mathrm{span}_{\mathbb{R}}\{X\}$ is a Lie subalgebra which is the Lie algebra of the dense subgroup $\{(e^{itx}, e^{ity}) \colon t \in \mathbb{R}\} \subset H$.

Thus, we will need to give H a topology that may not be the same as the topology of G to make it into a Lie group. As a first step, we will need the following "rational approximation" lemma, which will be used to show that there is a second countable topology on H.

LEMMA 15.38. *Let G be a Lie group with Lie algebra $\mathfrak{g}$, let $\mathfrak{h} \subseteq \mathfrak{g}$ be a Lie subalgebra, and let H be the subgroup generated by $\exp \mathfrak{h}$. Fix a linear basis of $\mathfrak{h}$, and call an element $X \in \mathfrak{h}$* **rational** *if its coefficients in the given basis are rational. Then for any neighborhood U_0 of 0 in $\mathfrak{g}$, and any $h \in H$, there is a finite collection $R_1, \ldots, R_n$ of rational elements in $\mathfrak{h}$ and an element $X \in U_0 \cap \mathfrak{h}$, such that*

$$h = \exp R_1 \exp R_2 \cdots \exp R_n \exp X.$$

PROOF. We use the Baker-Campbell-Hausdorff-Poincaré formula repeatedly here. Let U be the neighborhood of 0 in $\mathfrak{g}$ guaranteed by Theorem 14.21, so that, for $X, Y \in U$,

$$\exp X \exp Y = \exp C(X, Y) \tag{15.3}$$

where $C(X, Y)$ is defined in (14.7). Note from that equation that $C(\cdot, \cdot)$ is a continuous function. Let $U_1 \subseteq U$ be a neighborhood of 0 small enough that $C(X, Y) \in U$ whenever $X, Y \in U_1$. Wlog, we will shrink the given neighborhood U_0 to $U_0 \cap U_1$ (for if we can find $X \in U_0 \cap U_1 \cap \mathfrak{h}$, it is in $U_0 \cap \mathfrak{h}$ as desired).

Now, for any $X \in \mathfrak{h}$, there is some k so that $\frac{1}{k}X \in U_0$; thus $\exp X = (\exp \frac{1}{k}X)^k$ is a product of exponentials of elements of $U_0 \cap \mathfrak{h}$. Since any element of H is a product $\exp X_1 \cdots \exp X_n$

for some $X_1, \dots, X_n \in \mathfrak{h}$, it follows that we can write every element $h \in H$ in the form $h = \exp X_1 \cdots \exp X_m$ for some $m \in \mathbb{N}$ and $X_j \in U_0 \cap \mathfrak{h}$ for each $j \in \{1, \dots, m\}$. Our goal is to show that we can replace $X_1, \dots, X_m$ with rational elements in this decomposition, tacking on at most one more $\exp X$ for some $X \in U_0 \cap \mathfrak{h}$. We proceed by induction on m. For the base case, if $m = 0$ then $h = e = \exp 0 = \exp 0 \exp 0$, satisfying the conditions.

Proceeding by induction, suppose we know that for any element $h \in H$ which can be written in the form $\exp X_1 \cdots \exp X_m$ with $X_1, \dots, X_m \in U_0 \cap \mathfrak{h}$, there are rational elements $R_1, \dots, R_m \in \mathfrak{h}$ such that $h = \exp R_1 \cdots \exp R_m \exp X$ for some $X \in U_0 \cap \mathfrak{h}$. Consider, then, some element $h = \exp X_1 \cdots \exp X_m \exp X_{m+1}$ with $X_1, \dots, X_{m+1} \in U_0 \cap \mathfrak{h}$. By the inductive hypothesis, we can find rational elements $R_1, \dots, R_m$ and some $X \in U_0 \cap \mathfrak{h}$ so that

$$\begin{aligned} h = \exp X_1 \cdots \exp X_m \exp X_{m+1} &= \exp R_1 \cdots \exp R_m \exp X \exp X_{m+1} \\ &= \exp R_1 \cdots \exp R_m \exp C(X, X_{m+1}) \end{aligned}$$

by the Baker-Campbell-Hausdorff-Poincaré formula (14.7). Since $\mathfrak{h}$ is a Lie subalgebra of $\mathfrak{g}$ and $X, X_{m+1} \in \mathfrak{h}$, $C(X, X_{m+1}) \in \mathfrak{h}$, and by the assumption $U_0 \subseteq U_1$, $C(X, X_{m+1}) \in U$. This is not quite sufficient, since we would like to show it is in the smaller neighborhood U_0. To fix this, we introduce a new rational element $R_{m+1} \in \mathfrak{h}$ which is in U and is very close to $C(X, X_{m+1})$; since rational elements are dense in $\mathfrak{h}$, such elements clearly exist. In particular, Then we may apply the Baker-Campbell-Hausdorff-Poincaré formula once more, and we have

$$\begin{aligned} h &= \exp R_1 \cdots \exp R_m \exp R_{m+1} \exp(-R_{m+1}) \exp C(X, X_{m+1}) \\ &= \exp R_1 \cdots \exp R_{m+1} \exp C(-R_{m+1}, C(X, X_{m+1})). \end{aligned}$$

Since R_{m+1} and $C(X, X_{m+1})$ are in the Lie subalgebra $\mathfrak{h}$, so is $C(-R_{m+1}, C(X, X_{m+1}))$. Since C is continuous, and $C(-Z, Z) = 0$, we can choose R_{m+1} close enough to $C(X, X_{m+1})$ that $C(-R_{m+1}, C(X, X_{m+1})) \in U_0$. This concludes the induction, and thus the proof of the lemma. □

We can now prove that analytic subgroups are, in fact, Lie groups.

THEOREM 15.39. *Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let $H \subseteq G$ be an analytic subgroup. Let $\mathfrak{h} = \{X \in \mathfrak{g} \colon \exp tX \in H \text{ for all } t \in \mathbb{R}\}$ (which is therefore assumed to be a Lie subalgebra of $\mathfrak{g}$). There exists a unique smooth structure on H that makes it into a Lie subgroup of G, such that the inclusion map $H \hookrightarrow G$ is a smooth immersion. The Lie algebra of H is $\mathfrak{h}$.*

PROOF. Fix a neighborhood U of 0 in $\mathfrak{g}$ as in the Baker-Campbell-Hausdorff-Poincaré formula 14.7, and also small enough that exponential map is a diffeomorphism from U onto $\exp(U) \subseteq G$. For simplicity, we also assume U is a ball centered at 0. We define a topology on H by using as a base the collection

$$\mathcal{B} = \{h \cdot \exp(V \cap \mathfrak{h}) \colon h \in H, V \text{ an open nbhd of } 0 \text{ s.t. } \overline{V} \subset U\}.$$

Any set in $\mathcal{B}$ is contained in H, since $\exp \mathfrak{h} \subseteq H$. Open sets in H are those sets $W \subseteq H$ such that any point $w \in W$ is contained in some element $B \in \mathcal{B}$ with $B \subseteq W$. This means that two points $h_1, h_2 \in H$ are "close" if $h_2 = h_1 \exp X$ for some sufficiently small $X \in \mathfrak{h}$.

- *H is Hausdorff*: if $h_1 \neq h_2$ in H, since $h_1, h_2 \in G$ which is Hausdorff, we may find a subneighborhood V of U so that $h_1 \cdot \exp(V) \cap h_2 \cdot \exp(V) = \varnothing$. Then the two elements $h_1 \cdot \exp(V \cap \mathfrak{h}), h_2 \cdot \exp(V \cap \mathfrak{h}) \in \mathcal{B}$ separate h_1 and h_2, so H is Hausdorff.

- *H is second countable*: let $\mathcal{B}_0 \subset \mathcal{B}$ be the countable subset of all base elements of the form $\exp R_1 \cdots \exp R_m \cdot \exp(V_0 \cap \mathfrak{h})$ for some $m \in \mathbb{N}$, some rational elements $R_1, \ldots, R_m \in \mathfrak{h}$, and some open ball $V_0 \subseteq U$ centered at 0 with rational radius. Given any base element $B = h \cdot \exp(V \cap \mathfrak{h}) \in \mathcal{B}$, by Lemma 15.38 for any neighborhood U_0 of 0 in $\mathfrak{g}$ we can express $h = \exp R_1 \cdots \exp R_m \exp X$ for $R_1, \ldots, R_m$ rational and $X \in U_0 \cap \mathfrak{h}$. Taking $U_0 \subseteq U$, then for any $Y \in V \cap \mathfrak{h} \subset U_0 \cap \mathfrak{h}$, by the Baker-Campbell-Hausdorff-Poincaré formula we have $\exp X \exp Y = \exp C(X, Y)$ where $C(X, Y) \in U \cap \mathfrak{h}$. From (14.7), we have $C(0, Y) = Y$, and by continuity and compactness of $\overline{V}$, for any open ball V_0 containing $\overline{V}$, there is a neighborhood U_0 so that $C(X, Y) \in V_0$ for all $X \in U_0$ and $Y \in V$. It follows that B is contained in some element of $\mathcal{B}_0$; so $\mathcal{B}_0$ is a countable base for the topology of H.

Now, the topology is evidently locally Euclidean: the map $X \mapsto h \cdot \exp X$ is a homeomorphism from $U \cap \mathfrak{h}$ onto its image in H, and the former is an open subset of the vector space $\mathfrak{h}$. We also use this map as our local coordinate chart near h. To prove this forms an atlas, we must show the transition maps are smooth. Any element of an overlap of two such charts can therefore be written in either form $h_1 \exp X = h_2 \exp Y$ for some $X, Y \in U \cap \mathfrak{h}$, and so $h_2^{-1} h_1 = \exp Y \exp(-X)$, which is small since X, Y are small. Shrinking the neighborhood U to some U', we may arrange for $\exp Y \exp(-X) \in \exp(U)$ whenever $X, Y \in U'$, and so we will always have $h_2^{-1} h_1 \in \exp(U)$. This means that $\exp Y = h_2^{-1} h_1 \exp X$. Repeating the process above, we can arrange that this product is always in $\exp(U')$, and since $\exp$ is a diffeomorphism on U, it follows that the map $X \mapsto Y$ is smooth. This defines our smooth atlas, making H into a smooth manifold. In the chosen local coordinates, the inclusion map is simply $X \mapsto h \exp X$ on a neighborhood of h, and this is a smooth immersion as claimed. It now follows that H is a Lie group, since the group operations are the restrictions of those on G. Its Lie algebra is the tangent space at the identity, where the chart used was $X \mapsto \exp X$ on $U \cap \mathfrak{h}$; hence, the tangent space is $\mathfrak{h}$, as claimed. The uniqueness is left as an exercise. □

Our main application of analytic subgroups is the following theorem, often called Lie's Third Theorem.

THEOREM 15.40. *Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let $\mathfrak{h} \subseteq \mathfrak{g}$ be a Lie subalgebra. There exists a unique connected Lie subgroup H of G with Lie algebra $\mathfrak{h}$.*

To see how the proof goes, let us first address uniqueness. If H is a connected Lie subgroup with Lie algebra $\mathfrak{h}$, then $\exp X \in H$ for any $X \in \mathfrak{h}$. Since H is a subgroup, it follows that H contains the subgroup generated by $\exp \mathfrak{h}$. In fact, we will see that H *is* this subgroup:

$$H = \{\exp(X_1) \cdots \exp(X_n) \colon n \in \mathbb{N}, X_1, \ldots, X_n \in \mathfrak{h}\}. \tag{15.4}$$

If H' is any other Lie subgroup with Lie algebra $\mathfrak{h}$, it must therefore contain H, and since both have the same Lie algebra, they are the same dimension, so H is open in H'. But if H and H' are both connected, it follows that $H = H'$.

Thus, we must show that the subgroup of (15.4) is, in fact, a Lie group. It suffices to show it is an analytic subgroup: we must show that

$$\{X \in \mathfrak{g} \colon \exp tX \in H \text{ for all } t \in \mathbb{R}\} = \mathfrak{h}. \tag{15.5}$$

The reverse containment $\supseteq$ is clear from the definition: if $X \in \mathfrak{h}$, so is tX for any $t \in \mathbb{R}$, and hence $\exp tX \in H = \langle \exp \mathfrak{h} \rangle$. It is the forward containment that concerns us now: we must see that if $\exp tX \in H$ for all $t \in \mathbb{R}$, then in fact $X \in \mathfrak{h}$. To begin, decompose $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{b}$ for

some complementary subspace $\mathfrak{b}$ (that is not necessarily a Lie subalgebra). By Lemma 14.7, there are neighborhoods U, V of 0 in $\mathfrak{h}$ and $\mathfrak{b}$ so that the map $\Phi\colon U \times V \to G$ given by $\Phi(X, Y) = \exp X \exp Y$ is a diffeomorphism onto its image. Now, by definition of H, $\exp X \in H$; this does not preclude the possibility that $\exp Y \in H$, and this is entirely possible. As it happens, however, this only occurs for countably many $Y \in V$.

LEMMA 15.41. *Let U, V be neighborhoods as above, and let $E \subseteq V$ be the set of elements $Y \in \mathfrak{b}$ with $\exp Y \in H$:*

$$E = \{Y \in V\colon\ \exp Y \in H\}.$$

Then E is countable.

PROOF. Let C be the expression in the exponential on the right side of the Baker-Campbell-Hausdorff-Poincaré formual, cf. (15.3). As in the discussion following that equation, let $U_1 \subseteq U$ be a neighborhood of 0 in $\mathfrak{h}$ small enough that $C(X, Y) \in U$ for all $X, Y \in U_1$. Then for any sequence $R_1, \dots, R_n$ of rational elements of $\mathfrak{h}$, there is *at most one* $X \in U_1$ such that

$$\exp R_1 \exp R_2 \cdots \exp R_n \exp X \in \exp V.$$

Indeed: if there are $X_1, X_2 \in U_1$ and $Y_1, Y_2 \in V$ such that

$$\begin{aligned} \exp R_1 \exp R_2 \cdots \exp R_n \exp X_1 &= \exp Y_1, \\ \exp R_1 \exp R_2 \cdots \exp R_n \exp X_2 &= \exp Y_2 \end{aligned} \tag{15.6}$$

then $\exp Y_1 \exp(-X_1) = \exp Y_2 \exp(-X_2)$, and it follows that

$$\exp(-Y_1) = \exp(-X_1) \exp X_2 \exp(-Y_2) = \exp C(-X_1, X_2) \exp(-Y_2),$$

where $C(-X_1, X_2) \in U$ by assumption. But Φ is a diffeomorphism on $U \times V$, and is therefore injective; since $\exp(-Y_1) = \exp 0 \exp(-Y_1)$ with $(0, -Y_1) \in U \times V$, we must therefore have $C(-X_1, X_2) = 0$ and $-Y_2 = -Y_1$. In particular, $Y_1 = Y_2$, and so (15.6) implies that $\exp X_1 = \exp X_2$. Since $X_1, X_2 \in U$ where $\exp$ is a diffeomorphism, it follows that $X_1 = X_2$, as claimed.

Now, by Lemma 15.38, every element in H can be expressed as $\exp R_1 \cdots \exp R_n \exp X$ for some $n \in \mathbb{N}$, some rational elements $R_1, \dots, R_n$, and some $X \in U_1$. But there are only countably many elements of the form $\exp R_1 \cdots \exp R_n$ for some n, and as shown above, for each such expression, there is at most one $X \in U_1$ for which the product $\exp R_1 \cdots \exp R_n \exp X$ is in $\exp V$. Thus, the set of $Y \in V$ for which $\exp Y \in H$ is countable. $\square$

PROOF OF THEOREM 15.40. As noted above, it suffices to show that the subgroup $H = \langle \exp \mathfrak{h} \rangle$ is an analytic subgroup satisfying (15.5). I.e. let

$$\mathfrak{h}' \equiv \{X \in \mathfrak{g}\colon\ \exp tX \in H \text{ for all } t \in \mathbb{R}\}.$$

We will show that $\mathfrak{h}' = \mathfrak{h}$. By definition $\mathfrak{h} \subseteq \mathfrak{h}'$. Conversely, let $Z \in \mathfrak{h}'$; then $\exp tZ \in H$ for all $t \in \mathbb{R}$. Let $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{b}$ with neighborhoods U, V of 0 in $\mathfrak{h}$ and $\mathfrak{b}$ be as above, so that $\Phi(X, Y) = \exp X \exp Y$ is a diffeomorphism from $U \times V$ onto its image in G. For all sufficiently small $t \in \mathbb{R}$, $\exp tZ \in \Phi(U \times V)$; define $(X(t), Y(t)) = \Phi^{-1}(\exp tZ)$. So $X(t)$ is a smooth function from a small time interval into U, and $Y(t)$ is a smooth function from a small time interval into V.

Now, $\exp tZ = \exp X(t) \exp Y(t)$, and since $X(t) \in U \subseteq \mathfrak{h}$, $\exp X(t) \in H$ for all t. Thus $\exp Y(t) = \exp tZ \exp(-X(t)) \in H$ for all small $t \in \mathbb{R}$. It follows that $Y(t)$ is a constant function of t on its interval of definition. For, if not, then the continuous function $Y(\cdot)$ takes on uncountably many values; but then there would be uncountably many $Y(t) = Y \in V$ so that $\exp Y \in H$, contradicting Lemma 15.41. Thus $Y(t) = Y(0)$ for all small t, and since

$e = \exp(0Z) = \exp X(0) \exp Y(0)$, it follows (from the fact that Φ is injective) that $X(0) = Y(0) = 0$. Thus $Y(t) = 0$ for all small t, and so, for all small t, $\exp tZ = \exp X(t)$, which shows that $tZ = X(t)$ (again since Φ is injective). As $X(t) \in U \subseteq \mathfrak{h}$, it follows that $Z \in \mathfrak{h}$. We have thus shown that $\mathfrak{h}' \subseteq \mathfrak{h}$, concluding the proof. □

Finally, this brings is to the Lie-Cartan theorem, the main result of this section: every finite dimensional Lie algebra is the Lie algebra of some Lie group.

THEOREM 15.42 (Lie-Cartan). *Let $\mathfrak{g}$ be a finite dimensional Lie algebra. There is a unique simply connected Lie group G with* $\mathrm{Lie}(G) \cong \mathfrak{g}$.

PROOF. The vast majority of the proof relies on *Ado's Theorem*: if $\mathfrak{g}$ is any finite dimensional Lie algebra, there is an injective Lie algebra homomorphism $\phi\colon \mathfrak{g} \to \mathfrak{gl}(n, \mathbb{C})$ for some finite n (where n is not usually the dimension of $\mathfrak{g}$). This is a hard theorem that is completely algebraic and well beyond the scope of this course, so we take it on faith; the perturbed reader can then simply interpret the statement of this theorem to read "If $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{gl}(n, \mathbb{C})$ for some n, then there is a unique simply connected Lie group with Lie algebra isomorphic to $\mathfrak{g}$".

Thus, we may view $\mathfrak{g} \subseteq \mathfrak{gl}(n, \mathbb{C})$ as a Lie subalgebra. By Theorem 15.40, there is a connected Lie subgroup $H \subseteq \mathrm{GL}(n, \mathbb{C})$ with $\mathrm{Lie}(H) = \mathfrak{g}$. Now, define $G = \widetilde{H}$ to be the universal covering group of H. Then $\mathrm{Lie}(G) \cong \mathrm{Lie}(H) = \mathfrak{g}$, thus such a simply connected Lie group exists. If G' is another simply connected Lie group with $\mathrm{Lie}(G') \cong \mathfrak{g}$, then by the Lie correspondence $G' \cong G$, proving uniqueness. □

5. The Lie Correspondence Revisited

We can now completely answer the question: given a Lie algebra $\mathfrak{g}$, what connected Lie groups are there having $\mathfrak{g}$ as their Lie algebra? The Lie-Cartan theorem asserts that there is a unique simply connected G with Lie algebra $\mathfrak{g}$. By Corollary 15.13, if H is any discrete normal subgroup of G, then G/H also has Lie algebra $\mathfrak{g}$. In fact, this is an exhaustive list.

THEOREM 15.43. *Let $\mathfrak{g}$ be a finite dimensional Lie algebra. Let G denote the unique simply connected Lie group with* $\mathrm{Lie}(G) \cong \mathfrak{g}$*, cf. Theorem 15.42. The set of all connected Lie group whose Lie algebra is isomorphic to $\mathfrak{g}$ is the set $\{G/\Gamma \colon \Gamma$ is a discrete normal subgroup of $G\}$.*

PROOF. The preceding discussion verifies that all Lie groups G/Γ have Lie algebra isomorphic to $\mathfrak{g}$, so we need only concern ourselves with the converse. Let H be a connected Lie group with Lie algebra $\mathrm{Lie}(H) \cong \mathfrak{g}$. Then the universal covering group $\widetilde{H}$ also has Lie algebra $\cong \mathfrak{g}$. As both G and $\widetilde{H}$ are simply connected Lie groups with the same Lie algebra, by the Lie correspondence they are isomorphic; fix an isomorphism $\Phi\colon G \to \widetilde{H}$. Now, the quotient map $\pi\colon \widetilde{H} \to H$ is a Lie group homomorphism. Thus $\pi \circ \Phi\colon G \to H$ is a surjective Lie group homomorphism. Its kernel $\Gamma = \ker \pi \circ \Phi$ is a normal Lie subgroup of G, and by the first isomorphism theorem for groups, $\pi \circ \Phi$ descends to a group isomorphism $G/\Gamma \to H$; this map is smooth since its composition with π is smooth, cf. Lemma 15.5. To complete the proof, we must show that Γ is discrete.

Since Lie group homomorphisms are constant rank (cf. Theorem 11.9), and $\pi \circ \Phi$ is surjective, it follows from the Global Rank Theorem 9.10 that is is a submersion. By Corollary 9.25 (on level-sets of submersions), the level set $\Gamma = \ker(\pi \circ \Phi) = (\pi \circ \Phi)^{-1}(e)$ is an embedded submanifold of G with codimension equal to the dimension of H; but G and H have the same dimension (both having isomorphic Lie algebras), and thus Γ is a closed Lie subgroup of dimension 0. Thus, Γ is a discrete normal subgroup, concluding the proof. □

Whence, if G_1 and G_2 are any two connected Lie groups with the same Lie algebra, then both G_1 and G_2 are isomorphic to quotients of their common universal covering group $\widetilde{G}$ by discrete normal subgroups $\Gamma_1, \Gamma_2 \subset \widetilde{G}$: $G_1 \cong \widetilde{G}/\Gamma_1$, $G_2 \cong \widetilde{G}/\Gamma_2$. This completely classifies the equivalence relation "having isomorphic Lie algebras" in the connected setting.

Let us conclude by discussing the complete picture: what happens in the disconnected setting? We need one further group theory definition.

DEFINITION 15.44. *Let G and H be groups, and suppose there is a surjective group homomorphism $\Phi\colon G \to H$. Let $G_0 = \ker\Phi$. We say that G* **is an extension of** G_0 **by** H.

THEOREM 15.45. *Let $\mathfrak{g}$ be a finite dimensional Lie algebra. If G is any Lie group with $\mathrm{Lie}(G) \cong \mathfrak{g}$, then there is a connected Lie group G_0 with Lie algebra $\mathrm{Lie}(G_0) \cong \mathfrak{g}$ such that G is an extension of G_0 by a discrete Lie group.*

PROOF. This is left as a homework exercise. □

CHAPTER 16

Compact Lie Groups

1. Tori

The simplest example of a compact Lie group is $\mathbb{S}^1$. Since Cartesian products of compact spaces are compact, any torus $\mathbb{T}^k = (\mathbb{S}^1)^k$ is a compact Lie group. As we will see, tori play a key role in understanding all compact connected Lie groups. Of course, by "torus", we refer only to the Lie group isomorphism class.

DEFINITION 16.1. *A Lie group is called a* **torus** *group if it is isomorphic as a Lie group to* $\mathbb{T}^k$ *for some finite* k*. We usually use the letter* T *to refer to tori.*

EXAMPLE 16.2. Let T denote the set of diagonal matrices in $\mathrm{SU}(n)$; T is evidently a closed subgroup, and so is a Lie group. If $u_1, \dots, u_n$ are the diagonal entries of $U \in T$, then

$$I = UU^* = \begin{bmatrix} |u_1|^2 & 0 & \cdots & 0 \\ 0 & |u_2|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & |u_n|^2 \end{bmatrix}, \quad \text{and} \quad 1 = \det U = u_1 \cdots u_n.$$

Thus $u_1, \dots, u_n$ are complex numbers of unit length, and $u_n = (u_1 \cdots u_n)^{-1}$. In particular, if we define a map $\Phi\colon \mathbb{T}^{n-1} \to T$ by

$$\Phi(u_1, \dots, u_n) = \begin{bmatrix} u_1 & 0 & \cdots & 0 & 0 \\ 0 & u_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & u_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & (u_1 \cdots u_n)^{-1} \end{bmatrix}$$

then Φ is a bijection. It is also clearly smooth, and easily shown to be a group homomorphism. Thus Φ is a Lie group isomorphism, and T is a torus, isomorphic to $\mathbb{T}^{n-1}$.

Note that, in addition to being compact, tori are connected, and abelian. In fact, these properties characterize tori.

THEOREM 16.3. *Every compact, connected, abelian Lie group is a torus.*

To prove this theorem, we will need a few supporting lemmas. The first is that, in the connected abelian category, the exponential map is surjective.

LEMMA 16.4. *Let* G *be a connected abelian Lie group, with Lie algebra* $\mathfrak{g}$*. Consider* $\mathfrak{g}$ *is an (additive) Lie group. Then* $\exp\colon \mathfrak{g} \to G$ *is a surjective Lie group homomorphism.*

PROOF. Since G is abelian, $\mathfrak{g}$ has trivial bracket $[X, Y] = 0$ (cf. Proposition 12.48). Hence, by the Baker-Campbell-Hausdorff-Poincaré formula, there is a neighborhood of 0 in $\mathfrak{g}$ so that

$\exp X \exp Y = \exp(X+Y)$ for all $X, Y \in U$. In fact, this is true for all $X, Y \in \mathfrak{g}$: choose some integer k so that $\frac{1}{k}X, \frac{1}{k}Y \in U$. Then we have

$$\begin{aligned}\exp(X+Y) = \exp(\tfrac{1}{k}X + \tfrac{1}{k}Y)^k &= (\exp(\tfrac{1}{k}X)\exp(\tfrac{1}{k}Y))^k \\ &= (\exp(\tfrac{1}{k}X))^k(\exp(\tfrac{1}{k}Y))^k = \exp X \exp Y.\end{aligned}$$

This shows that $\exp$ is a group homomorphism $\mathfrak{g} \to G$; as it is smooth, it is a Lie group homomorphism. Finally, choose a connected neighborhood U_0 of 0 in $\mathfrak{g}$ such that $\exp\colon U_0 \to \exp(U_0)$ is a diffeomorphism. Since U_0 is connected, so is $\exp(U_0)$, and since G is connected, it follows from Proposition 11.12(c) that $\exp(U_0)$ generates G. Thus, any $g \in G$ can be expressed in the form

$$g = \exp X_1 \exp X_2 \cdots \exp X_k = \exp(X_1 + \cdots + X_k)$$

for some $X_1, \ldots, X_k \in U_0 \subseteq \mathfrak{g}$, which shows that $\exp$ is surjective. □

REMARK 16.5. We saw another example where the exponential map is surjective in Section 3: the exponential map of the Heisenberg group $\mathrm{H}(3, \mathbb{R})$ is actually a diffeomorphism. It is worth pointing out that these cases are very unusual: typically the exponential map is neither one-to-one or onto, even in the connected category. For example, consider the matrix

$$A = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

There is no matrix $X \in \mathfrak{sl}(2, \mathbb{C})$ with $\exp X = A$. Indeed: if X has distinct eigenvalues, then it is diagonalizable, in which case $\exp X$ is diagonalizable; while A is not diagonalizable. If, on the other hand, X has a repeated eigenvalue, this eigenvalue must be 0 since $X \in \mathfrak{sl}(2, \mathbb{C})$ so $\mathrm{Tr}(X) = 0$. Let v be an eigenvector with eigenvalue 0, so $Xv = 0$; then $(\exp X)v = (\exp 0)v = v$. Thus $\exp X$ has 1 as an eigenvalue, unlike the matrix A.

(There are 2×2 matrices Z with $\exp Z = A$, but none of them are in $\mathfrak{sl}(2, \mathbb{C})$; they are all upper triangular with diagonal entries both equal to $\frac{\pi}{2}i + 2n\pi i$ for some n, and upper-diagonal entry defined accordingly to give A as the exponential.)

Hence, by the first isomorphism theorem for groups, we can identify G with the quotient $\mathfrak{g}/\ker(\exp)$. Since $\exp$ is a Lie group homomorphism, and is surjective, it is a submersion, and its kernel is a Lie subgroup of codmension $\dim G = \dim \mathfrak{g}$, as in the proof of Theorem 15.43. Thus, the kernel is a discrete subgroup of $\mathfrak{g}$. Discrete additive subgroups of vector spaces are lattices.

LEMMA 16.6. *Let V be a finite dimensional inner product space, and let $\Gamma \subset V$ be a discrete additive subgroup. There are linearly independent vectors $v_1, \ldots, v_k \in V$ such that Γ is the integer linear span of $\{v_1, \ldots, v_k\}$:*

$$\Gamma = \{m_1 v_1 + \cdots m_k v_k \colon m_1, \ldots, m_k \in \mathbb{Z}\}.$$

PROOF. If $\Gamma = \{0\}$ there is nothing to prove, so we assume Γ is non-trivial. As Γ is discrete, we can find a non-zero element $\gamma_0 \in \Gamma$ with minimal length. (There will be more than one, since $-\gamma_0 \in \Gamma$ as well.) Let $P\colon V \to V$ be the orthogonal projection onto $\gamma_0^\perp$, and let $\Gamma' = P(\Gamma)$. Since P is linear, Γ' is an additive subgroup of the subspace $\gamma_0^\perp$. We now claim that Γ' is discrete.

Suppose, for a contradiction, that Γ' is not discrete. Thus, for every $\epsilon > 0$, we can find elements $\delta \neq \delta' \in \Gamma'$ with $\|\delta - \delta'\| < \epsilon$. Now, $\delta - \delta' \in \Gamma' = P(\Gamma)$, so there is some element $\gamma \in \Gamma$ with $P(\gamma) = \delta - \delta'$. Thus $\|P(\gamma)\| < \epsilon$, meaning that the distance from γ to $\mathrm{span}(\gamma_0)$ is $< \epsilon$. This distance is equal to $\|\gamma - \beta\|$ where $\beta = (I - P)\gamma$ is the orthogonal projection of γ onto the span of γ_0. The element β is on the line through γ_0, so it lies between $m\gamma_0$ and $(m+1)\gamma_0$ for some

$m \in \mathbb{Z}$. Thus there is some integer n (equal to either m or $m+1$) so that $\|n\gamma_0 - \beta\| \leq \frac{1}{2}\|\gamma_0\|$. But then we have

$$\|n\gamma_0 - \gamma\| \leq \|n\gamma_0 - \beta\| + \|\beta - \gamma\| < \frac{1}{2}\|\gamma_0\| + \epsilon.$$

Now, by taking $\epsilon < \frac{1}{2}\|\gamma_0\|$ at the beginning, we have found an element $\gamma \in \Gamma$ whose distance from some integer multiple $n\gamma_0$ is less than $\|\gamma_0\|$. That is: $\gamma - n\gamma_0 \in \Gamma$ with $\|\gamma - n\gamma_0\| < \|\gamma_0\|$. But γ_0 was chosen to have minimal nonzero length in Γ, and so it follows that $\gamma = n\gamma_0$. But then $\delta - \delta' = P(\gamma) = P(n\gamma_0) = 0$, contradicting the choice $\delta \neq \delta'$.

Thus, Γ' is indeed discrete. We can now setup an induction on the dimension of V, with base case $V = \{0\}$ where the statement is trivially true. Suppose we have verified the statement of the lemma for all vector spaces of dimension $< d$, and now let $\dim V = d$. Fix a discrete subgroup Γ of Γ; by the preceding paragraph, choosing a minimal length non-zero element $\gamma_0 \in \Gamma$ and letting P be the orthogonal projection from V onto $\gamma_0^\perp$, the subgroup $\Gamma' = P(\gamma)$ is a discrete subgroup of the $d-1$ dimensional vector space $\gamma_0^\perp$. By the inductive hypothesis, there are linearly independent vectors $u_1, \ldots, u_{k-1} \in \gamma_0^\perp$ such that Γ' is precisely the set of integer linear combinations of $u_1, \ldots, u_{k-1}$. Select any vectors $v_1, \ldots, v_{k-1} \in V$ with $P(v_j) = u_j$ for $1 \leq j \leq k-1$. Then $v_1, \ldots, v_{k-1}$ and γ_0 are linearly independent in V. If $\gamma \in \Gamma$, then $P(\gamma) = m_1u_1 + \cdots + m_{k-1}u_{k-1} = P(m_1v_1 + \cdots + m_{k-1}v_{k-1})$ for some $m_1, \ldots, m_{k-1} \in \mathbb{Z}$, and it follows that there is some $\sigma \in \operatorname{span}(\gamma_0)$ such that

$$\gamma = m_1v_1 + \cdots + m_kv_k + \sigma.$$

We must now prove that σ is an *integer* multiple of γ_0; if not, an argument entirely analogous to the one in the preceding paragraph would produce a non-zero element in Γ parallel to γ_0 with length $< \|\gamma_0\|$, contradicting the minimal-length choice of γ_0. This concludes the proof. □

PROOF OF THEOREM 16.3. Let T be a connected abelian Lie group, with Lie algebra $\mathfrak{t}$. By Lemma 16.4, the exponential map is a surjective Lie group homomorphism $\exp\colon \mathfrak{t} \to T$; since $\exp$ is a local diffeomorphism, $\exp$ is a submersion, and so its kernel is a discrete additive subgroup of $\mathfrak{t}$. By Lemma 16.6, it follows that $\Gamma = \ker(\exp)$ is a discrete lattice: there are linearly independent vectors $v_1, \ldots, v_k \in \mathfrak{t}$ such that Γ is the integer span of $\{v_1, \ldots, v_k\}$.

The first isomorphism theorem for groups shows that the quotient map $\pi\colon \mathfrak{t} \to \mathfrak{t}/\Gamma$ descents to a group isomorphism $T \to \mathfrak{t}/\Gamma$, and this map is surjective and smooth (since its composition with π is smooth). Hence, it is a Lie group isomorphism, and so $T \cong \mathfrak{t}/\Gamma$. By decomposing $\mathfrak{t} = \operatorname{span}\{v_1\} \oplus \cdots \oplus \operatorname{span}\{v_k\} \oplus \operatorname{span}\{v_1, \ldots, v_k\}^\perp$, we see that this quotient is isomorphic to $(\mathbb{S}^1)^k \times \mathbb{R}^{n-k}$, where $n = \dim(T)$. This space is only compact if $n = k$, in which case $T \cong \mathbb{T}^n$ as claimed. □

REMARK 16.7. The above proof in fact shows the nominally stronger result that the only connected abelian Lie groups are those of the form $\mathbb{T}^k \times \mathbb{R}^m$ for some $k, m \in \mathbb{N}$. For example, the Lie group $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is isomorphic as a Lie group to $\mathbb{S}^1 \times \mathbb{R}$, under the isomorphism $(e^{i\theta}, t) \mapsto e^{i\theta}e^t$. The *compact* abelian connected Lie groups are precisely those with no Euclidean space factor, $m = 0$.

Now, let us look briefly as Lie subgroups of tori. As we have noted in many examples, $\mathbb{R}$ is a Lie subgroup of $\mathbb{T}^2$, immersed (but not embedded) densely. In general it, is useful to know when a one-dimensional subgroup of a torus is dense. The following proposition gives an important group theoretic characterization.

PROPOSITION 16.8. *Let T be a torus group, and let $t \in T$. The subgroup $\langle t \rangle$ generated by t is dense in T if and only if the only Lie group homomorphism $\Phi\colon T \to \mathbb{S}^1$ with $\Phi(t) = 1$ is the constant homomorphism $\Phi \equiv 1$.*

PROOF. First, if there exists a nonconstant Lie group homomorphism with $\Phi(t) = 1$, then $\ker(\Phi)$ is a closed subgroup of T that contains t; it thus contains the group $\langle t \rangle$ generated by t. But since Φ is nonconstant, $\ker \Phi \neq T$, and hence $\langle t \rangle$ is contained in a closed strict subgroup of T, this is not dense in T.

Conversely, let S be the closure of the group $\langle t \rangle$. Since it is closed, it is an embedded Lie subgroup of T; as T is abelian, S is normal, and so T/S is a Lie group, and the quotient map $\pi\colon T \to T/S$ is a surjective Lie group homomorphism. By assumption, $S \neq T$, and so T/S is not the trivial group. It is compact and connected, since it is the image of the compact connected Lie group T under the continuous map π, and is it abelian since it is the image of the abelian group T under a homomorphism. Thus, by Theorem 16.3, T/S is a torus group. Fix an isomorphism $\Psi\colon T/S \to \mathbb{T}^m$; since T/S is nontrivial, $m > 0$. Let $\pi_1\colon \mathbb{T}^m \to \mathbb{S}^1$ be the projection onto the first factor. Set $\Phi = \pi_1 \circ \Psi \circ \pi$. This is a Lie group homomorphism which is non-trivial, and since $t \in S$, $\pi(t) = e \in T/S$ so $\Phi(t) = 1$. □

As a corollary, we can now give a characterization of when an element generates a dense subgroup that is more concrete, and related to the examples we know (involving irrational angles).

THEOREM 16.9. *Let $t = (e^{2\pi i\theta_1}, \dots, e^{2\pi i\theta_k})$ be an element of the torus $\mathbb{T}^k$. Then t generates a dense subgroup of $\mathbb{T}^k$ if and only if the numbers*

$$1, \theta_1, \dots, \theta_k$$

are linearly independent over the rational field $\mathbb{Q}$.

PROOF. In light of Proposition 16.8, we must show that $\{1, \theta_1, \dots, \theta_k\}$ is linearly independent over $\mathbb{Q}$ if and only if there is no nonconstant Lie group homomorphism $\Phi\colon \mathbb{T}^k \to \mathbb{S}^1$ with $(e^{2\pi i\theta_1}, \dots, e^{2\pi i\theta_k}) \in \ker\Phi$. We prove this in contrapositive form. First, assume that $\{1, \theta_1, \dots, \theta_k\}$ is linearly dependent over $\mathbb{Q}$: so there are rational numbers $r_0, r_1, \dots, r_k$ with $r_0 + r_1\theta_1 + \cdots + r_k\theta_k = 0$. Multiplying by a common denominator for $r_0, r_1, \dots, r_k$, we find an integer linear combination $m_0 + m_1\theta_1 + \cdots + m_k\theta_k = 0$ for $m_0, m_1, \dots, m_k \in \mathbb{Z}$ not all 0. Now, define $\Phi\colon \mathbb{T}^k \to \mathbb{S}^1$ by

$$\Phi(u_1, \dots, u_k) = u_1^{m_1} \cdots u_k^{m_k}. \tag{16.1}$$

Then Φ is a nonconstant Lie group homomorphism, and

$$\Phi(t) = (e^{2\pi i\theta_1})^{m_1} \cdots (e^{2\pi i\theta_k})^{m_k} = 2^{2\pi i(m_1\theta_1 + \cdots + m_k\theta_k)} = e^0 = 1.$$

Conversely, suppose such a nonconstant homomorphism exists. In fact, *every* Lie group homomorphism $\mathbb{T}^k \to \mathbb{S}^1$ is of the form (16.1) for some integeral $m_1, \dots, m_k$ (this is a homework exercise). Since Φ is assumed nonconstant, not all the m_j are 0. Hence

$$1 = \Phi(t) = \Phi(e^{2\pi i\theta_1}, \dots, e^{2\pi i\theta_k}) = e^{2\pi i(m_1\theta_1 + \cdots + m_k\theta_k)}.$$

It follows that $m\theta_1 + \cdots + m_k\theta_k \in \mathbb{Z}$, and hence there is a (non-trivial) integer linear combination of $1, \theta_1, \dots, \theta_k$ that equals 0. Hence, these numbers are linearly dependent over $\mathbb{Z}$, and hence over $\mathbb{Q}$. □

2. Maximal Tori

We will see that torus subgroups control the structure of all connected compact Lie groups. First we need a notion of the *maximal* ones. We first look at the Lie algebra level.

DEFINITION 16.10. *Let $\mathfrak{g}$ be a Lie algebra. A* **maximal abelian subalgebra** $\mathfrak{t} \subseteq \mathfrak{g}$ *is a subspace that is an abelian subalgebra ($[X, Y] = 0$ for all $X, Y \in \mathfrak{t}$) such that $\mathfrak{t}$ is not contained in any strictly large abelian subalgebra.*

EXAMPLE 16.11. In $\mathfrak{su}(n)$, consider the subspace $\mathfrak{t}$ of diagonal elements: diagonal matrices with imaginary entries, and having trace 0. All diagonal matrices commute, so this is an abelian subalgebra. Note that it contains all the elements $iE_{jj} - iE_{kk}$ for $1 \leq j < k \leq n$. Now, let $X \in \mathfrak{su}(n)$ be any element not in $\mathfrak{t}$. Then X is not diagonal, so there is some entry (j, k), with $j < k$, so that $X_{jk} \neq 0$. But we can compute that

$$[(iE_{jj} - iE_{kk})X]_{jk} = iX_{jk}, \quad [X(iE_{jj} - iE_{kk})]_{jk} = -iX_{jk}.$$

Since $X_{jk} \neq 0$, this shows that X does not commute with $iE_{jj} - iE_{kk} \in \mathfrak{t}$, and hence there is no abelian extension of $\mathfrak{t}$. Thus, $\mathfrak{t}$ is a maximal abelian subalgebra of $\mathfrak{su}(n)$.

Of course, the maximal abelian subalgebra $\mathfrak{t}$ in Example 16.11 is the Lie algebra of the torus subgroup of diagonal matrices in $\mathrm{SU}(n)$.

DEFINITION 16.12. *Let G be a Lie group. A* **maximal torus** *in G is a torus subgroup that is not contained in any strictly larger torus sugroup of G.*

EXAMPLE 16.13. Consider again the torus subgroup $T \subset \mathrm{SU}(n)$ of Example 16.2. It is, in fact, a maximal torus. Indeed, if S is another torus subgroup of $\mathrm{SU}(n)$ with $T \subseteq S$, then since torus groups are abelian, every $s \in S$ commutes with every $t \in T$. Thus, any pair (s, t) with $s \in S$ and $t \in T$ (which are both unitary matrices) are simultaneously unitarily diagonalizable.

Now, let $t \in T$ be an element with all distinct eigenvalues. Then the eigenspace for each is 1-dimensional (over $\mathbb{C}$); as t is unitary, any collection of eigenvectors is orthogonal. Hence, since t is already diagonal, and therefore the standard basis $\{e_1, \ldots, e_n\}$ is a diagonalizing basis, it follows that *any* diagonalizing basis of unit vectors is of the form $\{e^{i\theta_1}e_1, \ldots, e^{i\theta_n}e_n\}$. Hence, there is a basis of this form that also diagonalizes s. Thus, each $s \in S$ is diagonalized by a diagonal matrix, and it follows that s is diagonal. Therefore, $s \in T$, and so $S = T$. So T is a maximal torus.

EXAMPLE 16.14. If T is a maximal torus in K, then so is xTx^{-1} for any $x \in K$. Indeed, the conjugation map $y \mapsto xyx^{-1}$ is a group isomomrphism, so xTx^{-1} is a torus. Suppose it is contained in another torus S; then $T \subseteq x^{-1}Sx$. By the same reasoning, $x^{-1}Sx$ is a torus, so since T is maximal, $T = x^{-1}Sx$, and so $S = xTx^{-1}$, which proves that S is maximal.

We will soon see that the correspondence between Examples 16.11 and 16.13 works in general in a *compact* Lie group: maximal abelian subalgebras are precisely the Lie algebras of maximal tori. A first result en route to this result is the following.

PROPOSITION 16.15. *Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let $\mathfrak{h}$ be a maximal abelian subalgebera of $\mathfrak{g}$. Then the conneced Lie subgroup $H \subseteq G$ with $\mathrm{Lie}(H) = \mathfrak{h}$ is closed.*

PROOF. Since $\exp(\mathfrak{h})$ contains a connected neighborhood of the identity in H, and since H is connected, it follows that it generates $\mathfrak{h}$ (indeed, this is how we defined H in the proof of Theorem 15.40 that such H exists and is unique). Since $\mathfrak{h}$ is abelian, it follows that H is abelian. It then

follows easily that the closure $\overline{H}$ of H in G is also abelian. In addition, as the closure of a connected set, $\overline{H}$ is connected.

Let $\mathfrak{h}' = \mathrm{Lie}(\overline{H})$; then $\mathfrak{h} \subseteq \mathfrak{h}'$. Since $\overline{H}$ is abelian, $\mathfrak{h}'$ is also abelian (by Proposition 12.48), and by maximality of $\mathfrak{h}$, it therefore follows that $\mathfrak{h} = \mathfrak{h}'$. Thus, $\overline{H}$ is a connected Lie subalgebra of G with Lie algebra $\mathfrak{h}$, and hence, by uniqueness, it is equal to H. Hence, H is closed. □

THEOREM 16.16. *Let K be a compact Lie group with Lie algebra $\mathfrak{k}$. If $T \subseteq K$ is a maximal torus with Lie algebra $\mathfrak{t}$, then $\mathfrak{t}$ is a maximal abelian subalgebra of $\mathfrak{k}$. Conversely, if $\mathfrak{h}$ is any maximal abelian subalgebra, then the connected Lie subgroup H of K with $\mathrm{Lie}(H) = \mathfrak{h}$ is a maximal torus.*

PROOF. First, suppose T is a maximal torus, which is abelian. Thus, its Lie algebra $\mathfrak{t}$ is abelian. Let $\mathfrak{t}'$ be the maximal abelian subalgebra containing $\mathfrak{t}$ (i.e. the intersection of all subspaces of $\mathfrak{k}$ all of whose elements commute with $\mathfrak{t}$). Let T' be the connected Lie subalgebra of K with $\mathrm{Lie}(T') = \mathfrak{t}'$. Since T' is connected, it is generated by $\exp(\mathfrak{t}')$, and hence is abelian. By Proposition 16.15, T' is closed, and since K is compact, T' is therefore compact. Thus, T' is a compact, connected, abelian Lie group, and thus by Theorem 16.3 T' is a torus group. So $T \subseteq T'$ is contained in the torus group T', and since T is a maximal torus, it follows that $T = T'$. This shows that $\mathfrak{t} = \mathfrak{t}'$, so $\mathfrak{t}$ is a maximal abelian subalgebra.

Conversely, suppose $\mathfrak{h}$ is a maximal abelian subalgebra. By Proposition 16.15, the connectes subgroup $H \subseteq K$ with $\mathrm{Lie}(H) = \mathfrak{h}$ is a closed subgroup, hence compact. It is also abelian since $\mathfrak{h}$ is, and it is connected by definition, so again by Theorem 16.3, H is a torus subgroup. If S is a torus subgroup with $H \subseteq S$, then $\mathfrak{h} \subseteq \mathfrak{s}$ where $\mathfrak{s} = \mathrm{Lie}(S)$ is abelian; since $\mathfrak{h}$ is maximal abelian, it follows that $\mathfrak{h} = \mathfrak{s}$, and since S is connected, it then follows that $H = S$. Thus H is a maximal torus. □

Maximal tori will turn out to completely control the structure of compact connected Lie groups. The following theorem is the key.

THEOREM 16.17. *Let K be a connected, compact Lie group. Fix some maximal torus $T \subseteq K$. Then every element $y \in K$ can be written in the form $y = xtx^{-1}$ for some $x \in K$ and $t \in T$.*

EXAMPLE 16.18. Let $K = \mathrm{SU}(n)$, and let T be the maximal torus of diagonal elements, cf. Example 16.13. Theorem 16.17 in this setting is almost precisely the statement of the spectral theorem: for any $U \in \mathrm{SU}(n)$, we can find $V \in \mathrm{U}(n)$ such that $V^{-1}UV$ is diagonal. Since $\det(V^{-1}UV) = \det U = 1$, it follows that $V^{-1}UV$ is in the maximal torus T. Now, the conjugating matrix V is composed of eigenvectors of U, which can be independently scaled by any unit modulus constant without changing the fact that V is unitary; hence, we can rescale the first column so that $\det V = 1$, proving the theorem in this case.

Theorem 16.17 is often stated in the following equivalent form.

THEOREM 16.19 (Cartan's Torus Theorem). *Let K be a connected, compact Lie group.*

(1) *Any two maximal tori in K are conjugate: if S_1 and S_2 are maximal tori in K, there is an element $x \in K$ such that $S_2 = xS_1x^{-1}$.*
(2) *Every element of K is contained in some maximal torus.*

PROOF OF EQUIVALENCE TO THEOREM 16.17. • *Theorem 16.17 $\implies$ Theorem 16.19*
First, for any $y \in K$, there is some $x \in K$ with $y \in xTx^{-1}$; since xTx^{-1} is a maximal torus, this verifies (2). For (1), let S be any maximal torus, and fix some element $s \in S$

such that the subgroup $\langle s \rangle$ generated by s is dense in S. By Theorem 16.17, there is some $x \in K$ and $t \in T$ with $s = xtx^{-1}$, and so $t = x^{-1}sx$. Then for any $k \in \mathbb{Z}$ $t^k = x^{-1}s^kx$, which shows that $x^{-1}s^kx \in T$ so $s^k \in xTx^{-1}$ for all K. The torus xTx^{-1} is closed, and the set $\{s^k \colon k \in \mathbb{Z}\}$ is dense in S, so it follows thst $S \subseteq xTx^{-1}$. But since S is maximal, and xTx^{-1} is a torus, it follows that $S = xTx^{-1}$. Thus, S is conjugate to T. Since conjugacy is an equivalence relation, it follows that any two maximal tori are conjugate, verifying (1).

- *Theorem 16.19* $\implies$ *Theorem 16.17* Fix any maximal torus T. Given $y \in K$, by (2) there is some maximal torus S with $y \in S$. By (1), there is some $x \in K$ with $S = xTx^{-1}$. Thus $y \in xTx^{-1}$, as required.

□

Before proceeding to prove Theorem 16.17, let us note two important corollaries.

COROLLARY 16.20. *If K is a connected, compact Lie group with Lie algebra $\mathfrak{k}$, then the exponential map* $\exp \colon \mathfrak{k} \to K$ *is surjective.*

PROOF. Let $x \in K$, and choose a maximal torus T containing x. By Lemma 16.4, the restriction of $\exp$ to $\mathrm{Lie}(T)$ is surjective, and so $x \in \exp(\mathrm{Lie}(T)) \subseteq \exp(\mathrm{Lie}(K))$. □

COROLLARY 16.21. *If K is a connected, compact Lie group, and $z \in K$, then z is in the center of K if and only if z belongs to every maximal torus in K.*

(Note: the center of K, $Z(K)$, is the subgroup of K consisting of those elements that commute with everything in K.)

PROOF. If $z \in Z(K)$, let T be a maximal torus containing z. If T' is any other maximal torus, there is some $x \in K$ with $T' = xTx^{-1}$, which shows that $xzx^{-1} \in T'$. But z commutes with x, so $xzx^{-1} = z$, and so $z \in T'$. Thus z is in all maximal tori.

Conversely, suppose z belongs to every maximal torus. Then for any $x \in K$, there is some maximal torus $T \ni x$, and by assumption $z \in T$ as well. Since T is an abelian subgroup, it follows that x commutes with z. So $z \in Z(K)$. □

The proof of Theorem 16.17 is quite involved. We will need two main ingredients: the theory of *mapping degrees* of smooth maps between compact manifolds (returning to integration of differential forms), and the *Weyl group* of a compact Lie group. We begin with the latter.

3. The Weyl Group, and Ad-Invariant Inner Products

We now introduce an important discrete group that plays a central structural role in a compact connected Lie group.

DEFINITION 16.22. *Let K be a compact connected Lie group, and let T be a maximal torus in K. The* **normalizer** *of T, $N(T)$, is the group of elements $x \in K$ such that $xTx^{-1} = T$. Note that $T \subseteq N(T)$ is a subgroup, and in fact (essentially by definition) it is a normal subgroup. The quotient group*

$$W(T) \equiv N(T)/T$$

is the **Weyl group** *of T.*

EXAMPLE 16.23. Consider the maximal torus T of diagonal elements in $\mathrm{U}(n)$, (that this is a maximal torus is proved analogously as in Example 16.13). If $t_0 \in T$, and σ is a permutation matrix, then for any $t \in T$, $(\sigma t_0)t(\sigma t_0)^{-1} = \sigma t_0 t t_0^{-1}\sigma^{-1} = \sigma t \sigma^{-1}$ is diagonal: its diagonal entries are those of t permuted according to σ. Thus, letting $\Sigma_n \subset \mathrm{U}(n)$ denote the group of permutation matrices, we see that $\Sigma_n T \subseteq N(T)$. In fact, this is an equality. If $u \in \mathrm{U}(n)$ satisfies $utu^{-1} \in T$ for all $t \in T$, this must hold in particular for some t with all n diagonal entries distinct. Such a t is unitarily diagonalizable, with diagonalization unique up to the order of the eigenvalues and a unit length constant for each eigenvector. Hence, $utu^{-1} = t'$ where t' is a permutation σ of the diagonal entries of t, and this fixes u to be σ up to a scaling constant in $\mathbb{S}^1$ in each diagonal entry: i.e. $u \in \Sigma_n T$.

Thus, $N(T) = \Sigma_n T$, which means that $W(T) = N(T)/T \cong \Sigma_n$ is the finite group of permtuations.

We will prove (in two sections) that the Weyl group is always be a finite subgroup of an orthogonal group. To see why, we describe its actions on several spaces in the following. Note first that $N(T)$ is a closed subgroup of T (as can be quickly verified, using compactness of T). Thus, it is an embedded Lie subgroup. By definition, we have a left action of $N(T)$ on T by $x \cdot t = xtx^{-1}$. This action descends to an action of the quotient $W(T) = N(T)/T$ on T: if $x, x' \in N(T)$ are in the same coset $xT = x'T$, then for $t \in T$, there is some $t' \in T$ with $xt = x't'$, from which it follows that

$$x \cdot t = xtx^{-1} = x't'(x't't^{-1})^{-1} = x't'(x't^{-1}t')^{-1} = x't(x')^{-1} = x' \cdot t$$

where we have used the fact that T is abelian, so $t't^{-1} = t^{-1}t'$. Thus $W(T)$ acts on T. We will see later that this action is *effective*: if $w \cdot t = t$ for all $t \in T$, then w is the identity.

Now, if $x \in N(T)$, then by definition, the conjugation $C_x(t) = xtx^{-1}$ is in T for each $t \in T$; thus, C_x preserves the Lie group T. It follows that the derivative $\mathrm{Ad}(x)$ preserves the Lie algebra $\mathfrak{t}$ of T. This gives us an action of $N(T)$ on $\mathfrak{t}$: $x \cdot H \equiv \mathrm{Ad}(x)H$ for $H \in \mathfrak{t}$. This *also* descends to an action of W on $\mathfrak{t}$: this follows from differentiating $x(\exp sH)x^{-1} = y(\exp sH)y^{-1}$ at $s = 0$. The action of $W(T)$ on $\mathfrak{t}$ is also effective: if $w \cdot H = H$ for all $H \in \mathfrak{t}$, it follows that $\mathrm{Ad}(x)H = H$ for all $H \in \mathfrak{t}$ for all $x \in w$, meaning $\mathrm{Ad}(x) = \mathrm{Id}_{\mathfrak{t}}$. Since C_x is a Lie group homomorphism of the connected Lie group T, it then follows that it too acts trivially.

We summarize this discussion as follows.

PROPOSITION 16.24. *Let K be a compact, connected Lie group, and let $T \subseteq K$ be a maximal torus with Lie algebra $\mathfrak{t}$. Let $W(T) = N(T)/T$ be the Weyl group of T in K. The map*

$$[\mathrm{Ad}]\colon W(T) \to \mathfrak{gl}(\mathfrak{t}), \quad [x] \mapsto \mathrm{Ad}(x)$$

is a well-defined, injective group homomorphism.

(The only part left unproved is the injectivity, which, as described above, follows from the effectiveness of the conjugation action of $W(T)$ on T, which we will prove later.)

Hence, we can view $W(T)$ as a subgroup of $\mathfrak{gl}(\mathfrak{t})$. Let us now put an inner product on $\mathfrak{t}$: there are many choices, but the compactness of K means there exists an Ad-*invariant inner product*.

LEMMA 16.25. *If K is a compact Lie group with Lie algebra $\mathfrak{k}$, there exists an inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{k}$ that is invariant under the Adjoint action of K:*

$$\langle \mathrm{Ad}(g)X, \mathrm{Ad}(g)Y \rangle = \langle X, Y \rangle \qquad g \in K,\ X, Y \in \mathfrak{k}.$$

PROOF. Begin by fixing *any* inner product $\langle\cdot,\cdot\rangle_0$ on $\mathfrak{k}$. Then define the desired inner product by averaging over the Adjoint action:

$$\langle X, Y\rangle \equiv \int_K \langle \mathrm{Ad}(h)X, \mathrm{Ad}(h)Y\rangle_0 \, dh$$

where dh denotes the unique Haar probability measure on K, cf. Proposition 12.6. It is quick to verify that this is a bilinear form, and it is positive definite since

$$\langle X, X\rangle = \int_K \langle \mathrm{Ad}(h)X, \mathrm{Ad}(h)X\rangle_0 \, dh$$

is in integral of the strictly positive function $h \mapsto \langle \mathrm{Ad}(h)X, \mathrm{Ad}(h)X\rangle_0$. Hence, this defines an inner product, and we can quickly check that

$$\langle \mathrm{Ad}(g)X, \mathrm{Ad}(g)Y\rangle = \int_K \langle \mathrm{Ad}(h)\mathrm{Ad}(g)X, \mathrm{Ad}(h)\mathrm{Ad}(g)Y\rangle_0 \, dh = \int_K \langle \mathrm{Ad}(hg)X, \mathrm{Ad}(hg)Y\rangle_0 \, dh$$

and this is equal to $\int_K \langle \mathrm{Ad}(h')X, \mathrm{Ad}(h')Y\rangle_0 \, dh' = \langle X, Y\rangle$ by the left-invariance of the Haar measure. □

REMARK 16.26. In fact, there is often a *unique* Ad-invariant inner product up to scale: for example, on $\mathfrak{su}(n)$, the Hilbert-Schmidt inner product $\langle X, Y\rangle = -\mathrm{Tr}(XY)$ is the only Ad-invariant inner product up to scale. (This is not true on $\mathfrak{u}(n)$, and the difference is a delicate matter of representation theory we may get to later.) It is also not necessary for a Lie group to be compact for it to possess an Ad-invariant inner product: for example, it is clear that $\mathfrak{k} \oplus \mathbb{R}^n$ does whenever $\mathfrak{k}$ does. It turns out this is a characterization: the Lie algebra of a Lie group G possesses an Ad-invariant inner product if and only if $G \cong K \times \mathbb{R}^n$ for some compact K and finite n. Such Lie groups are called *compact-type*.

Thus, if we imbue $\mathfrak{k}$ with an Ad-invariant inner product, then we see that the homomorphism $[\mathrm{Ad}]\colon W(T) \to \mathfrak{gl}(\mathfrak{t})$ of Proposition 16.24 maps into the orthogonal group $\mathfrak{o}(\mathfrak{t})$. We will see later that it is finite: it is the collection of reflections through the hyperplanes orthogonal to a certain finite set of vectors in $\mathfrak{k}$ (called the *roots*).

While we are on the topic of Ad-invariant inner products, let us note a few useful facts about homogeneous spaces in that context.

PROPOSITION 16.27. *Let G be a connected Lie group with Lie algebra $\mathfrak{g}$, and let $H \subseteq G$ be a closed subgroup with Lie algebra $\mathfrak{h} \subseteq \mathfrak{g}$. Suppose that Lie algebra $\mathfrak{g}$ posses an inner product that is invariant under the adjoint action of H: $\langle \mathrm{Ad}(h)X, \mathrm{Ad}(h)Y\rangle = \langle X, Y\rangle$ for all $h \in H$, $X, Y \in \mathfrak{g}$. Let $\mathfrak{f} = \mathfrak{g} \ominus \mathfrak{h}$ be the orthogonal complement of $\mathfrak{h}$ in $\mathfrak{g}$ with respect to this inner product. Then, for each $g \in G$, the map*

$$\tau_g\colon \mathfrak{f} \mapsto T_{[g]}(G/H), \quad X \mapsto \left.\frac{d}{dt}\right|_{t=0} [g \exp tX]$$

is a linear isomorphism. Moreover, if $g, g' \in G$ are in the same coset $[g] = [g']$, then for any $h \in H$ with $g' = gh$, we have

$$\tau_{g'}(X) = \tau_g(\mathrm{Ad}(h)X).$$

Note: in the proof that G/H is a smooth manifold (Theorem 15.4), we constructed the tangent space at the identity in this fashion, using a complementary subspace to $\mathfrak{h}$ in $\mathfrak{g}$ to give our slice coordinates. The first statement of the proposition holds in that level of generality (no inner product needed to choose $\mathfrak{f}$); the Ad-invariance is needed to characterize the change $\tau_g \to \tau_{g'}$.

PROOF. From Theorem 15.4, we know that the quotient map $\pi\colon G \to G/H$ is a submersion; in particular, at the identity $e \in G$, $d\pi_e\colon \mathfrak{g} \to T_{[e]}(G/H)$ is surjective. By point (2) of the theorem, its kernel is precisely $\mathfrak{h}$. So, for any given vector $v \in T_{[e]}(G/H)$, fix some $X \in \mathfrak{g}$ with $d\pi_e(X) = v$. We may uniquely decompose $X = X_0 + X_1$ where $X_0 \in \mathfrak{h}$ and $X_1 \in \mathfrak{g}$; thus, $v = d\pi_e(X_0) + d\pi_e(X_1) = d\pi_e(X_1)$, and so we see that $d\pi_e|_{\mathfrak{f}}$ is a linear surjection $\mathfrak{f} \to T_[e](G/H)$. It is also injective: if $X_1, X_2 \in \mathfrak{f}$ satisfy $d\pi_e(X_1) = d\pi_e(X_2)$, then $X_1 - X_2 \in \ker(d\pi_e) = \mathfrak{h}$, and since $X_1 - X_2 \in \mathfrak{f} = \mathfrak{h}^\perp$, we have $X_1 - X_2 = 0$.

Thus, $d\pi_e|_{\mathfrak{f}}\colon \mathfrak{f} \to T_{[e]}(G/H)$ is a linear isomorphism. Since $t \mapsto \exp tX$ is a smooth curve with derivative X at $t = 0$, by definition we can compute

$$d\pi_e(X) = \frac{d}{dt}\bigg|_{t=0} \pi(\exp tX) = \tau_e(X)$$

which proves the first statement in the case $g = e$. For other $g \in G$, we use the left action of G on G/H which is smooth by Theorem 15.4(3), and thus $dL_g|_{[e]}\colon T_{[e]}(G/H) \to T_{[g]}(G/H)$ is a linear isomorphism, and we may compose to find an isomorphism

$$dL_g|_{[e]} \circ d\pi_e\colon \mathfrak{f} \to T_{[g]}(G/H).$$

But by the chain tule $dL_g|_{[e]} \circ d\pi_e = d(L_g \circ \pi)_e$, and so as above

$$d(L_g \circ \pi)_e(X) = \frac{d}{dt}\bigg|_{t=0} g \cdot \pi(\exp tX) = \frac{d}{dt}\bigg|_{t=0} [g \exp tX] = \tau_g(X).$$

Finally, if $h \in H$ and $g' = gh$, then for all $t \in \mathbb{R}$ and $X \in \mathfrak{f}$ we have

$$[g' \exp tX] = [gh \exp tX] = [gh(\exp tX)h^{-1}] = [gC_h(\exp tX)] = [g \exp(t\mathrm{Ad}(h)X)].$$

Differentiating at $t = 0$ yields $\tau_{g'} = \tau_g \circ \mathrm{Ad}(h)$. □

REMARK 16.28. Note: the identity $\tau_{g'} = \tau_g \circ \mathrm{Ad}(h)$ appears to be valid in general here without any assumption on the inner product, but the domain of $\tau_{g'}$ is, in general, $\mathrm{Ad}(h)(\mathfrak{f})$. Of course, in order for τ_g to really be well-defined (since we must choose some $g \in [g]$), we need the domain to be consistent.

To conclude, as a corollary, we show that homogeneous spaces G/H with *compact* connected H possess unique (up to scale) Haar measures.

PROPOSITION 16.29. *Let G be a connected Lie group, and let $K \subseteq G$ be a connected compact subgroup. Then there exists a left-invariant volume form on G/K, which is unique up to scale. In particular, G/K is orientable.*

PROOF. Following precisely the proof of Lemma 16.25, we may choose an $\mathrm{Ad}(K)$-invariant inner product on $\mathfrak{g} = \mathrm{Lie}(G)$. Let $\mathfrak{f} = \mathfrak{g} \ominus \mathfrak{k}$ be the orthogonal complement of $\mathfrak{k} = \mathrm{Lie}(K)$ in $\mathfrak{g}$. Then, for $h \in K$, $\mathrm{Ad}(h)|_{\mathfrak{f}} \in \mathrm{O}(\mathfrak{f})$ by the $\mathrm{Ad}(K)$-invariance of the inner product, and since H is connected and $\mathrm{Ad}(e) = I \in \mathrm{SO}(\mathfrak{f})$ (the connected component of the identity in $\mathrm{O}(\mathfrak{f})$), we must have $\mathrm{Ad}(h)|_{\mathfrak{f}} \in \mathrm{SO}(\mathfrak{f})$ for all $h \in K$.

Now, fix an orientation on $\mathfrak{f}$, and let ω_0 be the standard volume form associated to it: i.e. the unique top form which satisfies $\omega_0(e_1, \ldots, e_n) = 1$ whenever $\{e_1, \ldots, e_n\}$ is an orientation-preserving orthonormal basis for $\mathfrak{f}$. Since $\mathrm{Ad}(h) \in \mathrm{SO}(\mathfrak{f})$ for any $h \in K$, it follows that

$$\omega_0(\mathrm{Ad}(h)v_1, \ldots, \mathrm{Ad}(h)v_n) = \omega(X_1, \ldots, X_n), \quad h \in K,\ v_1, \ldots, v_n \in \mathfrak{f}.$$

We use this to define a volume form on G/H using the identification of all tangent spaces with $\mathfrak{h}$ in Proposition 16.27: for any $g \in G$, we define a n-form $\lambda_g \colon T_{[g]}(G/H)^n \to \mathbb{R}$ by

$$\lambda_g(X_1|_{[g]} \dots, X_n|_{[g]}) = \omega_0(\tau_g^{-1}(X_1|_{[g]}), \dots, \tau_g^{-1}(X_n|_{[g]})).$$

If we choose a different $g' \in [g]$, then then with $h \in K$ given such that $g' = gh$, Proposition 16.27 shows that $\tau_{g'} = \tau_g \circ \mathrm{Ad}(h)$, which means that for $v_1, \dots, v_n \in T_{[g]}(G/H)$,

$$\begin{aligned}\lambda_{g'}(v_1, \dots, v_n) &= \omega_0(\tau_{g'}^{-1}(v_1), \dots, \tau_{g'}^{-1}(v_n)) \\ &= \omega_0(\mathrm{Ad}(h^{-1})\tau_g^{-1}(v_1), \dots, \mathrm{Ad}(h^{-1})\tau_g^{-1}(v_n)) \\ &= \omega_0(\tau_g^{-1}(v_1), \dots, \tau_g^{-1}(v_n)) \\ &= \lambda_g(v_1, \dots, v_n).\end{aligned}$$

Hence, we may define a (rough) antisymmetrix contravariant tensor field ω on G/H by

$$\omega_{[g]}(X_1|_{[g]}, \dots, X_n|_{[g]}) = \lambda_g(X_1|_{[g]}, \dots, X_n|_{[g]})$$

for any choice of $g \in [g]$. It is left as a simple exercise (from the definition of τ_g) to show that $\omega \in \Omega^n(G/H)$ (i.e. it is smooth), and left-invariant.

As for uniqueness up to scale, if ω and ω' are two volume forms on G/H, then (as top forms) $\omega|_{[e]}$ and $\omega'|_{[e]}$ are proportional, $\omega'_{[e]} = \lambda\omega_{[e]}$ for some $\lambda > 0$. Since the left action of G on G/H is transitive, for any $x \in G/H$, we may find some $g \in G$ with $g \cdot [e] = x$, and thus if ω and ω' are left-invariant, we have $\omega'|_x = L_g^*\omega'|_{[e]} = \lambda L_g^*\omega_{[e]} = \lambda\omega|_x$ as well. Hence, $\omega' = \lambda\omega$. □

4. Interlude: Mapping Degrees

Let M, N be smooth manifolds. We will be concerned with smooth maps $F \colon M \to N$, and their level sets at *regular* values. In order for there to exist any regular values at all, we must assume that M and N have the same dimension. The first observation is that, if the domain manifold M is compact, then such level sets are finite.

LEMMA 16.30. *Let M, N be smooth manifolds of the same dimension, with M compact, and let $F \colon M \to N$ be smooth. If $q \in N$ is a regular value for F, then $F^{-1}(q)$ is finite.*

PROOF. Suppose, for a contradiction, that $F^{-1}(q)$ is infinite. Since M is compact, it follows that the infinite set $F^{-1}(q) \subset M$ has an accumulation point $p \in M$, and so by continuity of F, $F(p) = q$. Since q is a regular value, by definition dF_p is invertible. By the inverse function theorem, it follows that there is a neighborhood of p in M where F is a diffeomorphism onto its image; in particular, F is injective on a neighborhood of p. This is a contradiction: since $F^{-1}(q)$ accumulates at p, any neighborhood of p contains infinitely many points that map to q. □

Now, suppose that M and N are oriented manifolds. If $F \colon M \to N$, then for any regular value $q \in N$ and preimage $p \in F^{-1}(q)$, the (invertible) derivative dF_p is either orientation-preserving, or orientation-reversing. (Of course, if it is orientation-preserving at p, it is also orientation preserving at any point in a neighborhood of p, and likewise for the orientation-reversing case.)

DEFINITION 16.31. *Let M, N be smooth oriented manifolds of the same dimension, with M compact, and let $F \colon M \to N$ be smooth. For each regular value $q \in M$ for f, the* **degree** *of F at q, $\deg_q(F)$, is the integer*

$$\deg_q(F) = \sum_{p \in F^{-1}(q)} \mathrm{sgn}(dF_p)$$

where, for any linear map L between oriented vector spaces, $\operatorname{sgn}(L) = +1$ *if L is orientation-preserving and* $\operatorname{sgn}(L) = -1$ *if L is orientation-reversing.*

The main theorem of this section is that this degree is constant: it does not depend on which regular value we compute it at.

THEOREM 16.32. *Let M, N be smooth, oriented, connected, compact manifolds of the same dimension ≥ 1, and let $F\colon M \to N$ be a smooth map. There is an integer* $\deg(F)$ *such that, for all regular values $q \in N$ of F,* $\deg_q(F) = \deg(F)$.

The constant integer $\deg(F)$ is called the **mapping degree** of F. We will prove Theorem 16.32 by showing that it can be calculated in an analytic way, involving how integrals of differential forms change under pullback by F. To motivate how, consider the following example.

EXAMPLE 16.33. Fix an integer $k \in \mathbb{Z}$, and let $F\colon \mathbb{S}^1 \to \mathbb{S}^1$ be the map $F(u) = u^k$. We give $\mathbb{S}^1$ the usual ccw orientation in both cases; then F is orientation-preserving at all points if $k > 0$ and orientation-reversing at all points if $k < 0$. Note, the exponential map $\exp\colon \mathbb{R} \to \mathbb{S}^1$, given by $\exp\theta = e^{i\theta}$, is a diffeomorphism from $(-\pi, \pi)$ onto $\mathbb{S}^1 \setminus \{-1\}$; on this neighborhood, we have $\hat{F}(\theta) = F(e^{i\theta}) = e^{ki\theta}$, and the total derivative is $D\hat{F}_\theta(X) = kie^{ki\theta}X$ which is a surjective for all θ so long as $k \neq 0$. We can use the shifted exponential map $\theta \mapsto e^{i(\theta+\pi)}$ to make the same conclusion about $\mathbb{S}^1 \setminus \{1\}$. So, as long as $k \neq 0$, every value of F is a regular value. Note also that, for any point $u \in \mathbb{S}^1$, $F^{-1}(u)$ has size k: it is the collection of kth roots of u (which is, up to a fixed rotation, the same as the set of kth roots of unity).

Now, let $\omega \in \Omega^1(\mathbb{S}^1)$; then we may write $\omega = f\,d\theta$ for some $f \in C^\infty(\mathbb{S}^1)$. (Recall that, even though θ is not a smooth function on all of $\mathbb{S}^1$, the global vector field $\frac{\partial}{\partial\theta}$ is smooth; $d\theta$ is its dual 1-form. It can be expressed as the restriction to $\mathbb{S}^1 \subset \mathbb{R}^2$ of the global 1-form $x\,dy - y\,dx$.) We now compare

$$\int_{\mathbb{S}^1} \omega \qquad \text{vs.} \qquad \int_{\mathbb{S}^1} F^*\omega.$$

If F were a diffeomorphism, then by Proposition 10.15(d), these two would be equal (up to a minus sign of F is orientation-reversing). Indeed, this is what happens when $k = \pm 1$. But, for other k, F is not injective. Instead, we can compute as follows. Using the exponential map as our parametrization $\exp\colon [0, 2\pi] \to \mathbb{S}^1$, we have

$$\int_{\mathbb{S}^1} \omega \equiv \int_\alpha \omega = \int_0^{2\pi} \alpha^*\omega = \int_0^{2\pi} f(e^{i\theta})\,d\theta.$$

At the same time, by Proposition 6.23(d), we have

$$\int_{\mathbb{S}^1} F^*\omega = \int_\alpha F^*\omega = \int_{F\circ\alpha} \omega = \int_0^{2\pi} f(e^{ik\theta})\,d(k\theta).$$

For the latter, we make the substitution $\phi = k\theta$. Then this becomes

$$\int_0^{2k\pi} f(e^{i\phi})\,d\phi = \sum_{j=1}^{k} \int_{2(j-1)\pi}^{2j\pi} f(e^{i\phi})\,d\phi = k\int_0^{2\pi} f(e^{i\phi})\,d\phi.$$

Thus, we have $\int_{\mathbb{S}^1} F^*\omega = k\int_{\mathbb{S}^1} \omega$.

Example 16.33 demonstrates how we must modify Proposition 10.15(d) when the pullback function F is not a diffeomorphism. In this example, the value of the integral gets multiplied by

an integer, and this integer happens to be exactly the degree of F. This is, in fact, always true, and gives us a way to access the mapping degree of F using differential forms.

LEMMA 16.34. *Let M, N be compact, connected, oriented n-manifolds, and let $F\colon M \to N$ be a smooth map. Let $q \in N$ be a regular value for F. Then for all sufficiently small neighborhoods V of q, and all n-forms ω supported in V,*

$$\int_M F^*\omega = \deg_q(F) \int_N \omega.$$

PROOF. By Lemma 16.30, the preimage $F^{-1}(q)$ is a finite set $F^{-1}(q) = \{p_1, \dots, p_m\}$. By assumption dF_{p_j} is an isomorphism for each j, and so by the inverse function theorem there is some neighborhood U_j of p_j such that $F|_{U_j}$ is a diffeomorphism onto its image. By shrinking the neighborhoods as necessary, we may assume that dF_p is orientation-preserving for all $p \in U_j$ or orientation-reversing for all $p \in U_j$ for all j; and we may assume that the U_j are all disjoint.

Let V be any neighborhood of q contained in $\bigcap_j F(U_j)$. Thus, if ω is supported in V, then $F^*\omega$ is supported in $\bigsqcup_j U_j$, and we have

$$\int_M F^*\omega = \sum_{j=1}^m \int_{U_j} F^*\omega.$$

But F is a diffeomorphism on U_j, and so by Proposition 10.15(d), each term in the sum is equal to $\pm \int_{F(U_j)} \omega = \pm \int_V \omega$, with the sign determined by whether F is orientation-preserving or orientation-reversing. This concludes the proof. □

Now, fix a regular point q for F and a neighborhood V as in Lemma 16.34, and let ω be any top form supported in V for which $\int_N \omega \neq 0$. Then the lemma shows us that we can compute $\deg_q(F)$ as

$$\deg_q(F) = \frac{\int_M F^*\omega}{\int_N \omega}. \tag{16.2}$$

Thus, to show $\deg_q(F)$ does not depend on q, our approach will be to show that the ratio in (16.2) does not really depend on the form ω. In particular, given two regular values q, q' with neighborhoods V, V' and top forms ω, ω', we will show that we can deform ω to ω' while holding the numerator and denominator in (16.2) constant.

PROPOSITION 16.35. *Let N be a compact oriented n-manifold. Let $\Psi\colon [0,1] \times N \to N$ be a continuous map, and denote $\Psi_t(q) = \Psi(t,q)$. Suppose Ψ is piecewise smooth, in the sense that there is a partition $\{0 = t_0 < t_1 < \cdots < t_m = 1\}$ of $[0,1]$ so that $\Psi|_{(t_{j-1},t_j)\times N}$ is smooth for all $1 < j \leq m$. Suppose also that $\Psi_0 = \mathrm{Id}_N$, and Ψ_t is an orientation-preserving diffeomorphism for each $t \in [0,1]$.*

Let $\omega \in \Omega^n(N)$, and for each t, let $\omega_t = \Psi_t^*\omega$. Then, for $0 \leq t \leq 1$,

$$\int_N \omega_t = \int_N \omega. \tag{16.3}$$

Moreover, let M be another compact oriented n-manifold, and let $F\colon M \to N$ be smooth. Then, for $0 \leq t \leq 1$,

$$\int_M F^*\omega_t = \int_M F^*\omega. \tag{16.4}$$

PROOF. It suffices to prove the proposition assuming Ψ is smooth; for then we can apply the result on each subinterval $[t_{j-1}, t_j]$, and then use continuity to conclude the full result. Note, moreover, that (16.3) is an immediate consequence of Proposition 10.15(d), because Ψ_t is an orientation-preserving diffeomorphism. So we need only establish (16.4). For this purpose, we will use Stokes's theorem for manifolds with boundary (which we have not covered, so you will have to take some of this on faith).

Note that $F^*\omega_t = F^*(\Psi_t^*\omega) = (\Psi_t \circ F)^*\omega$. So, define $G\colon [0,1] \times M \to N$ by $G(t,p) = \Psi(t, F(p))$. Then for each $t \in [0,1]$, noting that $\omega = \omega_0$, we have

$$\int_M F^*\omega_t - \int_M F^*\omega = \int_{\partial([0,t]\times M)} G^*\omega = \int_{[0,t]\times M} d(G^*\omega) = \int_{[0,t]\times M} G^*(d\omega).$$

But ω is a top form on N, and so $d\omega = 0$. Thus, the above integral is 0, showing that $\int_M F^*\omega_t = \int_M F^*\omega$, as desired. □

We will then be able to prove our main theorem provided we can find such a piecewise-smooth family of diffeomorphisms exchanging any two regular values of F. In fact, we can exchange any two points.

LEMMA 16.36. *Let N be a connected, oriented manifold, and let $q, q' \in N$. There exists a piecewise smooth continuous map $\Psi\colon [0,1] \times N \to N$ as in Proposition 16.35 such that $\Psi(1, q') = q$.*

PROOF. This is a standard connectedness argument. We fix q, and consider the set E of all q' for which such a Ψ. Then one sees that E is open by working in local coordinates (where the family of diffeomorphisms can be constructed as a piecewise linear extension). Similarly, one sees that the set E is closed, by working in local coordinates near any limit point of E. The details are left to the interested reader. □

REMARK 16.37. In our intended use of all the results in this section, the manifold N will be a compact, connected Lie group. As such, Lemma 16.36 is very easy to prove. Let $\alpha\colon [0,1] \to N$ be a continuous, piecewise smooth curve connecting e to $q(q')^{-1}$. Then we can define $\Psi_t = L_{\alpha(t)}$, which is a smooth family of orientation-preserving diffeomorphisms (taking any left-invariant orientation), $\Psi_0 = \mathrm{Id}_N$, and $\Psi_1(q') = q$, as desired.

This brings us to the proof of Theorem 16.32.

PROOF OF THEOREM 16.32. It suffices to show that $\deg_q(F) = \deg_{q'}(F)$ for any two regular values $q, q \in N$ of F. According to Lemma 16.34, we may compute $\deg_q(F)$ by selecting a sufficiently small neighborhood V of q and a top form ω supported in V with $\int_N \omega \neq 0$; then

$$\deg_q(F) = \frac{\int_M F^*\omega}{\int_N \omega}. \tag{16.5}$$

Now, choose a map $\Psi\colon [0,1] \times N \to N$ as in Lemma 16.36. Since $\Psi_1(q') = q$, $\Psi_1^*\omega$ is supported in a neighborhood of q'. Shrinking the neighborhood V if necessary, we may assume that the support set of $\Psi_1^*\omega$ satisfies the conditions of Lemma 16.34 at q', and so

$$\deg_{q'}(F) = \frac{\int_M F^*\Psi_1^*\omega}{\int_N \Psi_1^*\omega}.$$

But, by Proposition 16.35, this ratio is equal to the one on the right-hand-side of (16.5). This concludes the proof. □

We now give a corollary to Theorem 16.32 which we will use in the proof of the torus theorem.

COROLLARY 16.38. *Let M, N be compact, connected, oriented manifolds, and let $F\colon M \to N$ be a smooth map with at least one regular value. If $\deg(F) \neq 0$, then $F(M) = N$: F is surjective.*

The idea is simply this: if $q \notin F(M)$, then q is vacuously a regular value of F, and since $F^{-1}(q) = \varnothing$, $\deg_q(F) = 0$; since the mapping degree is constant, this means $\deg(F) = 0$. This may feel like cheating, so we give a more detailed explanation below.

PROOF. Suppose that there is some point $q \in N \setminus F(M)$. Since M is compact, $F(M)$ is compact, and so $N \setminus F(M)$ is open. Thus, there is some neighborhood V of q contained in $N \setminus F(M)$. Fix some top form ω on N that is supported in V, such that $\int_N \omega \neq 0$. Since $F^*\omega = 0$,

$$\frac{\int_M F^*\omega}{\int_N \omega} = 0.$$

Now, fix a regular value $q' \in F(M)$, and let $\Psi\colon [0,1] \times N \to N$ be a family of diffeomorphisms with $\Psi_1(q') = q$ as in Lemma 16.36. Then by Theorem 16.32, Lemma 16.34, and Proposition 16.35, we have

$$\deg(F) = \deg_{q'}(F) = \frac{\int_M F^*\Psi_1^*\omega}{\int_N \Psi_1^*\omega} = \frac{\int_M F^*\omega}{\int_N \omega} = 0.$$

□

It is also worth recording that the identification of the mapping degree in Lemma 16.34 can be extended to all top forms, not just those supported in a sufficiently small neighborhood of a regular point.

PROPOSITION 16.39. *Let M, N be compact, connected, oriented manifolds of the same dimension $n \geq 1$, and let $F\colon M \to N$ be smooth. Then for every top form $\omega \in \Omega^n(N)$,*

$$\int_M F^*\omega = \deg(F) \int_N \omega.$$

The proof is left as a homework exercise.

5. Cartan's Torus Theorem

We now proceed to prove Theorem 16.17. Let K be a connected, compact Lie group, and fix a maximal torus $T \subseteq K$. Typically, T is not a normal subgroup, but it is an embedded Lie subgroup (as compact submanifolds are always embedded, cf. Proposition 9.16(3)), and so, by Theorem 15.4, the quotient K/T is a homogeneous K-space: a smooth manifold of dimension $\dim(K) - \dim(T)$ in possession of a transitive, smooth, left action of K.

DEFINITION 16.40. *Define a map*

$$\Phi\colon T \times K/T \to K, \quad \Phi(t, [x]) = xtx^{-1}.$$

This map is well-defined: for any other element $y \in K$ with $[y] = yT = xY = [x]$, there is some $s \in T$ with $y = xs$, and so

$$yty^{-1} = (xs)t(xs)^{-1} = x(sts^{-1})x^{-1} = xtx^{-1}$$

since $s, t \in T$ and T is abelian. Note, moreover, that the map $\pi\colon T \times K \to T \times K/T$ given by $\pi(t, x) = (t, [x])$ is a surjective submersion, and $\Phi \circ \pi(t, x) = xtx^{-1}$ is smooth; thus, by Lemma 15.5, Φ is smooth.

We will show that Φ is surjective; Theorem 16.17 then follows immediately (for then every $y \in K$ is of the form $y = \Phi(t, [x]) = xtx^{-1}$ for some $x \in K$ and $t \in T$). We will show that Φ is surjective using Corollary 16.38. We will show that K/T is compact and orientable, and that the degree of Φ is nonzero. In fact, we will see that elements $t \in T \subset K$ for which $\langle t \rangle$ is dense in T are regular values for Φ, and we will compute $\deg(\Phi) = \deg_t(\Phi)$ explicitly at such points. We require the following lemma, which identifies the preimage of Φ on such points t using the Weyl group of T.

LEMMA 16.41. *Let $t \in T$ be such that the subgroup $\langle t \rangle$ is dense in t. Then*

$$\Phi^{-1}(\{t\}) = \{(x^{-1}tx, [x])\colon [x] \in W(T) = N(T)/T \subseteq K/T\}.$$

In particular, if $xsx^{-1} = t$ for some $x \in K$ and $s \in T$, then s must be of the form $s = w^{-1} \cdot t$ for some $w \in W(T)$.

PROOF. If $x \in N(T)$, then $x^{-1}tx \in T$, and so $\Phi(x^{-1}tx, [x]) = t$, which shows the reverse containment $\supseteq$. Conversely, let $s \in T$ and $x \in K$ with $(s, [x]) \in \Phi^{-1}(t)$. Then $xsx^{-1} = \Phi(s, [x]) = t$, and so $x^{-1}t^kx = s^k \in T$ for all $k \in \mathbb{Z}$. Since $\{t^k\colon k \in \mathbb{Z}\}$ is dense in T, it follows that $x^{-1}Tx \subseteq T$. But T is a maximal torus, and therefore so is $x^{-1}Tx$; thus, $x^{-1}Tx = T$, which implies that $T = xTx^{-1}$, so $x \in N(T)$, as required.

Finally, if $xsx^{-1} = t$, then $(s, [x]) \in \Phi^{-1}(\{t\})$, and so by the preceding paragraph $[x] \in W(T)$. We then have $s = x^{-1}tx = w^{-1} \cdot t$ where $w = [x]$ and $\cdot$ denotes the usual conjugation action of $W(T)$ on T. $\square$

REMARK 16.42. The lemma gives a bijection between $\Phi^{-1}(\{t\})$ and $W(T)$; this will be important shortly.

Now we must compute $d\Phi_t$ at such points t. To succinctly express the differential, we identify tangent spaces as in Proposition 16.27. Fixing an Ad-invariant inner product on K, we decompose $\mathrm{Lie}(K) = \mathfrak{k} = \mathfrak{t} \oplus \mathfrak{f}$ where $\mathfrak{t} = \mathrm{Lie}(T)$ and $\mathfrak{f} = \mathfrak{t}^\perp$. At each point $[x] \in K/T$, we select some representative element $x \in K$ $g \in [g]$, and then identify $\tau_x\colon \mathfrak{f} \to T_{[x]}(K/T)$; thus, we identify the tangent space $T_{(t,[x])}(T \times (K/T)) \cong \mathfrak{t} \oplus \mathfrak{f}$. We may also use Proposition 16.27 to identify the tangent space $T_{[y]}K = T_{[y]}(K/\{e\})$ with $\mathfrak{k}$ for each $y \in K$, and also to identify $T_t(T) = T_t(T/\{e\})$ with $\mathfrak{t}$. To summarize:

$$T_{(t,[x])}(T \times (K/T)) \cong \mathfrak{t} \oplus \mathfrak{f}, \qquad T_{[x]}K \cong \mathfrak{k}, \qquad T_t(T) \cong \mathfrak{t}.$$

Hence, $d\Phi_{(t,[x])}\colon T_{(t,[x])}(T \times K/T) \to T_{xtx^{-1}}K$ can be viewed as a linear map

$$d\Phi_{(t,[x])}\colon \mathfrak{t} \oplus \mathfrak{f} \to \mathfrak{k} = \mathfrak{t} \oplus \mathfrak{f}.$$

That is, it can be represented as a 2×2 matrix of linear maps. In this form, we can easily compute it.

LEMMA 16.43. *Let $(t, [x]) \in T \times K/T$. Identify the tangent spaces with $\mathfrak{t}$, $\mathfrak{f}$, and $\mathfrak{t} \oplus \mathfrak{f}$ via Proposition 16.27 as above. Then*

$$d\Phi_{(t,[x])} = \mathrm{Ad}(x) \begin{bmatrix} I & 0 \\ 0 & \mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I \end{bmatrix}.$$

PROOF. For $X \in \mathfrak{t}$, we compute that

$$\begin{aligned}\left.\frac{d}{d\tau}\right|_{\tau=0} \Phi(t\exp\tau X,[x]) &= \left.\frac{d}{d\tau}\right|_{\tau=0} xt\exp(\tau X)x^{-1} \\ &= \mathrm{Ad}(x)\left(\left.\frac{d}{d\tau}\right|_{\tau=0} t\exp(\tau X)\right) = \mathrm{Ad}(x)\circ dL_t(X).\end{aligned}$$

Now, note that for any $y \in K$,

$$C_x(ty) = xtyx^{-1} = xtx^{-1}(C_x(y)).$$

Differentiating with respect to x, we have

$$\mathrm{Ad}(x)\circ dL_t = dL_{xtx^{-1}}\circ \mathrm{Ad}(x).$$

But the identification of $T_{\Phi(t,[x])}K$ with $\mathfrak{k}$ is precisely via the left action of xtx^{-1}, this shows that

$$d\Phi_{(t,[x])}(X,0) = \mathrm{Ad}(x)X.$$

On the other side, we use the product rule of Lemma 14.18.

$$\begin{aligned}\left.\frac{d}{d\tau}\right|_{\tau=0} \Phi(t,[x\exp\tau Y]) &= \left.\frac{d}{d\tau}\right|_{\tau=0} x\exp(\tau Y)t\exp(-\tau Y)x^{-1} \\ &= \left.\frac{d}{d\tau}\right|_{\tau=0} x\exp(\tau Y)tx^{-1}\cdot x\exp(-\tau Y)x^{-1} \\ &= \mathrm{Ad}(x)\circ dR_t(Y) - dL_{xtx^{-1}}\circ \mathrm{Ad}(x)Y.\end{aligned}$$

Differentiating the identity $xytx^{-1} = xtx^{-1}(xt^{-1}ytx^{-1})$ with respect to Y, we can rewrite

$$\mathrm{Ad}(x)\circ dL_t(Y) = dL_{xtx^{-1}}\circ \mathrm{Ad}(xt^{-1})Y = dL_{xtx^{-1}}\circ \mathrm{Ad}(x)\circ \mathrm{Ad}(t^{-1})(Y).$$

Hence, we have

$$\left.\frac{d}{d\tau}\right|_{\tau=0} \Phi(t,[x\exp\tau Y]) = dL_{xtx^{-1}}\circ \mathrm{Ad}(x)\left(\mathrm{Ad}(t^{-1}) - I\right)Y.$$

As above, since $T_{\Phi(t,[x])}K$ is identified with $\mathfrak{k}$ via the left action of xtx^{-1}, this shows that

$$d\Phi_{(t,[x])}(0,Y) = \mathrm{Ad}(x)\left(\mathrm{Ad}(t^{-1}) - I\right)Y.$$

Since $d\Phi_{(t,[x])}(X,Y) = d\Phi_{(t,[x])}(X,0) + d\Phi_{(t,[x])}(0,Y)$, this completes the calculation. □

Our goal is to show (that for a particular point $(t,[x])$) this differential is invertible (so $\Phi(t,[x])$ is a regular value), and then compute the degree of Φ at that point. To do so, we need orientations on T and K/T. Note that both are orientable: T is a Lie group, and so is orientable by Proposition 12.4, and K/T is orientable by Proposition 16.29. To fix orientations on these manifolds, we fix orientations on $\mathfrak{t}$ and $\mathfrak{f}$ (which then gives us an orientation on $\mathfrak{k} = \mathfrak{t}\oplus\mathfrak{f}$). We then use the identifications of all tangent spaces to T and K/T with these spaces in order to define orientations on the manifolds. Note that this is well-defined: for example, to identify $T_{[x]}(K/T)$ with $\mathfrak{f}$, we choose one of the maps τ_x with $x \in [x]$. If we were to choose another $\tau_{x'}$ with $[x'] = [x]$ (meaning $x' = xt$ for some $t \in T$), then (by Proposition 16.27) $\tau_{x'} = \tau_x \circ \mathrm{Ad}(t)$. As T is connected, $t \mapsto \det \mathrm{Ad}(t)$ has constant sign, and so all such identifications produce the same orientation on each tangent space. It gives rise to a global smooth orientation form because the maps $x \mapsto \tau_x$ are smooth.

Our next lemma shows when the second factor $\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I$ in Lemma 16.43 is invertible, and in that case, how it affects the orientations of the tangent spaces, in the special case that t generates a dense subgroup of T.

LEMMA 16.44. *Let $t \in T$ be an element that generates a dense subgroup. Then $\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I$ is an invertible linear transformation of $\mathfrak{f}$. Moreover, for all $w \in W(T)$,*

$$\det(\mathrm{Ad}(w \cdot t^{-1})|_{\mathfrak{f}} - I) = \det(\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I).$$

PROOF. For invertibility, we show that $\mathrm{Ad}(t^{-1})|_{\mathfrak{f}}$ does not have 1 as an eigenvalue. Suppose that there is some $Y \in \mathfrak{f}$ with $\mathrm{Ad}(t^{-1})Y = Y$. By infuction, it follows tha $\mathrm{Ad}(t^m)Y = Y$ for all $m \in \mathbb{Z}$. Since $\{t^m \colon m \in \mathbb{Z}\}$ is dense in T, by continuity we have $\mathrm{Ad}(s)Y = Y$ for all $s \in T$. It then follows that, for any $X \in \mathfrak{t}$,

$$[X, Y] = \mathrm{ad}(X)Y = \frac{d}{d\tau}\bigg|_{\tau=0} \mathrm{Ad}(\exp \tau X)Y = \frac{d}{d\tau}\bigg|_{\tau=0} Y = 0.$$

Since T is a maximal torus, $\mathfrak{t}$ is a maximal abelian subalgebra, and so since Y commutes with $\mathfrak{t}$, it follows that $Y \in \mathfrak{t} = \mathfrak{f}^{\perp}$. Thus, $Y = 0$. Hence, there is no non-zero eigenvector of 1 for $\mathrm{Ad}(t^{-1})|_{\mathfrak{f}}$, proving that $\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I$ is invertible.

For the second point, if $w \in W(T) = N(T)/T$ is represented by the element $x \in N(T)$, then

$$\mathrm{Ad}(w \cdot t^{-1}) = \mathrm{Ad}(xt^{-1}x^{-1}) = \mathrm{Ad}(x)\mathrm{Ad}(t^{-1})\mathrm{Ad}(x)^{-1}$$

and so $\mathrm{Ad}(w \cdot t^{-1}) - I = \mathrm{Ad}(x)(\mathrm{Ad}(t^{-1}) - I)\mathrm{Ad}(x)^{-1}$, showing that the two have the same determinant. □

We can now prove Theorem 16.17.

PROOF OF THEOREM 16.17. Let $t \in T$ be an element that generates a dense subgroup. By Lemma 16.41, the preimage $\Phi^{-1}(t)$ consists of those $(x^{-1}tx, [x])$ for which $[x] \in W(T)$. By Lemmas 16.43 and 16.44, for any $[x] \in K/T$, $(t, [x])$ is a regular point of Φ. Hence t is a regular value of Φ. It follows from Lemma 16.30 that $\Phi^{-1}(t)$ is a finite set. This set is in bijective correspondence with $W(T)$ by Lemma 16.41, and so we have proved that $W(T)$ is finite.

Now, referring to the formula in Lemma 16.43 for $d\Phi_{(t,[x])}$, the first factor $\mathrm{Ad}(x)$ is in $\mathrm{SO}(\mathfrak{k})$ because the inner product is Ad-invariant and K is connected, thus it has determinant 1 for any $x \in K$. As for the subfactor $\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I$: it appears to depend only on t, not on $[x]$. We should note, however, that the identification of the tangent space to K/T at $[x]$ is only defined up to conjugation. However, if $s, t \in T$ are in the same K-conjugacy class, then by Lemma 16.41 there is some $w \in W(T)$ such that $s = w \cdot t$, and then by Lemma 16.44, $\mathrm{Ad}(s^{-1}) - I$ and $\mathrm{Ad}(t^{-1}) - I$ have the same determinant over $\mathfrak{f}$. Hence, $d\Phi_{(t,[x])}$ has the same determinant at each point $(t, [x]) \in \Phi^{-1}(t)$. It follows that $\deg_t \Phi = |W(T)|$ or $= -|W(T)|$, depending on the (constant) sign of this determinant. In either case, since $|W(T)| \geq 1$, this shows that the mapping degree of Φ is not zero. Hence, by Corollary 16.38, Φ is surjective.

Thus, for every element $y \in K$, there is a pair $(t, [x]) \in T \times K/T$ so that $y = \Phi(t, [x]) = xtx^{-1}$ for any $x \in [x]$. This proves the torus theorem. □

In the midst of the proof, we saw that the Weyl group is finite. We record this as a separate corollary, with some other properties that also follow easily at this point.

COROLLARY 16.45. *Let K be a compact, connected Lie group. The Weyl groups $W(T)$ of all maximal tori in K are all isomorphic (by inner automorphisms of K). Henceforth, we denote this common finite group as $W = W(K)$.*

PROOF. As noted in the proof above, $W(T)$ is in bijective correspondence with the level set of a regular value of $\Phi\colon T \times K/T \to K$, and since $T \times K/T$ is compact, $W(T)$ is finite. Now, let T and S be two maximal tori in K. By Theorem 16.19, there is some $x \in K$ with $S = xTx^{-1}$. Then the conjugation map $C_{x^{-1}}$ restricts to an isomorphism from $N(S)$ onto $N(T)$, which we see as follows.

$C_{x^{-1}}$ is it (of course) an injective group homomorphism. Let $y \in N(S)$. Any $t \in T$ can be written as $t = x^{-1}sx$ for some $s \in S$. Then

$$C_{x^{-1}}(y) \cdot t \cdot C_{x^{-1}}(y)^{-1} = x^{-1}yx \cdot x^{-1}sx \cdot (x^{-1}yx)^{-1} = C_{x^{-1}}(C_y(s)).$$

Since $y \in N(S)$, $C_y(s) \in S$, and since $T = x^{-1}Sx$, it follows that $C_{x^{-1}}(C_y(s)) \in T$. Thus $C_{x^{-1}}(y) \in N(T)$. The same argument shows that C_x maps $N(T)$ into $N(S)$, and since $(C_x)^{-1} = C_{x^{-1}}$, this shows that $C_{x^{-1}}$ is a group isomorphism from $N(S)$ onto $N(T)$.

Finally, as shown in the proof of Proposition 16.24 (in the preceding discussion), $C_{x^{-1}}$ descends to a homomorphism from $W(S)$ to $W(T)$, and again since its inverse is C_x, it follows that this is an isomorphism. □

6. The Weyl Integration Formula

In the previous section, we saw that if K is a compact connected Lie group and $T \subseteq K$ is a maximal torus, then the map

$$\Phi\colon T \times K/T \to K, \quad \Phi(t, [x]) = xtx^{-1}$$

is smooth and surjective. In this section, we will use this map to give a decomposition of the Haar measure on K, which is (in a sense) a generalization of spherical coordinates on $\mathrm{SU}(2) \cong \mathbb{S}^3$, and will play an important role in the representation theory of compact Lie groups.

As usual, we fix an Ad-invariant inner product on K, and let $\mathfrak{f}$ be the orthocomplement of the Lie algebra $\mathfrak{t}$ of T in $\mathfrak{k} = \mathrm{Lie}(K)$. Identifying the tangent spaces or T, K/T, and K with $\mathfrak{t}$, $\mathfrak{f}$, and $\mathfrak{k} = \mathfrak{t} \oplus \mathfrak{f}$ using Proposition 16.27, we then expressed the differential of Φ as a 2×2 matrix of linear maps. Of greatest importance to us here is the entry of that matrix (up to a global Ad-factor) mapping $\mathfrak{f}$ into $\mathfrak{f}$. Let $\rho\colon T \to \mathbb{R}$ denote the function

$$\rho(t) = \det(\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I). \tag{16.6}$$

In Lemma 16.44, we showed that this function is non-zero whenever t generates a dense subgroup of T.

Now, the compact Lie groups K and T possess unique Haar (left-invariant) probability measures (cf. Proposition 12.6), and the homogeneous space K/T possesses a unique left-K-invariant probability measure (cf. Proposition 16.29). The main theorem of this section is the following integration formula, decomposing the Haar measure on K in terms of the measures on T and K/T.

THEOREM 16.46 (Weyl Integration Formula). *Let W be the Weyl group of K. For all continuous function $f\colon K \to \mathbb{R}$,*

$$\int_K f(x)\,dx = \frac{1}{|W|}\int_T \rho(t)\left(\int_{K/T} f(yty^{-1})\,d[y]\right)dt \tag{16.7}$$

where dx, dt, and $d[y]$ denote the left-invariant probability measures on K, T, and K/T respectively.

The theorem will be most useful in the special case that f is a **class function**, meaning that it is constant on conjugacy classes: $f(yxy^{-1}) = f(x)$ for all $x, y \in K$. IN this case, the Weyl integration formula takes the special form

$$\int_K f(x)\,dx = \frac{1}{|W|}\int_T \rho(t)f(t)\,dt. \tag{16.8}$$

(This follows immediately from (16.7), since the inside integral there reduces to $\int_{K/T} f(t)\,d[y]$ which equals $f(t)$ since the measure on K/T is normalized.)

Before proving the Weyl integration formula, we give an example of how it looks in the group $\mathrm{SU}(2)$.

EXAMPLE 16.47. Let $K = \mathrm{SU}(2)$, and let $T \subset K$ be the maximal torus of diagonal elements, cf. Examples 16.2 and 16.13. This torus is 1-dimensional, consisting of all matrices of the form

$$t = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}, \qquad \theta \in [-\pi, \pi].$$

The Lie algebra $\mathfrak{t}$ of this subgroup is the one-dimensional algebra spanned by the matrix $\mathrm{diag}(1, -1)$.

Now, fix the usual Hilbert-Schmidt inner product on $\mathfrak{su}(2)$. The orthogonal complement of $\mathfrak{t}$ is then the two-dimensional subspace

$$\mathfrak{p} = \mathfrak{t}^\perp = \left\{ X = \begin{bmatrix} 0 & x+iy \\ -x+iy & 0 \end{bmatrix} : x, y \in \mathbb{R} \right\}.$$

Direct computation shows that

$$\mathrm{Ad}(t^{-1})(X) = \begin{bmatrix} 0 & e^{-2i\theta}(x+iy) \\ e^{2i\theta}(-x+iy) & 0 \end{bmatrix}.$$

This shows that $\mathrm{Ad}(t^{-1})$ acts as a rotation by angle -2θ, and we can then compute that

$$\rho(t) = \det(\mathrm{Ad}(t^{-1}) - I) = \det \begin{bmatrix} \cos(2\theta) - 1 & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) - 1 \end{bmatrix} = 4\sin^2\theta.$$

As for the Weyl group: following the reasoning in Example 16.23, the Weyl group of $\mathrm{SU}(2)$ is the symmetric group Σ_2, acting via $\{I, -I\}$, so $|W| = 2$. Using the normalized Haar measure $d\theta/2\pi$ on $T \cong \mathbb{S}^1$, the Weyl integration formula for class functions in this case then reads

$$\int_{\mathrm{SU}(2)} f(x)\,dx = \frac{1}{2}\int_{-\pi}^{\pi} f(\mathrm{diag}(e^{i\theta}, e^{-i\theta})) \cdot 4\sin^2\theta\,\frac{d\theta}{2\pi}.$$

We can make one further simplification: the matrices $\mathrm{diag}(e^{i\theta}, e^{-i\theta})$ and $\mathrm{diag}(e^{-i\theta}, e^{i\theta})$ are conjugate in $\mathrm{SU}(2)$:

$$\begin{bmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}^{-1}.$$

So any class function takes the same value on both. Thus, we may restrict the integral to $[0, \pi]$, and by symmetry

$$\int_{\mathrm{SU}(2)} f(x)\,dx = \frac{2}{\pi}\int_0^{\pi} f(\mathrm{diag}(e^{i\theta}, e^{-i\theta})) \cdot \sin^2\theta\,d\theta. \tag{16.9}$$

Now, $\mathrm{SU}(2)$ can be identified with $\mathbb{S}^3$, as you showed on a homework set. In another homework exercise, you will show that, under this identification, two matrices in $\mathrm{SU}(2)$ are conjugate if and only if they represent two points on the sphere that have the same polar angle θ. Thus, class functions $\mathrm{SU}(2) \to \mathbb{R}$ are those function $\mathbb{S}^3 \to \mathbb{R}$ that depend *only* on the polar angle θ. The

interested reader can verify that (16.9) is precisely the usual spherical polar integration formula for the 3-sphere, in the special case that the function depends only on the polar angle. (The volume element in general is $\sin^2\theta \sin\phi_1\, d\theta\, d\phi_1\, d\phi_2$, with ϕ_2 ranging over $[0, 2\pi]$ and θ and ϕ_1 ranging over $[0, \pi]$. Note: the volume of $\mathbb{S}^3$ in these standard coordinates is $2\pi^2$. Integrating out ϕ_1 and ϕ_2 gives a factor of 4π, and renormalizing to give a probability measure on $\mathrm{SU}(2)$ then gives a constant of $4\pi \cdot \frac{1}{2\pi^2} = \frac{2}{\pi}$ as we see here.)

Let us now prove Theorem 16.46, up to a constant.

PROOF OF THEOREM 16.46, UP TO A CONSTANT. Let us, for the moment, assume f is smooth. Then $f(x)\, dx$ is a top form on K. Since $\Phi\colon T \times K/T \to K$ is smooth, Proposition 16.39 shows that

$$\deg(\Phi) \int_K f(x)\, dx = \int_{T\times K/T} \Phi^*(f(x)\, dx) = \int_{T\times K/T} (f \circ \Phi)\Phi^*(dx).$$

We showed in the proof of Theorem 16.17 that $\deg(\Phi) = \pm|W|$; we can certainly reverse the orientation on anyone of the three factors if need be to make it positive (and we assume that choice in what follows). Thus

$$\int_K f(x)\, dx = \frac{1}{|W|} \int_{T\times K/T} (f \circ \Phi)\, \Phi^*(dx). \tag{16.10}$$

Equation (16.10) holds for smooth f. Since K and $T \times K/T$ are compact topological spaces, the Weierstrass approximation theorem shows that smooth functions are uniformly dense in continuous functions on both sides, and since the integrals behave well with respect to uniform limits, it follows that (16.10) holds for all continuous functions.

Since $\Phi(t, [y]) = yty^{-1}$, it now suffices to show that $\Phi^*(dx) = \rho(t)\, d[y] \wedge dt$. To see why this is true, up to a constant, we refer to the proof of Proposition 16.29, where we constructed the left-invariant volume form on K/T. The form $\omega_{K/T}$ is defined by selecting a volume form on $\mathfrak{f}$, and extending it to K/T using the identifications of tangent spaces of K/T with $\mathfrak{f}$ via Proposition 16.27. So, fix a volume form ω_0 on $\mathfrak{f}$, and also fix a volume form ω_1 on $\mathfrak{f}$. By the same procedure, we may regard ω_1 as a left-invariant volume form on T. At the same time, $\omega_1 \wedge \omega_0$ is a volume form on $\mathfrak{t} \oplus \mathfrak{f} = \mathfrak{k}$, and again we may regard this is a left-invariant volume form on K. By uniqueness of left-invariant volume forms in all three cases, ω_1, ω_0, and $\omega_1 \wedge \omega_0$ are equal to the volume forms dt, $d[y]$, and dx on T, K/T, and K up to scale.

Now, fix an orthonormal basis $\{X_1, \dots, X_r\}$ for $\mathfrak{t}$ and an orthonormal basis $\{Y_1, \dots, Y_m\}$ for $\mathfrak{f}$; their union is an orthonormal basis for $\mathfrak{k} = \mathfrak{t} \oplus \mathfrak{f}$. We may choose the volume forms ω_1 and ω_0 to be normalized in these basis, in which case $\omega_1 \wedge \omega_0$ is noramlized in the joint basis. we can then compute, in these coordinates, that

$$\begin{aligned}
&\Phi^*(\omega_1 \wedge \omega_0)(X_1, \dots, X_r, Y_1, \dots, Y_m) \\
&= \omega_1 \wedge \omega_0(d\Phi(X_1), \dots, d\Phi(X_r), d\Phi(Y_1), \dots, d\Phi(Y_m)) \\
&= \det[d\Phi]\omega_1 \wedge \omega_0(X_1, \dots, X_r, Y_1, \dots, Y_m).
\end{aligned}$$

From Lemma 16.43, we can compute this determinant. Since $\mathrm{Ad}(x) \in \mathrm{SO}(\mathfrak{k})$, its determinant is 1. The determinant of a block diagonal matrix is the product of the determinants of the diagonal blocks. Thus $\det[d\Phi_{t,[x]}] = \det(\mathrm{Ad}(t^{-1})|_{\mathfrak{f}} - I) = \rho(t)$. So we have

$$\Phi^*(\omega_1 \wedge \omega_0) = \rho(t)\omega_1 \wedge \omega_0.$$

Up to the the identifications above (which are only valid up to scale), this shows that $\Phi^*(dx) = \rho(t)\, d[y] \wedge dt$, as claimed. $\square$

REMARK 16.48. It is possible to show that the scaling constant found in the above proof is 1 directly, being more careful about the tangent space identifications and working in local slice coordinates. We will not give this level of detail here; the proof that the scaling constant is 1 will be given in a different, algebraic way in the next chapter.

CHAPTER 17

Basic Representation Theory

1. Introduction and Examples

Most of the examples of Lie groups we've considered have been matrix Lie groups: subgroups of $\mathrm{GL}(n, \mathbb{C})$. We've seen only one example, the universal covering group $\widetilde{\mathrm{SL}(2, \mathbb{R})}$ that is not a matrix group, cf. Corollary 15.36. To be more precise, what we showed is that there is no Lie subgroup of $\mathrm{GL}(n, \mathbb{C})$ that is isomorphic to $\widetilde{\mathrm{SL}(2, \mathbb{R})}$.

The question of whether a given Lie group is isomorphic to a subgroup of $\mathrm{GL}(n, \mathbb{C})$ is an important one, which began the largest piece of Lie theory: **representation theory**. This applies equally well to Lie groups and Lie algebras.

DEFINITION 17.1. *Let G be a Lie group. A* **finite-dimensional complex representation** *of G is a pair (V, Π) where V is a finite-dimensional non-zero complex vector space and Π Lie group homomorphism*

$$\Pi\colon G \to \mathrm{GL}(V).$$

If V is a real vector space instead, we call Π a **finite-dimensional real representation**.

Similarly, if $\mathfrak{g}$ is a Lie algebra, a **finite-dimensional complex representation** *of $\mathfrak{g}$ is a pair (V, π) where V is a finite-dimensional non-zero complex vector space and π is a Lie algebra homomorphism*

$$\pi\colon \mathfrak{g} \to \mathfrak{gl}(V).$$

If V is a real vector space instead, we call π a **finite-dimensional real representation**.

If a representation is injective, we call it **faithful**.

The question of whether a given Lie group (or Lie algebra) *is* a matrix group (or algebra) is thus the question of whether it possesses a faithful representation. But, as we will see, understanding *all* representations of a given Lie group / Lie algebra gives much more information than this humble question.

NOTATION 17.2. In many instances, we will de-emphasize the representation homomorphism Π or π, and refer to V itself as the representation. In the case of a Lie group homomorphism, a representation $\Pi\colon G \to \mathrm{GL}(V)$ gives rise to a left action of G on V: $g \cdot v = \Pi(g)v$. So we may describe V as a G-space. It is important that this is no ordinary action: it is a *linear* action.

Let us now consider several examples of representations, beginning with the stupidest one.

EXAMPLE 17.3. The constant homomorphism $\Pi\colon G \to \mathrm{GL}(1, \mathbb{C})$ given by $\Pi(g) = I$ for all $g \in G$ is (trivially) a complex representation of the Lie group G. Similarly, $\pi\colon \mathfrak{g} \to \mathfrak{gl}(1, \mathbb{C})$ given by $\pi(X) = 0$ for all $X \in \mathfrak{g}$ is (trivially) a complex representation of the Lie algebra $\mathfrak{g}$. We call these representations the **trivial representations**.

EXAMPLE 17.4. If $G \subseteq \mathrm{GL}(n, \mathbb{C})$ is a matrix Lie group, or if $\mathfrak{g} \subseteq \mathfrak{gl}(n, \mathbb{C})$ is a matrix Lie algebra, then the inclusion maps $\Pi(g) = g$ and $\pi(X) = X$ are faithful representations. They are called the **standard representations**.

EXAMPLE 17.5. Given a Lie group G with Lie algebra $\mathfrak{g}$, the **Adjoint representation** $\mathrm{Ad}\colon G \to \mathrm{GL}(\mathfrak{g})$ is a Lie group representation (complex if $\mathfrak{g}$ happens to be a complex Lie algebra). It is typically not faithful, although it can be, for example on $SO(3)$. Similarly, $\mathrm{ad}\colon \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ (given by $\mathrm{ad}(X)Y = [X, Y]$) is a **adjoint representation** of $\mathfrak{g}$. Again, it is typically not faithful, although it can be, for example on $\mathfrak{so}(3)$.

EXAMPLE 17.6. For $m \in \mathbb{N}$, let V_m denote the space of homogeneous polynomials of degree m in two complex variables. Elements of V_m have the form

$$f(z_1, x_2) = a_0 z_1^m + a_1 z_m^{m-1} z_2 + a_2 z_1^{m-2} z_2^2 + \cdots + a_m z_2^m, \qquad a_0, a_1, \ldots, a_m \in \mathbb{C}$$

and so we see that V_m is a complex vector space of dimension $m + 1$. Now, for any matrix $A \in \mathrm{GL}(2, \mathbb{C})$ and any $f \in V_m$, we can compute that the function $(z_1, z_2) \mapsto f(A^{-1}[z_1, z_2]^\top)$ is again in V_m. Indeed, by linearity, it suffices to show this is true for $f(z_1, z_2) = z_1^k z_2^\ell$ for any $k, \ell \in \mathbb{N}$ with $k + \ell = m$. For $A \in \mathrm{GL}(2, \mathbb{C})$ with entries A_{ij}^{-1},

$$\begin{aligned}
f(A^{-1}[z_1, z_2]^\top) &= f(A_{11}^{-1} z_1 + A_{12}^{-1} z_2, A_{21}^{-1} z_1 + A_{22}^{-1} z_2) \\
&= (A_{11}^{-1} z_1 + A_{12}^{-1} z_2)^k (A_{21}^{-1} z_1 + A_{22}^{-1} z_2)^\ell \\
&= \sum_{i=0}^{k} \binom{k}{i} (A_{11}^{-1} z_1)^i (A_{12}^{-1} z_2)^{k-i} \cdot \sum_{j=0}^{\ell} (A_{21}^{-1} z_1)^j (A_{22}^{-1} z_2)^{\ell-j} \\
&= \sum_{i=0}^{k} \sum_{j=0}^{\ell} (A_{11}^{-1})^i (A_{12}^{-1})^{k-i} (A_{21}^{-1})^j (A_{22}^{-1})^{\ell-j} z_1^{i+j} z_2^{k+\ell-(i+j)}
\end{aligned}$$

This allows us to define a representation

$$\Pi_m\colon \mathrm{GL}(2, \mathbb{C}) \to \mathrm{GL}(V_m), \quad [\Pi_m(A) f](z) = f(A^{-1} z).$$

Indeed, $\Pi_m(A)$ is a linear map on f, as can be immediately checked. It is invertible, since $\Pi_m(A^{-1}) \circ \Pi_m(A) = \mathrm{Id}_{V_m}$. And it is a homomorphism:

$$[\Pi_m(AB) f](z) = f((AB)^{-1} z) = f(B^{-1} A^{-1} z) = [(\Pi_m(B) f)(A^{-1} \cdot)](z) = [\Pi_m(A) \Pi_m(B) f](z).$$

(Indeed, this is why we have $\mathrm{GL}(2, \mathbb{C})$ act by inverse on the left; if we had defined it by $f(z) \mapsto f(Az)$, it would have been an anti-homomorphism – i.e. a right action, instead of a left action.)

Note that the restriction of Π_m to any matrix Lie group $G \subseteq \mathrm{GL}(2, \mathbb{C})$ is a representation of G on V_m. We will soon look at the case $G = \mathrm{SU}(2)$.

Since any representation $\Pi\colon G \to \mathrm{GL}(V)$, of a Lie group G is a Lie group homomrphism, we may differentiate it to give the induced Lie algebra homomorphism $\pi_*\colon \mathfrak{g} \to \mathfrak{gl}(V)$. We summarize this and the usual properties of the exponential map (cf. Theorem 13.9) in the present context as follows.

PROPOSITION 17.7. *Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let (V, Π) be a finite-dimensional (real or complex) representation of G. Then $\pi = \Pi_*\colon \mathfrak{g} \to \mathfrak{gl}(V)$ is the unique representation of $\mathfrak{g}$ satisfying*

$$\Pi(\exp X) = e^{\pi(X)}, \quad \forall\, X \in \mathfrak{g}. \tag{17.1}$$

The representation π can be computed as

$$\pi(X) = \frac{d}{dt}\bigg|_{t=0} \Pi(\exp tX). \tag{17.2}$$

It satisfies

$$\pi(\mathrm{Ad}(g)X) = \mathrm{Ad}(\Pi(g))\pi(X) = \Pi(g)\pi(X)\Pi(g)^{-1}, \qquad \forall\, g \in G, X \in \mathfrak{g}. \tag{17.3}$$

PROOF. Equation 17.2 is just the definition of $\pi = \Pi_* = d\Pi_e$, and (17.1) is Theorem 13.9(g) applied to the homomorphism Π. For (17.3), we note that, for $t \in \mathbb{R}$, $g \in G$, and $X \in \mathfrak{g}$,

$$e^{t\pi(\mathrm{Ad}(g)X)} = e^{\pi(\mathrm{Ad}(g)tX)} = \Pi(\exp \mathrm{Ad}(g)tX)$$

by (17.1). But $\mathrm{Ad}(g) = (C_g)_*$, and so again applying Theorem 13.9(g) to the homomorphism C_g we have $\exp \mathrm{Ad}(g)tX = C_g(\exp tX)$. Since Π is a homomorphism, we therefore have

$$e^{t\pi(\mathrm{Ad}(g)X)} = \Pi(g \exp(tX)g^{-1}) = \Pi(g)\Pi(\exp tX)\Pi(g^{-1}) = \Pi(g)e^{t\pi(X)}\Pi(g)^{-1}$$

again using (17.1). Differentiating both sides at $t = 0$ yields (17.3). □

Of course, not every Lie algebra homomorphism arises this way in general (cf. the Lie Correspondence), but in the special case that G is simply connected, there is a one-to-one correspondence between representations of G and representations of g.

In Examples 17.3, 17.4, and 17.5, the two representations given (of a Lie group and its Lie algebra) are related as in Proposition 17.7. Let us now consider the induced Lie algebra represention from the Lie group representation in Example 17.6.

EXAMPLE 17.8. Let (V_m, Π_m) be the representation in Example 17.6, restricted to the group $\mathrm{SU}(2)$. Let us calculate the induced representation $\pi_m \colon \mathrm{su}(2) \to \mathrm{gl}(V_m)$. It is given by

$$[\pi_m(X)f](z) = \frac{d}{dt}\bigg|_{t=0} f(e^{-tX}z) = -\frac{\partial f}{\partial z_1} \cdot (X_{11}z_1 + X_{12}z_2) - \frac{\partial f}{\partial z_2} \cdot (X_{21}z_1 + X_{22}z_2). \tag{17.4}$$

The (real) Lie algebra $\mathfrak{su}(2)$ is spanned by the three vectors

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad C = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},$$

where $[A, B] = 2C$, $[B, C] = 2A$, and $[C, A] = 2B$. Then we have

$$\begin{aligned} \pi_m(A) &= -z_2\frac{\partial}{\partial z_1} + z_1\frac{\partial}{\partial z_2} \\ \pi_m(B) &= -iz_2\frac{\partial}{\partial z_1} - iz_1\frac{\partial}{\partial z_2} \\ \pi_m(C) &= -iz_1\frac{\partial}{\partial z_1} + iz_2\frac{\partial}{\partial z_2}. \end{aligned}$$

The interested reader can check that these satisfy the same commutation relations as A, B, C, and so π_m is indeed a Lie algebra homomorphism. Applying these operators to the basis elements $z_1^k z_2^\ell$ ($k + \ell = m$) of V_m, we have

$$\begin{aligned} \pi_m(A)(z_1^k z_2^\ell) &= -kz_1^{k-1}z_2^{\ell+1} + \ell z_1^{k+1}z_2^{\ell-1} \\ \pi_m(B)(z_1^k z_2^\ell) &= -ikz_1^{k-1}z_2^{\ell+1} - i\ell z_1^{k+1}z_2^{\ell-1} \\ \pi_m(C)(z_1^k z_2^\ell) &= i(\ell - k)z_1^k z_2^\ell \end{aligned}$$

with the understanding that when an exponent drops below 0 or rises above m the result is the 0 vector. So the given basis of V_m consists of eigenvectors for $\pi_m(C)$, while the others act as combinations of raising- and lowering-operators on the two factors.

In Example 17.8, since Π_m and π_m are complex representations of a real Lie group and Lie algebra. At least in the case of the Lie algebra, it is straightforward to complexify, as we did in the proof of Lemma 15.33. Let us formalize this.

DEFINITION 17.9. *Let $\mathfrak{g}$ be a real Lie algebra. Its* **complexification** $\mathfrak{g}_{\mathbb{C}}$ *is the complex Lie algebra $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} + i\mathfrak{g}$, where the bracket is given by distributing:*

$$[X + iY, X' + iY'] = [X, X'] - [Y, Y'] + i([X, Y'] + [X', Y]).$$

It is easy to check that $\mathfrak{g}_{\mathbb{C}}$ is a Lie algebra. If $\mathfrak{g}$ was already the "real version" of a complex Lie algebra (for example $\mathfrak{sl}(2, \mathbb{R})$ as opposed to $\mathfrak{sl}(2, \mathbb{C})$), its complexification is the "complex version", but there are many examples where the complexification is quite different.

EXAMPLE 17.10. In Lemma 15.33, we showed that $\mathfrak{sl}(2, \mathbb{R})_{\mathbb{C}} = \mathfrak{sl}(2, \mathbb{C})$; in fact, the same argument shows that $\mathfrak{sl}(n, \mathbb{R})_{\mathbb{C}} = \mathfrak{sl}(n, \mathbb{C})$. More interestingly, consider $\mathfrak{su}(n)_{\mathbb{C}}$. This is the space of complex matrices in $\mathfrak{gl}(n, \mathbb{C})$ of the form $X + iY$ where $X, Y \in \mathfrak{su}(n)$. Both X and Y have trace 0, and therefore so does $X + iY$; but subject to this constrant, the diagonal entries can be any complex numbers now. Moreover, the off-diagonal entries are now completely unconstrained: since Y is skew Hermitian, iY is Hermitian, and since any matrix is a sum of a Hermitian and a skew-Hermitian part, it follows that

$$\mathfrak{su}(n)_{\mathbb{C}} = \{Z \in \mathfrak{gl}(n, \mathbb{C}) \colon \operatorname{Tr}(Z) = 0\} = \mathfrak{sl}(n, \mathbb{C}).$$

Thus, the two very different real Lie algebras $\mathfrak{sl}(n, \mathbb{R})$ and $\mathfrak{su}(n)$ have the same complexification $\mathfrak{sl}(n, \mathbb{C})$.

The following lemma was essentially proved in Lemma 15.33; the proof is immediate.

LEMMA 17.11. *Let $\mathfrak{g}$ be a (real) Lie algebra, with complexification $\mathfrak{g}_{\mathbb{C}}$. If (V, π) is a complex representation of $\mathfrak{g}$, it has a unique extension to a complex representation $(V, \pi_{\mathbb{C}})$ of $\mathfrak{g}_{\mathbb{C}}$ (on the same space). The complexification is given by*

$$\pi_{\mathbb{C}}(X + iY) = \pi(X) + i\pi(Y).$$

EXAMPLE 17.12. Consider again Example 17.8. The representation π_m of $\mathfrak{su}(2)$ given there extends uniquely to a representation $(\pi_m)_{\mathbb{C}}$ of the complex Lie algebra $\mathfrak{su}(2)_{\mathbb{C}} = \mathfrak{sl}(2, \mathbb{C})$, cf. Example 17.10. By linearity, it is immediate that $(\pi_m)_{\mathbb{C}}$ is also given by the right-hand-side of (17.4).

The complex Lie algebra $\mathfrak{sl}(2, \mathbb{C})$ is spanned (over $\mathbb{C}$) by the three vectors

$$X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

where $[X, Y] = H$, $[Y, H] = 2Y$, and $[X, H] = -2X$. Now we have

$$\begin{aligned} (\pi_m)_{\mathbb{C}}(X) &= -z_2 \frac{\partial}{\partial z_1} \\ (\pi_m)_{\mathbb{C}}(Y) &= -z_1 \frac{\partial}{\partial z_2} \\ (\pi_m)_{\mathbb{C}}(H) &= -z_1 \frac{\partial}{\partial z_1} + z_2 \frac{\partial}{\partial z_2}. \end{aligned}$$

Acting on the basis elements $z_1^k z_2^\ell$ ($k+\ell=m$) of V_m, we have

$$(\pi_m)_{\mathbb{C}}(X)(z_1^k z_2^\ell) = -k z_1^{k-1} z_2^{\ell+1}$$
$$(\pi_m)_{\mathbb{C}}(Y)(z_1^k z_2^\ell) = -\ell z_1^{k+1} z_2^{\ell-1}$$
$$(\pi_m)_{\mathbb{C}}(H)(z_1^k z_2^\ell) = (\ell-k) z_1^k z_2^\ell$$

with the understanding that when an exponent drops below 0 or rises above m the result is the 0 vector. Again, we see that the standard basis vectors are eigenvectors for $(\pi_m)_{\mathbb{C}}(H)$, while $(\pi_m)_{\mathbb{C}}(X)$ and $(\pi_m)_{\mathbb{C}}(Y)$ are (somewhat less complicated) shift operators, exchanging exponents between the two variables. In particular, note that $(\pi_m)_{\mathbb{C}}(X)$ *increases* the eigenvalue of $(\pi_m)_{\mathbb{C}}(H)$ by 2, while $(\pi_m)_{\mathbb{C}}(Y)$ *decreases* the eigenvalue of $(\pi_m)_{\mathbb{C}}(H)$ by 2.

Following is the notion of isomorphism for Lie Group representations.

DEFINITION 17.13. *Let G be a matrix Lie group with two representations (V,Π) and (W,Θ). An* **intertwining map** *for the two representations is a linear map $\phi\colon V\to W$ with the property that*

$$\phi(\Pi(g)v) = \Theta(g)\phi(v), \qquad \forall\, g\in G, v\in V.$$

If ϕ is additionally a linear isomorphism, we call it an **isomorphism of representations***, and we call the two representations* **isomorphic**.

The precisely analogous notion and language is used for Lie algebra representations.

Viewing the two representations as (linear) actions of G, the linear map $\phi\colon V\to W$ is an intertwining map iff $\phi(g\cdot v) = g\cdot\phi(v)$; in the language of group actions on manifolds (cf. Definition 11.18), ϕ is *equivariant*. (We also used the term *intertwines* in that context.)

It is worth noting that the presence of (multiple) isomorphisms between representations depends on whether they are real or complex.

EXAMPLE 17.14. Consider the following two *real* representations of $\mathbb{C}^* = \mathrm{GL}(1,\mathbb{C})$, both on the vector space $V=\mathbb{C}$:

$$\Pi(z)w = zw, \qquad \Theta(z)w = \overline{z}w.$$

Both are easily seen to be representations $\mathbb{C}^*\to\mathrm{GL}(V)$. Π is the standard representation of $\mathrm{GL}(1,\mathbb{C})$. These two representations are isomorphic: the map $\phi\colon V\to V$ given by $\phi(v)=\overline{v}$ is a(n $\mathbb{R}$-)linear isomorphism, and $\phi(\Pi(z)w) = \phi(zw) = \overline{zw} = \Theta(z)\phi(w)$.

Note, however, that although Π and Θ are also both complex representations, ϕ is not a $\mathbb{C}$-linear map. In fact, we will soon see (as a consequence of Schur's lemma) that there are no other complex representations of $\mathbb{C}^*$ isomorphic to Π (other than Π itself).

EXAMPLE 17.15. The standard representation and the Adjoint represention of $\mathrm{SO}(3)$ are isomorphic; this is a homework exercise.

The notions of isomorphism for Lie group representations and Lie algebra representations are equivalent (for connected groups), in the following sense.

LEMMA 17.16. *Let G be a connected Lie group, with two representations (V,Π) and (W,Θ). Let (V,π), and (W,ϑ) be the induced Lie algebra homomorphisms. Then (V,π) and (W,ϑ) are isomorphic iff (V,Π) and (W,Θ) are isomorphic.*

PROOF. Suppose first that (V,Π) and (W,Θ) are isomorphic, via intertwining isomorphicm $\phi\colon V\to W$. Then we have, for each $v\in V$, $X\in\mathfrak{g}$, and $t\in\mathbb{R}$,

$$\phi(\Pi(\exp tX)v) = \Theta(\exp tX)\phi(v).$$

Differentiating both sides at $t = 0$ shows that $\phi(\pi(X)v) = \vartheta(X)\phi(v)$, showing that ϕ is also an intertwining isomorphism for the Lie algebra representations.

Conversely, suppose $\phi\colon V \to W$ is a Lie algebra isomorphism. Since G is connected, every element $g \in G$ has the form $g = \exp X_1 \exp X_2 \cdots \exp X_n$ for some $X_1, \ldots, X_n \in \mathfrak{g}$. We make use of the Baker-Campbell-Hausdorff formula, so (by writing each term $\exp X = \exp(\frac{1}{k}X)^k$ for large enough k if needed) we assume that $X_1, \ldots, X_n$ are all in a sufficiently small neighborhood of 0 so that the BCH formula applies jointly to both $X_1, \ldots, X_n$ and to $\pi(X_1), \ldots, \pi(X_n)$. Now, for $v \in V$,

$$\phi(\Pi(g)v) = \phi(\Pi(\exp X_1 \cdots \exp X_n)v) = \phi(\Pi(\exp X_1) \cdots \Pi(\exp X_n)v) = \phi(e^{\pi(X_1)} \cdots e^{\pi(X_n)}v).$$

By the BCH formula there is a functions C in n variables, built out of brackets, so that

$$e^{\pi(X_1)} \cdots e^{\pi(X_n)} = e^{C(\pi(X_1),\ldots,\pi(X_n))}.$$

Because π is a Lie algebra homomorphism, we also have

$$C(\pi(X_1), \ldots, \pi(X_n)) = \pi(C(X_1, \ldots, X_n)).$$

Thus

$$\begin{aligned}\phi(\Pi(g)v) = \phi(e^{\pi(C(X_1,\ldots,X_n))}v) &= \phi(\Pi(\exp C(X_1, \ldots, X_n))v) \\ &= \Theta(\exp C(X_1, \ldots, X_n))\phi(v).\end{aligned}$$

Reversing the preceding calculation shows that $\exp C(X_1, \ldots, X_n) = \exp X_1 \cdots \exp X_n = g$, and this concludes the proof that ϕ intertwines the Lie group representations. □

REMARK 17.17. It is possible to prove the preceding lemma without the BCH formula, by expanding the matrix exponential as a power series and setting up and inductions argument; despite the fact that it uses "heavy machinery", the above proof is much cleaner.

2. Irreducible and Completely Reducible Representations

As with most mathematical structures, there are basis "indecomposable" objects that all others (ideally) can be built from in representation theory. They are called **irreducible representation**.

DEFINITION 17.18. *Let (V, Π) be a finite-dimensional real or complex representation of a Lie group G. A subspace $W \subseteq V$ is called* **invariant** *for the representation if $\Pi(g)w \in W$ for all $w \in W$ and $g \in G$. Clearly $\{0\}$ and V are invariant subspaces; all others are called* **nontrivial** *invariant subspaces. The representation is* **irreducible** *if it has no nontrivial invariant subspaces. For short, irreducible representations are sometimes called* **irreps**.

The precsiely analogous notion and language is used for Lie algebra representations.

REMARK 17.19. If (V, π) is a complex representation of $\mathfrak{g}$, an invariant subspace is requires to be a complex subspace. It is entirely possible for a complex representation to be irreducible, but to have invariant real subspaces when viewed as a real representation.

EXAMPLE 17.20. Any 1-dimensional representation is trivially irreducible; so the trivial representation of any Lie group / Lie algebra is irreducible.

EXAMPLE 17.21. The standard representation of $\mathrm{SO}(n)$ is irreducible. Indeed, let $W \subseteq \mathbb{R}^n$ be an invariant subspace. For any $w \in W$, the orbit $\mathrm{SO(n)} \cdot w$ is the sphere of radius $\|w\|$. Hence W must contain all vectors of length $\|w\|$, and so if $w \neq 0$, since W is a subspace, $W = \mathbb{R}^n$.

On the other hand, the standard representation of $\mathrm{SO}(n) \times \mathrm{SO}(m)$ on $\mathbb{R}^{n+m}$ is not irreducible: the subspaces $\mathbb{R}^n \times \{0\}$ and $\{0\} \times \mathbb{R}^m$ are invariant.

Let us record as a lemma a fact that is elementary to prove and left to the reader.

LEMMA 17.22. *If two representations are isomorphic, then one is irreducible iff the other is irreducible.*

An important example of irreducible representations are those in Example 17.12.

PROPOSITION 17.23. *The representations $(V_m, (\pi_m)_{\mathbb{C}})$ of $\mathfrak{sl}(2, \mathbb{C})$ in Example 17.12 are irreducible.*

PROOF. Let $W \subseteq V_m$ be a non-zero invariant subspace, and let $w \in W$. Write it in the form

$$w = a_0 z_1^m + a_1 z_1^{m-1} z_2 + a_2 z_1^{m-2} z_2^2 + \cdots + a_m z_2^m.$$

Since $w \neq 0$, at least one of the coefficients is nonzero, so let k_0 be the smallest index of a nonzero coefficient, meaning that all nonzero terms have $z_1^{m-k} z_2^k$ with $k \geq k_0$. Now, repeatedly apply $(\pi_m)_{\mathbb{C}}(X)$ to w. Since this operator has the effect of increasing the exponent of z_2, if we apply it $m - k_0$ times, we kill all the terms except $a_{k_0} z_1^{m-k_0} z_2^{k_0}$, so we have

$$(\pi_m)_{\mathbb{C}}(X)^{m-k_0} w = a_{k_0}(m - k_0)! z_2^m.$$

Now we apply the lowering operator $(\pi_m)_{\mathbb{C}}(Y)$ repeatedly: for $0 \leq j \leq m$, we have

$$(\pi_m)_{\mathbb{C}}(Y)^j z_2^m = m(m-1)\cdots(m-j+1) z_1^j z_2^{m-j}.$$

Thus, for each j, there is a non-zero constant $c_j = a_{k_0}(m - k_0)! m(m-1)\cdots(m-j+1)$ such that

$$(\pi_m)_{\mathbb{C}}(Y)^j (\pi_m)_{\mathbb{C}}(X)^{m-k_0} w = c_j z_1^j z_2^{m-j}.$$

Since W is an invariant subspace, and $w \in W$, it follows that $c_j z_1^j z_2^{m-j} \in W$ for each j. The span of these vectors is all of V_m, so it follows that $W = V_m$. □

We will see later that, up to isomorphism, the representations $(V_m, (\pi_m)_{\mathbb{C}})$ are the only irreducible complex representations of $\mathfrak{sl}(2, \mathbb{C})$.

From Proposition 17.23, we can actually conclude that the corresponding real representations (V_m, π_m) of $\mathfrak{su}(2)$ and (V_m, Π_m) of $\mathrm{SU}(2)$ are irreducible, due to the following lemmas.

LEMMA 17.24. *Let $\mathfrak{g}$ be a Lie algebra with complexification $\mathfrak{g}_{\mathbb{C}}$, and let (V, π) be a complex representation of $\mathfrak{g}$ with complexification $(V, \pi_{\mathbb{C}})$. Then (V, π) is an irreducible representation of $\mathfrak{g}$ iff $(V, \pi_{\mathbb{C}})$ is an irreducible representation of $\mathfrak{g}_{\mathbb{C}}$.*

PROOF. Suppose $(V, \pi_{\mathbb{C}})$ is irreducible. If $W \subseteq V$ is an invariant subspace for π, then for any $X, Y \in \mathfrak{g}$ and $w \in W$, $\pi_{\mathbb{C}}(X + iY)w = \pi(X)w + i\pi(Y)w \in W + iW = W$ since W. Thus W is invariant for $\pi_{\mathbb{C}}$, and so by assumption $W = V$ or $W = 0$. Hence, (V, π) is irreducible. Conversely, suppose (V, π) is irreducible, and let $W \subseteq V$ be an invariant subspace for $\pi_{\mathbb{C}}$. Then for any $X \in \mathfrak{g}$ and $w \in W$, $W \ni \pi_{\mathbb{C}}(X + i0)w = \pi(X)w$, which shows that W is an invariant subspace for π, and so by assumption $W = V$ or $W = 0$. Hence, $(V, \pi_{\mathbb{C}})$ is irreducible. □

LEMMA 17.25. *Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. Let (V, Π) be a Lie group representation with induced Lie algebra representation (V, π). Then (V, Π) is irreducible iff (V, π) is irreducible.*

PROOF. Suppose (V, Π) is irreducible, and let $W \subseteq V$ be an invariant subspace for π. Since G is connected, any $g \in G$ can be written in the form $g = \exp X_1 \exp X_2 \cdots \exp X_n$ for some $X_1, \ldots, X_n \in \mathfrak{g}$. Then we have

$$\Pi(g) = \Pi(\exp X_1 \cdots \exp X_n) = \Pi(\exp X_1) \cdots \Pi(\exp X_n) = e^{\pi(X_1)} \cdots e^{\pi(X_n)}.$$

Now, since W is invariant under $\pi(X_j)$, it is also invariant under $\pi(X_j)^k$ for all $k \in \mathbb{N}$, and so by continuity, W is invariant under $e^{\pi(X_j)}$ as well. Thus, W is invariant under $\Pi(g)$ for all $g \in G$, which shows that W is an invariant subspace for Π. Since Π is irreducible, it follows that $W = 0$ or $W = V$. Thus, (V, π) is irreducible.

Conversely, suppose (V, π) is irreducible, and let $W \subseteq V$ be an invariant subpace for Π. For any $X \in \mathfrak{g}$ and $t \in \mathbb{R}$, W is invariant under $\Pi(\exp tX)$, and so for any $w \in W$

$$\pi(X)w = \frac{d}{dt}\bigg|_{t=0} \Pi(\exp tX)w \in W.$$

Hence W is invariant for each $\pi(X)$ for $X \in \mathfrak{g}$, and so is invariant for π. Since (V, π) is irreducible, it follows that $W = 0$ or $W = V$, and so (V, Π) is irreducible. □

Thus, Proposition 17.23 shows that the representation (V_m, π_m) of $\mathfrak{su}(2)$ (cf. Example 17.8) is irreducible, and hence the representation (V_m, Π_m) of $\mathrm{SU}(2)$ (cf. Example 17.6) is irreducible.

Now, not every representation is irreducible, of course. Naively, one would hope that every representation can be decomposed into a finite collection of irreps in some way. The notion of decomposition is direct sum.

DEFINITION 17.26. *Let (V_1, Π_1), …, (V_m, Π_m) be representations of a Lie group G. The* **direct sum** *representation $(V_1 \oplus \cdots \oplus V_m, \Pi_1 \oplus \cdots \Pi_m)$ is defined by*

$$[\Pi_1 \oplus \cdots \oplus P_m(g)](v_1, \ldots, v_m) = (\Pi_1(g)v_1, \ldots, \Pi_m(g)v_m), \qquad g \in G.$$

Similar notation and definitions apply to direct sums of representations of Lie algebras.

It is straightforward to verify that the direct sum really is a represention. Note that it is never irreducible (unless all factors but one are the 0 vector space), since $0 \oplus \cdots \oplus 0 \oplus V_j \oplus 0 \oplus \cdots \oplus 0$ are all nontrivial invariant subspaces.

DEFINITION 17.27. *A finite-dimensional representation of a Lie group or Lie algebra is called* **completely reducible** *if it is isomorphic to a (finite) direct sum of irreducible representations. A Lie group or Lie algebra is said to have the* **complete reducibility property** *if every finite-dimensional representation is completely reducible.*

If (V, Π) is completely reducible, then fix an isomorphism $\phi\colon V \to W_1 \oplus \cdots \oplus W_m$ and irrep $\Theta_1, \ldots, \Theta_m$ such that ϕ intertwines Π and $\Theta_1 \oplus \cdots \oplus \Theta_m$. Since the subspaces $\{0\} \oplus \cdots \oplus \{0\} \oplus W_j \oplus \{0\} \oplus \cdots \oplus \{0\}$ are invariant for $\Theta_1 \oplus \cdots \oplus \Theta_m$, it follows that the their preimages V_j under ϕ are invariant subspaces for Π in V. Hence, if (V, Π) is completely reducible, then there are invariant subspaces $V_1, \ldots, V_m \subseteq V$ for Π such that $V_1 \oplus \cdots \oplus V_m = V$; moreover, note that $\Pi|_{V_j}$ is irreducible (since the isomorphism ϕ intertwines it with the irreducible representation (W_j, Θ_j)). Conversely, if V can be decomposed as a direct sum of invariant subspaces for Π such that the restriction of Π to each subspace is irreducible, then one can decompose $\Pi = \Pi|_{V_1} \oplus \cdots \oplus \Pi|_{V_m}$

(up to the isomorphism between internal and external direct sum) to see that (V, Π) is completely reducible. That is:

LEMMA 17.28. *A representation (V, Π) is completely reducible iff there are invariant subspaces $V_1, \dots, V_m \subseteq V$ so that $V = V_1 \oplus \cdots \oplus V_m$ such that $(V_j, \Pi|_{V_j})$ is irreducible for each j.*

The naive hope expressed above would be that every representation is completely reducible (meaning every Lie group or Lie algebra has the complete reducibility property). As it turns out, this is very false.

EXAMPLE 17.29. Let $\Pi\colon \mathbb{R} \to \mathrm{GL}(2, \mathbb{C})$ be the map

$$\Pi(x) = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}.$$

It is easy to check that $(\mathbb{C}^2, \Pi)$ is a representation of $\mathbb{R}$. It is not irreducible: letting $\{e_1, e_2\}$ denote the standard basis of $\mathbb{C}^2$, the subspace $\mathbb{C}e_1$ is invariant for Π, since $\Pi(x)e_1 = e_1$. However, $\mathbb{C}e_1$ is the *only* nontrivial invariant subspace. Indeed, suppose W is a nonzero invariant subspace, containing some vector $w = ae_1 + be_2$ not in $\mathbb{C}e_1$ (so $b \neq 0$). Then

$$\Pi(1)w - w = (\Pi(1) - I)v = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = be_1$$

and since W is invariant and a subspace, $\Pi(1)w - w \in W$. Hence $e_1 \in W$, and so $W \supseteq \mathrm{span}\{e_1, w\} = \mathbb{C}^2$.

It then follows that $(\mathbb{C}^2, \Pi)$ is not completely reducible: if it were, Lemma 17.28 shows that $\mathbb{C}^2$ could be decomposed as a direct sum of invariant subspaces for Π; but the unique invariant subspace for Π is $\mathbb{C}e_1$, which does not span $\mathbb{C}^2$.

In some sense, *most* representations are not completely reducible. One notably exceptional class are *unitary representations*.

DEFINITION 17.30. *Let V be a finite-dimensional Hilbert space. A representation (V, Π) of a Lie group G is called* **unitary** *of $\Pi(g)$ is a unitary operator on V for each $g \in G$. (If the representation is real, we interpret "unitary" to mean "orthogonal" here.)*

Similarly, if $\mathfrak{g}$ is a Lie algebra, then a representation (V, π) is called **unitary** *(or* **skew-Hermitian***) if $\pi(X)^* = -\pi(X)$ for all $X \in \mathfrak{g}$.*

As usual, the notions of unitarity for Lie group vs. Lie algebra representations are equivalent, modulo connectivity of the group.

LEMMA 17.31. *Let (V, Π) be a representation of a connected Lie group, with induced Lie algebra representation (V, π). Then (V, Π) is unitary iff (V, π) is unitary.*

PROOF. Let G denote the Lie group, and $\mathfrak{g}$ its Lie algebra. If (V, Π) is unitary, then for all $g \in G$, $\Pi(g) \in \mathrm{U}(V)$. In particular, for $X \in \mathfrak{g}$ and $t \in \mathbb{R}$, $\Pi(\exp tX) = e^{t\pi(X)} \in \mathrm{U}(V)$. Differentiating at $t = 0$ shows that $\pi(X) \in \mathfrak{u}(V)$ which (of couse) consists of skew-Hermitian matrices. So (V, π) is unitary. Conversely, suppose (V, π) is unitary. Since G is connected, any $g \in G$ has the form $g = \exp X_1 \cdots \exp X_n$ for some $X_1, \dots, X_n \in \mathfrak{g}$. Then

$$\Pi(g) = \Pi(\exp X_1 \cdots \exp X_n) = \Pi(\exp X_1) \cdots \Pi(\exp X_n) = e^{\pi(X_1)} \cdots e^{\pi(X_n)}.$$

Since each $\pi(X_j) \in \mathfrak{u}(V)$, $e^{\pi(X_j)} \in \mathrm{U}(V)$, and as $\mathrm{U}(V)$ is a group, the product $\Pi(g) \in \mathrm{U}(V)$. So (V, Π) is unitary. $\square$

Unitary representations abound, in the sense that it is often possible to define an inner product on the representation space that makes the given representation unitary. This is always the case, for example, for representations of compact Lie groups and their Lie algebras.

LEMMA 17.32. *Let (V, Π) be a representation of a compact Lie group K. There exists an inner product on V with respect to which (V, Π) is a unitary representation.*

PROOF. To begin, fix any inner product $\langle \cdot, \cdot \rangle_0$ on V. Define a new inner product by averaging the action of Π over the Haar masure of K:

$$\langle v, w \rangle \equiv \int_K \langle \Pi(x)v, \Pi(x)w \rangle_0 \, dx.$$

This is a generalization of the same trick we used in Lemma 16.25 to construct an Ad-invariant inner product; the same proof here shows that this averaged bilinear form is an inner product that is preserved by $\Pi(g)$ for all $g \in G$; in other words, with respect to this inner product, (V, Π) is a unitary representation. □

This is very useful, because unitary representations are always completely reducible.

PROPOSITION 17.33. *Unitary representations (of Lie groups or Lie algebras) are completely reducible.*

PROOF. We write the proof here for group representations; the argument is the same for Lie algebra representations is analogous. Let (Π, V) be a unitary representation of a Lie group G on a finite-dimensional Hilbert space V; denote the inner product by $\langle \cdot, \cdot \rangle$. If the representation is irreducible, it is completely reducible vacuously. If not, there is some nontrivial subspace $W \subset V$ that is invariant. Let $W^\perp$ be the orthogonal complement in V with respect to the inner product. Then $W^\perp$ is *also* an invariant subspace. To see thus, note that for any $g \in G$ $\Pi(g^{-1}) = \Pi(g)^{-1} = \Pi(g)^*$. Thus, for any $w \in W$ and $v \in W^\perp$,

$$\langle \Pi(g)v, w \rangle = \langle v, \Pi(g)^* w \rangle = \langle v, \Pi(g^{-1})w \rangle = 0$$

because $w \in W$ is invariant for Π and so $\Pi(g^{-1})w \in W$ which is orthogonal to $W^\perp$. Thus $\Pi(g)v$ is orthogonal to W for all $g \in G$, and so is in $W^\perp$.

Now, $V = W \oplus W^\perp$ is a direct sum of invariant subspaces. We proceed by induction. Suppose we have decomposed $V = V_1 \oplus \cdots \oplus V_m$ into invariant subspaces. If all are irreducible for Π, then we have shown (V, Π) is completely reducible. Otherwise, there is at least on V_j that is not irreducible, so it has a nontrivial invariant subspace W_j. The preceding argument shows that $W_j^\perp$ is also invariant, and so $V = V_1 \oplus \cdots \oplus V_{j-1} \oplus W_j \oplus W_j^\perp \oplus V_{j+1} \cdots \oplus V_m$ is a finer decomposition of V into invariant subspaces. This process must terminate finitely: at each step of the process, the maximal dimension of any non-irreducible invariant subspace decreases, and since V is finite dimensional, we eventually decompose V into a direct sum of invariant irreducible subspaces. By Lemma 17.28, it follows that (Π, V) is completely reducible. □

COROLLARY 17.34. *Compact Lie groups have the complete reducibility property.*

PROOF. Let (Π, V) be a representation of a compact Lie group. By Lemma 17.32, there is an inner product on V with respect to which Π is unitary. Then by Proposition 17.33, (Π, V) is completely reducible. □

COROLLARY 17.35. *If G is a Lie group with Lie algebra $\mathfrak{g}$, and if there is a simply-connected compact Lie group K with Lie algebra $\mathfrak{k}$ so that $\mathfrak{g} \cong \mathfrak{k}_{\mathbb{C}}$, then G and $\mathfrak{g}$ have the complete reducibility property.*

Note: we have not formally defined complex Lie algebras, but in this context we simply mean that the Lie algebra of G happens to be a complex Lie algebra.

PROOF. Let (V,Π) be a complex representation of G, and let (V,π) be the induced Lie algebra representation of $\mathfrak{g}$. Since $\mathfrak{g} \cong \mathfrak{k}_{\mathbb{C}} = \mathfrak{k} + i\mathfrak{k}$, we may view $\mathfrak{k}$ as a Lie subalgebra, and so $\phi = \pi|_{\mathfrak{k}}$ is a Lie algebra representation of $\mathfrak{k}$ on V. Since K is simply-connected, by the Lie correspondence (Theorem 14.35), there is a unique representation $\Phi\colon K \to \mathrm{GL}(V)$ with $\Phi_* = \phi$. By Corollary 17.34, Φ is completely reducible, so $V = V_1 \oplus \cdots \oplus V_m$ is a direct sum of invariant subspaces each of which is irreducible for Φ. It follows from Lemma 17.25 that the subspaces $V_1,\ldots,V_m$ are invariant for ϕ and $\phi|_{V_j}$ is irreducible, and then it follows from Lemma 17.24 that $\pi = \phi_{\mathbb{C}}$ has the same properties. Thus, (V,π) is completely reducible, and one more application of Lemma 17.25 shows that (V,Π) is completely reducible. $\square$

EXAMPLE 17.36. The Lie group $\mathrm{SU}(2)$ is compact and simply-connected (as it is diffeomorphic to $\mathbb{S}^3$). Note that $\mathfrak{sl}(2,\mathbb{C}) = \mathfrak{su}(2)_{\mathbb{C}}$. Hence, by Corollary 17.35, $\mathfrak{sl}(2,\mathbb{C})$ and $\mathrm{SL}(2,\mathbb{C})$ have the complete reducibility property.

From Example 17.36, we see that to understand *all* representations of $\mathfrak{sl}(2,\mathbb{C})$, it suffices to understand the irreps (since all representations are then isomorphic to direct sums of these). We will classify all such irreps. First, we need the most important tool in representation theory: *Schur's Lemma*, which is the topic of the next section.

3. The Representation Theory of $\mathfrak{sl}(2,\mathbb{C})$

In this section, we will prove that the irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$ in Example 17.12 and Proposition 17.23 are, up to isomorphism, *all* of the irreps. Since $\mathrm{SL}(2,\mathbb{C})$ has the complete reducibility property (cf. Example 17.36), this gives us a complete picture of all representations of this group and its Lie algebra.

Before proceeding with the proof, a few words on why we pay this special case special attention. First, since $\mathfrak{sl}(2,\mathbb{C}) = \mathfrak{su}(2)_{\mathbb{C}}$, and $\mathfrak{su}(2) \cong \mathfrak{so}(3)$, this will also give us a good understanding of the representation theory of $\mathrm{SO}(3)$, which is a group of great physical significance. In short: if we are trying to study the solutions of a (linear) PDE, if that PDE happens to be rotationally invariant, then its solution space forms a representation of $\mathrm{SO}(3)$, and therefore also of $\mathfrak{so}(3)$. Having a complete understanding of all representations of $\mathfrak{so}(3)$ then gives us tools to understand the solutions of the PDE. This applies in particular to the Schrödinger equation for the Hydrogen atom, which is $\mathrm{SO}(3)$-invariant. In fact, the computations we will do in this section are done in every standard quantum mechanics textbookm under the heading "angular momentum".

Second, and more important for us: in understanding the representation theory of other (semisimple) Lie algebras, we will make frequent use of the representations of $\mathfrak{sl}(2,\mathbb{C})$ which appears as a common subalgebra. In particular, the approach we use presently (via commutation relations) to determine the representations is illuminating of how we will approach the problem in general.

THEOREM 17.37. *For each integer $m \geq 0$, there is (up to isomorphism) precisely one complex irrep of $\mathfrak{sl}(2,\mathbb{C})$ of dimension $m+1$; it is isomorphic to the representation $(V_m,(\pi_m)_{\mathbb{C}})$ of Example 17.12.*

Since two representations of different dimensions are trivially non-isomorphic, Theorem 17.37 completely classifies all irreps of $\mathfrak{sl}(2,\mathbb{C})$.

Before proceeding with the proof, we remind ourselves of the notation used in Example 17.12. The Lie algebra is spanned by

$$X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and these three matrices satisfy the commutation relations

$$[X, Y] = H, \qquad [Y, H] = 2Y, \qquad [X, H] = -2X. \tag{17.5}$$

Note, then, that if V is any vector space, and $A, B, C \in \mathfrak{gl}(V)$ are three linear operators that happen to satisfy

$$[A, B] = C, \qquad [B, C] = 2B, \qquad [A, C] = -2A$$

then the linear map $\pi\colon \mathfrak{sl}(2, \mathbb{C}) \to \mathfrak{gl}(V)$ defined by $\pi(X) = A$, $\pi(Y) = B$, and $\pi(H) = C$ will be a Lie algebra representation (thanks to the bilinearity and skew-symmetry of brackets). Conversely, the images of $\pi(X)$, $\pi(Y)$, and $\pi(H)$ in any representation must satisfy the same relations. (They may satisfy additional representations, of course.) As a result, we have the following Lemma which shows that the same "raising" and "lowering" properties shown to be satisfied by $(\pi_m)_{\mathbb{C}}(X)$ and $(\pi_m)_{\mathbb{C}}(Y)$ hold in any representation of $\mathfrak{sl}(2, \mathbb{C})$.

LEMMA 17.38. *Let (V, π) be a representation of $\mathfrak{sl}(2, \mathbb{C})$. Let u be an eigenvector of $\pi(H)$, with eigenvalue $\alpha \in \mathbb{C}$. Then*

$$\pi(H)\pi(X)u = (\alpha + 2)\pi(X)u, \quad \textit{and} \quad \pi(H)\pi(Y)u = (\alpha - 2)\pi(Y)u.$$

Thus, $\pi(X)u$ and $\pi(Y)u$ are either 0 vectors or are eigenvectors of $\pi(H)$ with eigenvalues $\alpha \pm 2$.

PROOF. We deal with the case $\pi(X)$; $\pi(Y)$ is analogous. Since $[\pi(X), \pi(H)] = \pi([X, H]) = \pi(-2X) = -2\pi(X)$, we have

$$\pi(H)\pi(X)u = \pi(X)\pi(H)u - [\pi(X), \pi(H)]u = \pi(X)\alpha u - (-2\pi(X))u = (\alpha + 2)\pi(X).$$

□

With this in hand, we now prove Theorem 17.37.

PROOF OF THEOREM 17.37. Let (V, π) be a complex irrep of $\mathfrak{sl}(2, \mathbb{C})$. Our strategy is to diagonalize the operator $\pi(H)$. Since the representation is complex, $\pi(H)$ has an eigenvalue α; let $u \in V$ be an eigenvector with this eigenvalue. Applying Lemma 17.38 repeatedly, we find that

$$\pi(H)\pi(X)^k u = (\alpha + 2k)\pi(X)^k u, \qquad k \in \mathbb{N}.$$

This says that, if $\pi(X)^k u \neq 0$, it is an eigenvector of $\pi(H)$ with eigenvalue $\alpha + 2k$. Only finitely many of these can therefore be nonzero, since $\pi(H)$ can have only finitely-many distinct eigenvalues. Thus, there is some $n \geq 0$ such that $u_0 \equiv \pi(X)^n u \neq 0$ but $\pi(X)^{n+1}u = 0$. Let $\lambda_0 = \alpha + 2n$. Thus $u_0 \neq 0$, and

$$\pi(H)u_0 = \lambda_0 u_0, \qquad \pi(X)u_0 = 0.$$

Now, for $k \in \mathbb{N}$, define

$$u_k = \pi(Y)^k u_0.$$

Then by Lemma 17.38, we have

$$\pi(H)u_k = (\lambda_0 - 2k)u_k. \tag{17.6}$$

As above, since $\pi(H)$ has only finitely-many distinct eigenvalues, it follows that only finitely-many of the u_k are nonzero, and hence there is some $m \geq 0$ such that $u_m \neq 0$ but $u_{m+1} = 0$. Since

$u_k = \pi(Y)^k u_0$, if $u_k = 0$ then $u_{k+1} = \pi(Y)u_k = 0$ as well, and this we see that $u_0, u_1, \ldots, u_m$ are all nonzero, while $u_k = 0$ for $k \geq m+1$.

We wish to compute $\pi(X)u_k$ for each k. For $k = 0$, we have $\pi(X)u_0 = \pi(X)^{N+1}u = 0$. For $k = 1$,

$$\begin{aligned}\pi(X)u_1 = \pi(X)\pi(Y)u_0 &= (\pi(Y)\pi(X) + [\pi(X), \pi(Y)])\, u_0 \\ &= \pi(Y)\pi(X)u_0 + \pi([X,Y])u_0 \\ &= 0 + \pi(H)u_0 = \lambda_0 u_0. \end{aligned} \tag{17.7}$$

Similarly, we have

$$\begin{aligned}\pi(X)u_2 = \pi(X)\pi(Y)u_1 &= (\pi(Y)\pi(X) + [\pi(X), \pi(Y)])\, u_1 \\ &= \pi(Y)\pi(X)u_1 + \pi(H)u_1.\end{aligned}$$

Applying (17.7) and (17.6), we have

$$\pi(X)u_2 = \pi(Y)(\lambda_0 u_0) + (\lambda_0 - 2)u_1 = \lambda_0 u_1 + (\lambda_0 - 2)u_1 = 2(\lambda_0 - 1)u_1.$$

Proceeding by induction, we can quickly show that

$$\pi(X)u_k = k(\lambda_0 - (k-1))u_{k-1}, \qquad k \in \mathbb{N}. \tag{17.8}$$

Applying this with $k = m+1$, we have

$$0 = \pi(X)0 = \pi(X)u_{m+1} = (m+1)(\lambda_0 - m)u_m.$$

Since $u_m \neq 0$, it follows that $\lambda_0 = m$ is a non-negative integer.

Let us summarize what we have shown so far: for every complex irrep (V, π) of $\mathfrak{sl}(2,\mathbb{C})$, there is an integer $m \geq 0$ and nonzero vectors $u_0, \ldots, u_m$ such that:

(a) $\pi(H)u_k = (m - 2k)u_k$ for $0 \leq k \leq m$,
(b) $\pi(Y)u_k = u_{k+1}$ for $0 \leq k < m$ and $\pi(Y)u_m = 0$, and
(c) $\pi(X)u_k = k(m - (k-1))u_{k-1}$ for $0 < k \leq m$ and $\pi(X)u_0 = 0$.

Since the nonzero vectors $u_0, \ldots, u_m$ are eigenvectors of $\pi(H)$ with all distinct eigenvalues, they are linearly independent. What's more, $W = \operatorname{span}_{\mathbb{C}}\{u_0, \ldots, u_m\}$ is manifestly invariant under each of $\pi(X)$, $\pi(Y)$, and $\pi(H)$, and since X, Y, H generate $\mathfrak{sl}(2,\mathbb{C})$ and π is a Lie algebra homomorphism, it follows that W is invariant under $\pi(H)$ for all $H \in \mathfrak{sl}(2,\mathbb{C})$. Since π is irreducible, and $W \neq 0$, we must therefore have $W = V$. So $\dim(V) = m+1$. Points (a)–(c) above thus completely determine the representation π.

It is now a simple exercise to show that, considering the representation $(V_m, (\pi_m)_{\mathbb{C}})$ of Example 17.12 and letting $u_k = z_1^k z_2^{m-k}$, (a)–(c) are satisfied. Since we proved that this representation is irreducible, this shows that there is a unique (up to isomorphism) irrep of $\mathfrak{sl}(2,\mathbb{C})$ of dimension $m+1$, completing the proof. □

As noted above, since $\mathrm{SL}(2,\mathbb{C})$ has the complete reducibility property, every representation is isomorphic to a direct sum of irreps. Since we now have a complete understanding of all irreps, we therefore also understand all representations. The following properties of all $\mathfrak{sl}(2,\mathbb{C})$ representations follow easily.

COROLLARY 17.39. *Let (V, π) be a finite-dimensional complex representation of $\mathfrak{sl}(2,\mathbb{C})$ (not necessarily irreducible).*

(1) *Every eigenvalue of $\pi(H)$ is an integer. Furthermore, if v is an eigenvector for $\pi(H)$ with eigenvalue λ, and if $v \in \ker \pi(X)$, then $\lambda \in \mathbb{N}$ is a non-negative integer.*

(2) *If $k \in \mathbb{Z}$ is an eigenvalue for $\pi(H)$, then so are $-|k|, -|k|+2, \ldots, |k|-2, |k|$.*
(3) *The matrices $\pi(X)$ and $\pi(Y)$ are nilpotent – there are positive integers k, ℓ with $\pi(X)^k = \pi(Y)^\ell = 0$.*
(4) *Let $S \in \mathrm{End}(V)$ be the matrix $S = e^{\pi(X)}e^{-\pi(Y)}e^{\pi(X)}$. Then $S\pi(H)S^{-1} = -\pi(H)$.*

PROOF. First note that all four properties are invariant under isomorphism, as is trivial to check. So we may assume that the representation (V, π) actually is a direct sum of irreps $(V_j, \pi_j)_{1\le j\le r}$.

(1) If $v \in V$ is an eigenvector of $\pi(H)$ with eigenvalue λ, then $v = (v_1, \ldots, v_r)$ and
$$(\lambda v_1, \ldots, \lambda v_r) = \lambda v = \pi(H)v = (\pi_1(H)v_1, \ldots, \pi_k(H)v_r)$$
showing that λ is an eigenvalue of each $\pi_j(H)$. In the proof of Theorem 17.37, we showed that there is a basis $\{u_0, \ldots, u_m\}$ for V_j in which $\pi_j(H)$ is diagonal, with $\pi_j(H)u_k = (m-2k)u_k$; thus the eigenvalues of $\pi(H)$ are all integers, and so $\lambda \in \mathbb{Z}$. If $\pi(X)v = 0$ then $\pi_j(X)v_j = 0$ for all j, and we showed that $\pi_j(X)u_k = k(m-(k-1))u_{k-1}$ for $0 < k \le m$ while $\pi_j(X)u_0 = 0$; this shows that $\ker \pi_j(X) = \mathbb{C}u_0$, and since u_0 is an eigenvector of $\pi(H)$ with eigenvalue $m \ge 0$, it follows that $\lambda \ge 0$ in this case.
(2) If k is an eigenvalue of $\pi(H)$, then it is an eigenvalue for $\pi_j(H)$ as in part (1). Since the set of eigenvalues of $\pi_j(H)$ has the form $-m, -m+2, \ldots, m-2, m$ for some integer m, the result follows immediately.
(3) It follows immediately from (b) and (c) in the proof of Theorem 17.37 that, with $m+1 = \dim(V_j)$, $\pi_j(X)^{m+1} = \pi_j(X)^{m+1} = 0$. A finite direct sum of nilpotent operators is nilpotent, hence $\pi(X)$ and $\pi(Y)$ are nilpotent.
(4) In this case, it is simpler not to use the direct sum approach, but rather argue directly. Note that
$$\begin{aligned} S\pi(H)S^{-1} &= e^{\pi(X)}e^{-\pi(Y)}e^{\pi(X)}\pi(H)e^{-\pi(X)}e^{\pi(Y)}e^{-\pi(X)} \\ &= \mathrm{Ad}(e^{\pi(X)})\mathrm{Ad}(e^{-\pi(Y)})\mathrm{Ad}(e^{\pi(X)})\pi(H) \\ &= e^{\mathrm{ad}(\pi(X))}e^{-\mathrm{ad}(\pi(Y))}e^{\mathrm{ad}(\pi(X))}\pi(H). \end{aligned}$$
Now, $[X, H] = -2X$, and so $(\mathrm{ad}X)H = -2X$ and $(\mathrm{ad}X)^k H = 0$ for $k > 1$. Since π is a Lie algebra homomorphism, we similarly have $\mathrm{ad}(\pi(X))\pi(H) = -2\pi(X)$ and $(\mathrm{ad}(\pi(X)))^k\pi(H) = 0$ for $k > 1$. Thus
$$e^{\mathrm{ad}(\pi(X))}\pi(H) = \sum_{k=0}^{\infty} \frac{1}{k!}(\mathrm{ad}(\pi(X)))^k\pi(H) = \pi(H) - 2\pi(X).$$
For the next term, since $[Y, H] = 2Y$ and $[Y, X] = -H$, thus $(-\mathrm{ad}(\pi(Y)))^k\pi(H) = -2\pi(Y)\mathbb{1}_{k=1}$, while $(-\mathrm{ad}(\pi(Y)))^k\pi(X) = \pi(H)\mathbb{1}_{k=1} - 2\pi(Y)\mathbb{1}_{k=2}$. Thus
$$\begin{aligned} e^{-\mathrm{ad}(\pi(Y))}e^{\mathrm{ad}(\pi(X))}\pi(H) &= e^{-\mathrm{ad}(\pi(Y))}(\pi(H) - 2\pi(X)) \\ &= (\pi(H) - 2\pi(Y)) - 2(\pi(X) + \pi(H) - \frac{1}{2}\cdot 2\pi(Y)) \\ &= -\pi(H) - 2\pi(X). \end{aligned}$$
Finally, this means that
$$\begin{aligned} e^{\mathrm{ad}(\pi(X))}e^{-\mathrm{ad}(\pi(Y))}e^{\mathrm{ad}(\pi(X))}\pi(H) &= e^{\mathrm{ad}(\pi(X))}(-\pi(H) - 2\pi(X)) \\ &= (-\pi(H) + 2\pi(X)) - 2\pi(X) = -\pi(H) \end{aligned}$$
as claimed.

□

We conclude this section with the comparable statement of Theorem 17.37 for Lie *group* representations of $\mathrm{SU}(2)$.

COROLLARY 17.40. *For each integer $m \geq 0$, there is (up to isomorphism) precisely one complex irrep of $\mathrm{SU}(2)$ of dimension $m+1$; it is isomorphic to the representation (V_m, Π_m) of Example 17.6.*

PROOF. Let (V, Π) be an irrep of $\mathrm{SU}(2)$, and let $\pi = \Pi_*$ be the induced Lie algebra representation of $\mathfrak{su}(2)$. By Lemma 17.25, (V, π) is irreducible. Then, by Lemma 17.24, the complexification $(V, \pi_{\mathbb{C}})$ is an irrep of $\mathfrak{sl}(2,\mathbb{C}) = \mathfrak{su}(2)_{\mathbb{C}}$. It follows from Theorem 17.37 that $(V, \pi_{\mathbb{C}}) \cong (V_m, (\pi_m)_{\mathbb{C}})$ where $m = \dim(V)$. Restricting to the real form $\mathfrak{su}(2) \subset \mathfrak{sl}(2,\mathbb{C})$, it follows that $(V, \pi) \cong (V_m, \pi_m)$. Note that $\pi_m = (\Pi_m)_*$. Since $\mathrm{SU}(2)$ is connected, it therefore follows from Lemma 14.36 that $(V, \Pi) \cong (V_m, \Pi_m)$. □

EXAMPLE 17.41. The standard representation $\Pi(U) = U$ of $\mathrm{SU}(2)$ on $\mathbb{C}^2$ is irreducible (the proof is the same as for $\mathrm{SO}(n)$, cf. Example 17.21). Since it is 2-dimensional, it follows from Corollary 17.40 that $(\mathbb{C}^2, \Pi) \cong (V_1, \Pi_1)$. we can, in fact, see this explicitly: the isomorphism $\phi\colon \mathbb{C}^2 \to V_1$ can be chosen as $\phi(e_1) = f_2$ and $\phi(e_2) = -f_1$ where $f_j(z_1, z_2) = z_j$. Then, for $U \in \mathrm{SU}(2)$ and $v = v_1 e_1 + v_2 e_2$,

$$\begin{aligned}\phi[\Pi(U)v](z_1, z_2) &= (U_{11}v_1 + U_{12}v_2)\phi(e_1)(z_1, z_2) + (U_{21}v_1 + U_{22}v_2)\phi(e_2)(z_1, z_2)\\ &= (U_{11}v_1 + U_{12}v_2)z_2 - (U_{21}v_1 + U_{22}v_2)z_1.\end{aligned}$$

On the other hand,

$$\begin{aligned}[\Pi_1(U)\phi(v)](z_1, z_2) &= v_1 f_2(U^{-1}[z_1, z_2]^\top) - v_2 f_1(U^{-1}[z_1, z_2]^\top)\\ &= v_1(U_{21}^{-1}z_1 + U_{22}^{-1}z_2) - v_2(U_{11}^{-1}z_1 + U_{12}^{-1}z_2)\\ &= v_1(-U_{21}z_1 + U_{11}z_2) - v_2(U_{22}z_1 - U_{12}z_2),\end{aligned}$$

where the last equality comes from the fact that $\det U = 1$, so that

$$U^{-1} = \begin{bmatrix} U_{22} & -U_{12} \\ -U_{21} & U_{11} \end{bmatrix}.$$

Thus $\phi(\Pi(U)v) = \Pi_1(U)\phi(v)$ for all $U \in \mathrm{SU}(2)$ and $v \in \mathbb{C}^2$. Since ϕ is a linear isomorphism (it is the "ccw right-angle rotation"), we have $(\mathbb{C}^2, \Pi) \cong (V_1, \Pi_1)$ as claimed. Note that the exact same computation shows that these two representations of $\mathrm{SL}(2,\mathbb{C})$ are isomorphic; this also follows from the same reasoning as in Corollary 17.40, since $(\pi_m)_{\mathbb{C}}$ is the induced Lie algebra represention from Π_m on $\mathrm{SL}(2,\mathbb{C})$.

4. The Representation Theory of $\mathfrak{sl}(3,\mathbb{C})$

To motivate the general techniques we will use to analyze representations of many Lie groups, let's take a brief look at how the pictures gets complicated moving from $\mathfrak{sl}(2,\mathbb{C})$ to $\mathfrak{sl}(3,\mathbb{C})$. To begin, we need a basis for the Lie algebra. As there is only one (complex) constraint, $\mathfrak{sl}(3,\mathbb{C}) = \{X \in \mathfrak{gl}(3,\mathbb{C})\colon \mathrm{Tr}\,(X) = 0\}$, the algebra is 8-dimensional (over $\mathbb{C}$). The following basis naturally generalizes the basis $\{X, Y, H\}$ of $\mathfrak{sl}(2,\mathbb{C})$ used in the preceding section.

DEFINITION 17.42. *The following is the* **standard basis** *of* $\mathfrak{sl}(3,\mathbb{C})$.

$$X_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad X_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad X_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$Y_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad Y_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad Y_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad H_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Since $\mathfrak{sl}(3,\mathbb{C})$ is a Lie algebra, we know the Lie bracket of any two of these is a linear combination of them. There are $\binom{8}{2} = 28$ brackets in total, much more than the $\binom{3}{2} = 3$ in the case of $\mathfrak{sl}(2,\mathbb{C})$! To begin enumerating them, note that the first two columns each give an isomorphic copy of $\mathfrak{sl}(2,\mathbb{C})$ inside $\mathfrak{sl}(3,\mathbb{C})$:

$$\begin{aligned} [X_1, H_1] &= -2X_1 & [X_2, H_2] &= -2X_2 \\ [Y_1, H_1] &= -2Y_1 & [Y_2, H_2] &= -2Y_2 \\ [X_1, Y_1] &= H_1 & [X_2, Y_2] &= H_2 \end{aligned}$$

One might hope that these two copies of $\mathfrak{sl}(2,\mathbb{C})$ commute with each other, but this is not the case:

$$\begin{aligned} [X_1, X_2] &= X_3 & [X_1, H_2] &= X_1 \\ [X_1, Y_2] &= 0 & [Y_1, H_2] &= -Y_1 \\ [X_2, Y_1] &= 0 & [X_2, H_1] &= X_2 \\ [Y_1, Y_2] &= -Y_3 & [Y_2, H_1] &= -Y_2 \end{aligned}$$

$$[H_1, H_2] = 0.$$

The remaining commutation relations in $\mathfrak{sl}(3,\mathbb{C})$ are as follows:

$$\begin{aligned} [X_1, X_3] &= 0 & [Y_1, Y_3] &= 0 \\ [X_2, X_3] &= 0 & [Y_2, Y_3] &= 0 \\ [Y_1, X_3] &= X_2 & [X_1, Y_3] &= -Y_2 \\ [Y_2, X_3] &= -X_1 & [X_2, Y_3] &= Y_1 \\ [H_1, X_3] &= X_3 & [H_1, Y_3] &= -Y_3 \\ [H_2, X_3] &= X_3 & [H_2, Y_3] &- Y_3 \end{aligned}$$

$$[X_3, Y_3] = H_1 + H_2.$$

If (V, π) is any representation of $\mathfrak{sl}(3,\mathbb{C})$, then all the matrices $\pi(X_j), \pi(Y_j), \pi(H_j)$ will have the same commutation relations as above (since π is a Lie group homomorphism). We hope to mimic some features of the irreps of $\mathfrak{sl}(2,\mathbb{C})$ we saw in Section 3, where a standard basis was given by the eigenvectors of $\pi(H)$ and $\pi(X)$ and $\pi(Y)$ acted as "raising" and "lowering" operators through this basis. Hence, a starting point will be to try to diagonalize $\pi(H_1)$ and $\pi(H_2)$. Fortunately, since $[\pi(H_1), \pi(H_2)] = \pi([H_1, H_2]) = 0$, if both are diagonalizable they will be *simultaneously* diagonalizable (i.e. they have a common basis of eigenvectors). This motivates the following definition.

DEFINITION 17.43. *Let (V,π) be a complex representation of $\mathfrak{sl}(3,\mathbb{C})$. An ordered pair $\mu = (m_1,m_2) \in \mathbb{C}^2$ is called a* **weight** *for π if there is a common eigenvector $v \in V \setminus \{0\}$ such that*

$$\begin{aligned} \pi(H_1)v &= m_1 v \\ \pi(H_2)v &= m_2 v. \end{aligned} \tag{17.9}$$

Any such v is called a **weight vector** *for the weight μ, and the space of all such common eigenvectors is called the* **weight space** *of μ. The* **multiplicity** *of the weight μ is the dimension of its weight space.*

The weights and multiplicities are invariant under isomorphism, as can be easily checked.

LEMMA 17.44. *If (V,π) and (V',π) are isomorphic representations of $\mathfrak{sl}(3,\mathbb{C})$, then they have the same weights with the same multiplicities.*

If $\pi(H_1),\pi(H_2)$ are simultaneously diagonalizable, then (following the language of Definition 17.43) there is a basis of V consisting of weight vectors. In general, this will not be true for every representation (although it is true for the adjoint representation $(\mathfrak{sl}(3,\mathbb{C}),\mathrm{ad})$, which will be important later). However, it is always true that $\pi(H_1),\pi(H_2)$ have at least one common eigenvector.

PROPOSITION 17.45. *Every representation of $\mathfrak{sl}(3,\mathbb{C})$ has at least one weight.*

PROOF. Let (V,π) be a complex representation of $\mathfrak{sl}(3,\mathbb{C})$. Then $\pi(H_1)$ has an eigenvalue $m_1 \in \mathbb{C}$. Let $W \subseteq V$ be the eigenspace of $\pi(H_1)$ for eigenvalue m_1. Since $[\pi(H_1),\pi(H_2)] = 0$, it follows that for any $w \in W$

$$m_1\pi(H_2)w = \pi(H_2)\pi(H_1)w = \pi(H_1)\pi(H_2)w$$

which shows that $\pi(H_2)w$ is an eigenvector of $\pi(H_1)$ with eigenvalue m_1 – i.e. $\pi(H_2)w \in W$. Thus W is an invariant subspace for $\pi(H_2)$, and therefore $\pi(H_2)|_W$ is an operator which again must have an eigenvalue $m_2 \in \mathbb{C}$. If $v \in W$ is any eigenvector of $\pi(H_2)$ with eigenvalue m_2, then since $v \in W$, we see that (m_1,m_2) is a weight for π with weight vector v. □

Let us also note that all weights are in the integer lattice.

PROPOSITION 17.46. *If $\mu = (m_1,m_2)$ is a weight of a complex representation (V,π) of $\mathfrak{sl}(3,\mathbb{C})$, then $m_1,m_2 \in \mathbb{Z}$.*

PROOF. The restriction of (V,π) to the Lie subalgebra $\langle X_1,Y_1,H_1\rangle \cong \mathfrak{sl}(2,\mathbb{C})$ is a complex representation, and hence by Corollary 17.39(1), the eigenvalue m_1 of $\pi(H_1)$ is an integer. The same argument applied to the restriction to $\langle X_2,Y_2,H_2\rangle \cong \mathfrak{sl}(2,\mathbb{C})$ shows that $m_2 \in \mathbb{Z}$. □

Now that we have a common eigenvector for $\pi(H_1)$ and $\pi(H_2)$, if we mimic the construction of all irreps of $\mathfrak{sl}(2,\mathbb{C})$, we should begin applying $\pi(X_j)$ and $\pi(Y_j)$ to see what happens to this eigenvector. The calculations needed to characterize this in the proof of Theorem 17.37 relied on the commutation relations $[X,H] = -2X$ and $[Y,H] = 2Y$. Note that these say that X and Y are eigenvectors of $\mathrm{ad}H$, with eigenvalues ± 2. This motivates the following definition.

DEFINITION 17.47. *A non-zero weight for the adjoint representation $(\mathfrak{sl}(3,\mathbb{C}),\mathrm{ad})$ is called a* **root***; its weight vector is called a* **root vector**. *In other words: a $\alpha = (a_1,a_2) \in \mathbb{C}^2 \setminus \{(0,0)\}$ is called a* **root** *for $(3,\mathbb{C})$ if there is a vector $Z \in (3,\mathbb{C}) \setminus \{0\}$ such that*

$$\begin{aligned} [H_1,Z] &= a_1 Z \\ [H_2,Z] &= a_2 Z. \end{aligned}$$

Such a Z is called a **root vector** *for the root α.*

Considering all the commutation relations listed above, we can pick off 6 roots immediately: Note that H_1 and H_2 are also both weight vectors for the adjoint representation, but each has weight

weight α	weight vector Z	α	weight vector Z
$(2,-1)$	X_1	$(-2,1)$	Y_1
$(-1,2)$	X_2	$(1,-2)$	Y_2
$(1,1)$	X_3	$(-1,-1)$	Y_3

$(0,0)$, so it is not a root by Definition 17.47. Since the vectors $X_1, X_2, X_3, Y_1, Y_2, Y_3, H_1, H_2$ form a linear basis for $\mathfrak{sl}(3,\mathbb{C})$, it is not hard to check that the 6 roots listed above are the only roots.

The following lemma is the generalization of Lemma 17.38 to $\mathfrak{sl}(3,\mathbb{C})$, showing how, in any representation (V,π), $\pi(X_j)$ and $\pi(Y_j)$ act as raising and loweing operators for weight vectors (common eigenvectors of $\pi(H_1), \pi(H_2)$).

LEMMA 17.48. *Let $\alpha = (a_1, a_2)$ be a root of $\mathfrak{sl}(3,\mathbb{C})$ with root vector Z_α. Let (V,π) be a complex representation of $\mathfrak{sl}(3,\mathbb{C})$, with a weight $\mu = (m_1, m_2)$ and weight vector $v \neq 0$. Then*

$$\pi(H_1)\pi(Z_\alpha)v = (m_1 + a_1)\pi(Z_\alpha)v,$$
$$\pi(H_2)\pi(Z_\alpha)v = (m_2 + a_2)\pi(Z_\alpha)v.$$

Thus, either $\pi(Z\alpha)v = 0$ of $\pi(Z_\alpha)v$ is a new weight vector with weight $\mu+\alpha = (m_1+a_1, m_2+a_2)$.

PROOF. Since Z_α is a root vector with root α, by definition $[H_j, Z_\alpha] = a_j Z_\alpha$ for $j = 1, 2$. Thus

$$\pi(H_j)\pi(Z_\alpha)v = (\pi(Z_\alpha)\pi(H_j) + a_j\pi(Z_\alpha))v = \pi(Z_\alpha)m_j v + a_j\pi(Z_\alpha)v = (m_j + a_j)\pi(Z_\alpha)v.$$

□

Now, if we are to proceed mimicking the classification of irreps of $\mathfrak{sl}(2,\mathbb{C})$, we want to start with some weight vector v for π, and apply the root vectors until we get to a "top" weight. Things are more complicated now, with two directions to move, however. So we need a more sophisticated notion of "top".

Let us single out the roots of X_1 and X_2:

$$\alpha_1 = (2,-1), \qquad \alpha_2 = (-1,2). \tag{17.10}$$

We will call these the **positive simple roots** (although this is a fairly arbitrary choice, as we'll see). Note that $(1,1) = \alpha_1 + \alpha_2$, and so all of the roots are integer linear combinations of the positive simple roots, with all coefficients of the same sign. (Any other pair of roots with the same property would suffice for our purposes.) We use them to introduce a partial order on all weights.

DEFINITION 17.49. *Let $\mu_1, \mu_2 \in \mathbb{Z}^2$. We say μ_1 is* **higher** *than μ_2, written $\mu_1 \succcurlyeq \mu_2$ or $\mu_2 \preccurlyeq \mu_1$, if $\mu_1 - \mu_2$ is "nonnegative" in the sense that it is a nonnegative linear combination of the positive simple roots $\{\alpha_1, \alpha_2\}$ of* (17.10)*:*

$$\mu_1 \succcurlyeq \mu_2 \qquad \Longleftrightarrow \qquad \mu_1 - \mu_2 = a\alpha_1 + b\alpha_2,\ \exists\, a, b \geq 0.$$

Let (V,π) be a representation of $\mathfrak{sl}(3,\mathbb{C})$. A weight μ_0 for (V,π) is called a **highest weight** *if $\mu \preccurlyeq \mu_0$ for all weights μ of (V,π).*

Note that $\preccurlyeq$ is not a total order: for example, $\alpha_1 - \alpha_2$ cannot be written as an integer linear combination of α_1, α_2 with both coefficients of the same sign (as you can check), so it is neither $\preccurlyeq (0,0)$ or $\succcurlyeq (0,0)$. The relation $\preccurlyeq$ is, however, a partial order:

- *Reflexive*: for any $\mu \in \mathbb{Z}^2$, $\mu - \mu = (0,0) = 0\alpha_1 + 0\alpha_2$, so $\mu \preccurlyeq \mu$.

- *Antisymmetric*: if $\mu_1, \mu_2 \in \mathbb{Z}^2$ satisfy $\mu_1 \preccurlyeq \mu_2$ and $\mu_2 \preccurlyeq \mu_1$, then $\mu_1 - \mu_2 = a\alpha_1 + b\alpha_2$ and $\mu_1 - \mu_2 = a'\alpha_1 + b'\alpha_2$ with $a, a', b, b' \geq 0$; thus $a\alpha_1 + b\alpha_2 = -a'\alpha_1 - b'\alpha_2$ so $(a + a')\alpha_1 + (b + b')\alpha_2$. Since α_1, α_2 are linearly independent, it follows that $a + a' = b + b' = 0$, and as they are all ≥ 0, it follows that $a = a' = b = b' = 0$; thus $\mu_1 - \mu_2 = 0$, so $\mu_1 = \mu_2$.
- *Transitive*: if $\mu_1, \mu_2, \mu_3 \in \mathbb{Z}^2$ with $\mu_1 \preccurlyeq \mu_2$ and $\mu_2 \preccurlyeq \mu_3$, then $\mu_3 - \mu_1 = (\mu_3 - \mu_2) + (\mu_2 - \mu_1)$ is a sum of nonnegative linear combinations of the simple positive roots, and thus is a nonnegative linear combination of the positive simple roots. Hence $\mu_1 \preccurlyeq \mu_3$.

Note also that the coefficients $a, b \geq 0$ need not be integers, simply nonnegative real numbers. For example: $(1, 0) = \frac{2}{3}\alpha_1 + \frac{1}{3}\alpha_2$ is $\succcurlyeq (0, 0)$.

REMARK 17.50. The definition of the partial order depends on the choice of positive simple roots, but any other choice will give an isomorphic partial order.

As a partial order on $\mathbb{Z}^2$, $\preccurlyeq$ has no global maximal or minimal elements, of course. It will turn out the, for a given irrep of $\mathfrak{sl}(3,\mathbb{C})$, the set of weights (which is a finite subset of $\mathbb{Z}^2$) has a unique maximal element. This is part of the statement of the main theorem we now state, which generalizes Theorem 17.37, classifying the irreps of $\mathfrak{sl}(3,\mathbb{C})$.

THEOREM 17.51 (The Theorem on Highest Weights for $\mathfrak{sl}(3,\mathbb{C})$). *Fix simple positive roots of* $\mathfrak{sl}(3,\mathbb{C})$ *as in* (17.10)*, and define the partial order* $\preccurlyeq$ *as in Definition 17.49 accordingly.*

(1) *Every irrep of* $\mathfrak{sl}(3,\mathbb{C})$ *has a unique highest weight; this highest weight* μ *is in* $\mathbb{N}^2$.
(2) *Every* $\mu \in \mathbb{N}^2$ *is the highest weight of some irrep of* $\mathfrak{sl}(3,\mathbb{C})$. *Two irreps are isomorphic iff they have the same highest weight.*
(3) *Every irrep of* $\mathfrak{sl}(3,\mathbb{C})$ *is the direct sum of its weight spaces.*

This is the precise analog of the structure theorem for irreps of $\mathfrak{sl}(2,\mathbb{C})$ in Theorem 17.37. In that case, weights and weight vectors are simply eigenvalues and eigenvectors of $\pi(H)$. The "highest weight" of $(V_m, (\pi_m)_{\mathbb{C}})$ is just m. $(V_m, (\pi_m)_{\mathbb{C}})$ is the only representation with this weight up to isomorphism, giving uniqueness. And, in any irrep, the eigenvectors of $(\pi_m)_{\mathbb{C}}(H)$ form a basis for V_m; this is the analog of (3) (which is the statement that, in any irrep (V, π) of $\mathfrak{sl}(3,\mathbb{C})$, $\pi(H_1), \pi(H_2)$ are simultaneously diagonalizable). Note that the "only if" direction of (2) follows immediately from Lemma 17.44.

We will not prove Theorem 17.51. While we have all the tools to give a complete proof, it would take most of our remaining time, and this is just one more Lie algebra. The theorem generalizes considerably, not to all Lie algebras, but to all *semisimple Lie algebras* (which we will define in the next section). The interested reader is urged to study the book [**2**] for the proof of Theorem 17.51 and an excellent general introduction to the representation theory of Lie groups and Lie algebras.

We conclude this section by recasting weights in a different light. Our present Definition 17.43 defines a weight μ as a pair of eigenvalues of $\pi(H_1)$ and $\pi(H_2)$ sharing a common eigenvector v. Any such v is also a common eigenvector of any linear combination $aH_1 + bH_2 \in \mathfrak{sl}(3,\mathbb{C})$; since $[H_1, H_2] = 0$, this linear span is a Lie subalgebra. It is, in fact, a maximal abelian subalgebra.

DEFINITION 17.52. *The standard* **Cartan subalgebra** *of* $\mathfrak{sl}(3,\mathbb{C})$ *is* $\mathfrak{h} = \operatorname{span}_{\mathbb{C}}\{H_1, H_2\}$. *It is a maximal abelian subalgebra.*

Note: H_1 and H_2 are also in the real form $\mathfrak{su}(3)$; Example 16.11 showed that $\mathfrak{h}$ is a maximal abelian subalgebra of $\mathfrak{su}(3)$. It is easy to show that it is therefore also a maximal abelian subalgebra of $\mathfrak{sl}(3,\mathbb{C}) = \mathfrak{su}(2)_{\mathbb{C}}$, as claimed in Definition 17.52.

Thus, any weight vector v of a representation (V, π) is an eigenvector for all elements of the Cartan subalgebra $\mathfrak{h}$; moreover, if v's weight is (m_1, m_2), then for any element $H = aH_1 + bH_2 \in \mathfrak{h}$, we have $\pi(aH_1+bH_2)v = a\pi(H_1)v+b\pi(H_2)v = am_1v+bm_2v = (am_1+bm_2)v$. In particular, the eigenvalue of $\pi(H)$ with eigenvector v depends linearly on H. So we may think of a weight as a linear functional on $\mathfrak{h}$: if $\mu = (m_1, m_2)$, we may identify the weight as the linear functional $\mu^* \in \mathfrak{h}^*$ defined by $\mu^*(H_j) = m_j$.

Now, by choosing an inner product on $\mathfrak{h}$, we can identify $\mathfrak{h}^*$ with $\mathfrak{h}$ in the usual way. In fact, we have a nice inner product not only on $\mathfrak{h}$ but on all of $\mathfrak{sl}(3, \mathbb{C})$: the Hilbert-Schmidt inner product $\langle H, H' \rangle = \mathrm{Tr}(H^*H')$. This inner product is $\mathrm{Ad}(\mathrm{SU}(3))$-invariant. Restricted to the diagonal matrices in $\mathfrak{h}$, this is just the usual Euclidean inner product

$$\langle \mathrm{diag}(a, b, c), \mathrm{diag}(a', b', c') \rangle = \overline{a}a' + \overline{b}b' + \overline{c}c'.$$

Every linear functional $\mu^* \in \mathfrak{h}^*$ can be represented in the form

$$\mu^*(H) = \langle \lambda, H \rangle$$

for a unique element $\lambda \in \mathfrak{h}$. Hence, we may think of a weight (in the new light above) as an element of $\mathfrak{h}$; then Definition 17.43 becomes the following.

DEFINITION 17.53. *Let $\mathfrak{h} \subset \mathfrak{sl}(3, \mathbb{C})$ be the standard Cartan subalgebra. Let (V, π) be a complex representation of $\mathfrak{sl}(3, \mathbb{C})$. An element $\lambda \in \mathfrak{h}$ is called a* **weight** *for this representation if there is a nonzero vector $v \in V$ such that*

$$\pi(H)v = \langle \lambda, H \rangle v$$

for all $H \in \mathfrak{h}$. Such a vector v is a **weight vector** *of λ; the linear span of all weight vectors is the* **weight space***, and its dimension is the* **multiplicity** *of λ.*

To match up with our previous terminology: if λ is a weight in our new sense, then the corresponding old weight (m_1, m_2) is given by

$$(m_1, m_2) = (\langle \lambda, H_1 \rangle, \langle \lambda, H_2 \rangle).$$

Note: the roots are just the nonzero weights of the adjoint representation, and so in the new language, the roots are also elements of $\mathfrak{h}$. Let's calculate them for the positive simple roots α_1, α_2: in fact, $\alpha_1 \cong H_1$ and $\alpha_2 \cong H_2$, since

$$(\langle H_1, H_1 \rangle, \langle H_1, H_2 \rangle) = (2, -1), \quad (\langle H_2, H_1 \rangle, \langle H_2, H_2 \rangle) = (-1, 2).$$

Hence, the 6 roots of $\mathfrak{sl}(3, \mathbb{C})$ are the elements $\pm H_1, \pm H_2, \pm(H_1 + H_2)$ in $\mathfrak{h}$.

Now, consider a representation (V, Π) of the Lie group $\mathrm{SU}(3)$. Then Π_* is a representation of $\mathfrak{su}(3)$, and so its complexification $\pi = (\Pi_*)_{\mathbb{C}}$ is a representation of $\mathfrak{sl}(3, \mathbb{C})$. Hence, it has weights $\lambda \in \mathfrak{h}$, using the new Definition 17.53. The weights thus live in $\mathfrak{h} = \mathfrak{t}_{\mathbb{C}}$, where $\mathfrak{t}$ is the maximal abelian subalgebra of $\mathfrak{su}(3)$ corresponding to the maximal torus T of diagonal elements.

Recall the **Weyl group** from Section 16.3: $W(T) = N(T)/T$, where $N(T)$ is the normalizer of the maximal torus $T \subset \mathrm{SU}(3)$. It is a finite discrete group. By Proposition 16.24, there is a well-defined action of W on $\mathfrak{t} = \mathrm{Lie}(T)$, given by $w \cdot H = \mathrm{Ad}(U)H$ for any $H \in \mathfrak{t}$ and any $U \in w$. This, of course, extends to an action of $W(T)$ on $\mathfrak{t}_{\mathbb{C}} = \mathfrak{h}$, given by the same formula. As discussed following Lemma 16.25, the linear operator $H \mapsto w \cdot H$ is in $\mathrm{U}(\mathfrak{t})$ whenever $\mathfrak{t}$ is imbued with an $\mathrm{Ad}(\mathrm{SU}(3))$-invariant (complex) inner product.

The next proposition, with which we conclude this introductory discussion, shows that the Weyl group acts as a symmetry group of the weights of any representation.

PROPOSITION 17.54. *Let (V, Π) be a complex representation of $\mathrm{SU}(3)$, with induced representation (V, π) of $\mathfrak{sl}(3, \mathbb{C})$. Let $T \subset \mathrm{SU}(3)$ denote the maximal torus of diagonal elements, with Lie algebra $\mathfrak{t}$, and let $\mathfrak{h} = \mathfrak{t}_{\mathbb{C}}$ denote the standard Cartan subalgebra of $\mathfrak{sl}(3, \mathbb{C})$. If $w \in W(T)$ is in the Weyl group, and if $\lambda \in \mathfrak{h}$ is a weight for (V, π), then $w \cdot \lambda$ is also a weight for (V, π), with the same multiplicity. In particular, taking $(V, \pi) = (\mathfrak{sl}(3, \mathbb{C}), \mathrm{ad})$, the roots are invariant under the action of the Weyl group.*

PROOF. First, note that for any $U \in \mathrm{SU}(2)$ and $H \in \mathfrak{h}$,

$$\pi(H)\Pi(U) = \Pi(U)(\Pi(U)^{-1}\pi(H)\Pi(U)) = \Pi(U)C_{\Pi(U^{-1})}(\pi(H)) = \Pi(U)\pi(\mathrm{Ad}(U^{-1})H).$$

Now, let $U \in N(T)$; then $U^{-1} \in N(T)$ as well, so that $U^{-1}gU \in T$ for all $g \in T$. Differentiating with respect to g, this shows that $\mathrm{Ad}(U^{-1})H \in \mathfrak{t}$ for all $H \in \mathfrak{t}$, and so $\mathrm{Ad}(U^{-1})H \in \mathfrak{h}$ for any $H \in \mathfrak{h}$.

Now, let $\lambda \in \mathfrak{h}$ be a weight for (V, π) with weight vector v. Then for $U \in N(T)$ and $H \in \mathfrak{h}$, since $\mathrm{Ad}(U^{-1})H \in \mathfrak{h}$ we have $\pi(\mathrm{Ad}(U^{-1})H)v = \langle \lambda, \mathrm{Ad}(U^{-1})H\rangle v$, and hence the above calculation shows that

$$\pi(H)\Pi(U)v = \Pi(U)\pi(\mathrm{Ad}(U^{-1})H)v = \langle \lambda, \mathrm{Ad}(U^{-1})H\rangle \Pi(U)v.$$

Thus, if $w = [U] \in W(T)$, then $w^{-1} \cdot H = \mathrm{Ad}(U^{-1})H$, and we have

$$\pi(H)\Pi(U)v = \langle \lambda, w^{-1} \cdot H\rangle \Pi(U)v = \langle w \cdot \lambda, H\rangle \Pi(U)v$$

where, in the last equality, we have used the fact that the inner product is $\mathrm{Ad}(\mathrm{SU}(2))$-invariant, and so the action of $w \in W(T)$ is unitary (i.e. $w^{-1} = w^*$). Thus, we see that $\Pi(U)v$ is a weight vector for (V, π) with weight $w \cdot \lambda$.

Since $\Pi(U)$ is an isomorphism of V, it follows that it maps the weight space of λ onto the weight space of $w \cdot \lambda$ isomorphically. Thus, the two weights have the same multiplicities. □

Hence, the Weyl group can be viewed as a group of symmetries of the weights of any representation, and in particular of the roots. Time permitting, we will later see that, in this form, the Weyl group consists exactly of the group generated by the reflections across the hyperplanes orthogonal to the roots.

5. Semisimple Lie Groups and Representations: an Overview

We now examine how far we can push the strategy of the previous two sections in classifying irreps of Lie algebras. The jump in complexity from $\mathfrak{sl}(2, \mathbb{C})$ to $\mathfrak{sl}(3, \mathbb{C})$ was significant (indeed, we left out most of the proofs). It turns out this was the biggest jump. It is not hard to imagine how to proceed to work out the representation theory of $\mathfrak{sl}(n, \mathbb{C})$ in general. In fact, the same ideas work for a very large class of Lie algebras. Let us define them now.

DEFINITION 17.55. *A finite-dimensional Lie algebra $\mathfrak{g}$ is called* **simple** *if it is nonabelian and has no nontrivial ideals. It is* **semisimple** *if it is a finite direct sum of simple Lie algebras.*

Many of the Lie algebras we've studied turn out to be semisimple (or even simple). Nevertheless, the question of verifying this is a laborious one that is very algebraic. In better sync with our goals, we consider an ostensibly different class of Lie algebras.

DEFINITION 17.56. *A Lie algebra $\mathfrak{k}$ is called* **compact** *if there is a compact Lie group K with $\mathrm{Lie}(K) \cong \mathfrak{k}$. A Lie algebra $\mathfrak{g}$ is called* **reductive** *if there is a compact Lie algebra $\mathfrak{k}$ with $\mathfrak{g} = \mathfrak{k}_{\mathbb{C}}$.*

Conversely, given a (complex) Lie algebra $\mathfrak{g}$, a real Lie subalgebra $\mathfrak{k} \subset \mathfrak{g}$ is called a **compact real form** *of $\mathfrak{g}$ if it is a compact Lie algebra and $\mathfrak{g} = \mathfrak{k} + i\mathfrak{k}$.*

In particular, a Lie algebra is reductive iff it possesses a compact real form.

REMARK 17.57. The algebraically-minded reader may be unsatisfied with the preceding definition of *compact Lie algebra*, as it is extrinsic to the world of Lie algebras. In fact, there is an intrinsic characterization. The **Killing form** of a Lie algebra $\mathfrak{g}$ is the bilinear form $B\colon \mathfrak{g} \times \mathfrak{g} \to \mathbb{C}$ defined by $B(X, Y) = \mathrm{Tr}(\mathrm{ad}(X)\mathrm{ad}(Y))$ (where the trace is on the space $\mathrm{End}(\mathfrak{g})$ where $\mathrm{ad}(X)$ and $\mathrm{ad}(Y)$ live). It is a theorem that a Lie algebra is compact if and only if its Killing form is negative semidefinite. (Among these, the only ones that are degenerate are the abelian Lie algebras; all other Killing forms of compact Lie algebras are strictly negative definite.)

EXAMPLE 17.58. Since $\mathfrak{su}(n)_{\mathbb{C}} = \mathfrak{sl}(n, \mathbb{C})$ and $\mathfrak{su}(n)$ is a compact Lie algebra, $\mathfrak{sl}(n, \mathbb{C})$ is reductive. The same is true of $\mathfrak{gl}(n, \mathbb{C}) = \mathfrak{u}(n)_{\mathbb{C}}$. On their face, these two seem pretty similar, but there is an easy structural difference: the center of $\mathfrak{gl}(n, \mathbb{C})$ is nontrivial, consisting of the 1-dimensional span of the identity matrix. On the other hand, $\mathrm{sl}(n, \mathbb{C})$ has trivial center, consisting only of $\{0\}$ when $n \geq 2$.

To see this, recall Example 16.11 where we showed that the torus Lie algebra $\mathfrak{t}$ of diagonal matrices in $\mathfrak{su}(n, \mathbb{C})$ is a maximal abelian subalgebra. The proof came from showing that, for any matrix X and any j, k,

$$[(E_{jj} - E_{kk})X]_{jk} = X_{jk} = -[X(E_{jj} - E_{kk})X]_{jk}.$$

It then follows that if X commutes with all the elements $E_{jj} - E_{kk} \in \mathfrak{sl}(n, \mathbb{C})$, then $X_{jk} = -X_{jk} = 0$ for $j \neq k$, and so X must be diagonal. But then for $j < k$ we may test it against the matrix unit $E_{jk} \in \mathfrak{sl}(n, \mathbb{C})$ instead, to find that

$$0 = [X, E_{jk}] = (X_{jj} - X_{kk})E_{jk}$$

which shows that X has all equal diagonal entries, and so is a scalar multiple of the identity. Finally, since $X \in \mathfrak{sl}(n, \mathbb{C})$, it has trace 0, and so $X = 0$.

This difference turns out to be precisely the difference between reductive and semisimple.

THEOREM 17.59. *A finite-dimensional complex Lie algebra $\mathfrak{g}$ is semisimple if and only if it is reductive and has trivial center.*

The "if" direction of Theorem 17.59 is not too hard to prove. The idea is similar to the proof of Proposition 17.33 (that unitary representations have the complete reducibility property). Since $\mathfrak{g}$ is reductive, choose a compact real form $\mathfrak{k} \subset \mathfrak{g}$, and a compact Lie algebra K with Lie algebra $\mathfrak{k}$. Fix an $\mathrm{Ad}(K)$-invariant inner product on $\mathfrak{k}$; then $\mathrm{ad}X$ is skew self-adjoint for each $X \in \mathfrak{k}$. Since ideals in $\mathfrak{g}$ are simply invariant subspaces for the adjoint representation, this can be used in short order to show that any ideal in $\mathfrak{g}$ has a complementary ideal. Iterating this shows that any reductive Lie algebra is a finite direct sum of maximal ideals. The only wrinkle is that some of these ideals may well be abelian. The trivial center condition rules this out (since, as can be quickly calculated, complementary ideals commute with each other, so if any one is abelian, it is contained in the center).

On the other hand, the "only if" direction is very involved algebra, and far beyond the scope of this course. So, while we will discuss the representation theory of semisimple Lie algebras, the skeptical reader may content themselves only with the "special case" of reductive Lie algebras

with trivial center. Nevertheless, we will use the theorem to identify the two notions, and therefore use the term "semisimple" throughout.

Now, to mimic our approach to classifying representations, we need something like the standard Cartan subalgebra of $\mathfrak{sl}(3, \mathbb{C})$.

DEFINITION 17.60. *Let $\mathfrak{g}$ be a complex semisimple Lie algebra. A* **Cartan subalgebra** $\mathfrak{h}$ *of $\mathfrak{g}$ is a maximal abelian subalgebra with the property that $\mathrm{ad}(H)$ is diagonalizable for each $H \in \mathfrak{h}$.*

Note that, since $\mathfrak{h}$ is abelian, all elements commute, and therefore so do all $\mathfrak{ad}(H)$; thus, since they are all diagonalizable, they are simultaneously diagonalizable by the spectral theorem. This will allow us to define roots and weights just as in Definition 17.53.

Definition 17.60 makes sense in any Lie algebra, semisimple or not. We will see in a moment that semisimple Lie algebras always have Cartan subalgebras; by this definition, others may not. In fact, this is not the usual definition of Cartan subalgebra. The more general notion, introduced in Élie Cartan's PhD thesis, is as follows.

DEFINITION 17.61. *Let $\mathfrak{g}$ be a Lie algebra. A Lie subalgebra $\mathfrak{h} \subseteq \mathfrak{g}$ is called* **nilpotent** *if there is some finite n so that $\mathrm{ad}(X_1)\mathrm{ad}(X_2)\cdots\mathrm{ad}(X_n) = 0$ for all $X_1, \ldots, X_n \in \mathfrak{g}$. A* **general Cartan subalgebra** $\mathfrak{h}$ *of $\mathfrak{g}$ is a nilpotent subalgebra that is* self-normalizing*: if $X \in \mathfrak{g}$ is such that $\mathrm{ad}(X)$ maps $\mathfrak{h}$ into $\mathfrak{h}$, then $X \in \mathfrak{h}$.*

The two notions of Cartan subalgebra in Definitions 17.60 and 17.61 are not equivalent in general. It turns out that, in a semisimple Lie algebra, they *are* equivalent. Again, showing this would be a long algebraic discussion, immaterial to our goals. For our purposes, Definition 17.60 gives the structure we need to classify irreps.

PROPOSITION 17.62. *Let $\mathfrak{g} = \mathfrak{k}_{\mathbb{C}}$ be a complex semisimple Lie algebra, and let $\mathfrak{t} \subset \mathfrak{k}$ be a maximal abelian subalgebra. Then $\mathfrak{h} \equiv \mathfrak{t}_{\mathbb{C}} = \mathfrak{t} + i\mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{g}$.*

PROOF. Since $\mathfrak{t}$ is abelian, clearly so is $\mathfrak{h} = \mathfrak{t}_{\mathbb{C}}$. We wish to see that it is maximal abelian. So, suppose $X \in \mathfrak{g}$ commutes with $\mathfrak{h}$. Since $\mathfrak{t} \subset \mathfrak{h}$, X therefore commutes with $\mathfrak{t}$. Decomposing $X = X_1 + iX_2 \in \mathfrak{k} + i\mathfrak{k}$, this shows that

$$0 = [X, H] = [X_1, H] + i[X_2, H], \qquad \forall\, H \in \mathfrak{t}.$$

Since the decomposition of any element in $\mathfrak{g}$ (in this case 0) into $\mathfrak{k}$ and i$\mathfrak{k}$ is unique, it follows that $[X_1, H] = [X_2, H] = 0$, and so $X_1, X_2 \in \mathfrak{t}$ since $\mathfrak{t}$ is maximal abelian. Thus $X \in \mathfrak{t} + i\mathfrak{t} = \mathfrak{h}$, so $\mathfrak{h}$ is maximal abelian.

Now, let K be a compact Lie group with Lie algebra $\mathfrak{k}$, and fix any $\mathrm{Ad}(K)$-invariant inner product on $\mathfrak{k}$. Since $\langle \mathrm{Ad}(g)X, \mathrm{Ad}(g)Y\rangle = \langle X, Y\rangle$ for all $g \in K$, differentiating with respect to g shows that

$$\langle \mathrm{ad}(Z)X, Y\rangle = -\langle X, \mathrm{ad}(Z)Y\rangle, \qquad \forall\, X, Y, Z \in \mathfrak{k}.$$

Thus, for each $H \in \mathfrak{t} \subset \mathfrak{k}$, $\mathrm{ad}H$ is skew self-adjoint, and hence diagonalizable. So if $H \in \mathfrak{h} = \mathfrak{t}_{\mathbb{C}}$, then $H = H_1 + iH_2$ for $H_1, H_2 \in \mathfrak{t}$, and since $\mathfrak{t}$ is abelian, $[H_1, H_2] = 0$, and thus $\mathrm{ad}H_1, \mathrm{ad}H_2$ also commute. Since they are both diagonalizable, they are therefore simultaneously diagonalizable, and hence $\mathrm{ad}H = \mathrm{ad}H_1 + i\mathrm{ad}H_2$ is diagonalizable. □

So we now have plenty of examples of Cartan subalgebras. In fact, these are the only ones, due to the following theorem.

THEOREM 17.63. *Let $\mathfrak{g}$ be a Lie algebra, and let $\mathfrak{h}_1$ and $\mathfrak{h}_2$ be Cartan subalgebras. Then there is an automorphism of $\mathfrak{g}$ that maps $\mathfrak{h}_1$ onto $\mathfrak{h}_2$.*

(Again, we will make no attempt to prove this purely algebraic result.) A complex automorphism maps a compact real form to a compact real form, and so it follows that all Cartan subalgebras of a semisimple Lie algebra are complexifications of maximal abelian subalgebras. These all have the same dimension (by Cartan's torus Theorem 16.19 and Theorem 16.16), and so we may talk about the **rank** of a Lie algebra being the dimension of any Cartan subalgebra. (In the semisimple case, this naturally generalizes our earlier notion of rank for compact Lie groups and their Lie algebras: the dimension of any maximal torus.) For example: $\mathfrak{sl}(n)$ has rank $n-1$.

Now that we have a maximal abelian subalgebra of elements whose adjoint representations are diagonalizable, we can proceed to try to construct the framework for the classification of irreps we discussed for $\mathfrak{sl}(3,\mathbb{C})$.

DEFINITION 17.64. *Let $\mathfrak{g}$ be a semisimple Lie algebra, with a compact real form $\mathfrak{k}$. Fix a Cartan subalgebra $\mathfrak{h} = \mathfrak{t}_{\mathbb{C}}$ where $\mathfrak{t}$ is a maximal abelian subalgebra of $\mathfrak{k}$. Let K be a compact Lie group with Lie algebra $\mathfrak{k}$, and fix an $\mathrm{Ad}(K)$-invariant inner product in $\mathfrak{k}$, and thus on $\mathfrak{t}$. Extend it to a (complex) $\mathrm{Ad}(K)$-invariant inner product on $\mathfrak{h}$ by*

$$\langle X+iY, X'+iY'\rangle \equiv \langle X, X'\rangle + \langle Y, Y'\rangle + i(\langle X, Y'\rangle - \langle Y, X'\rangle).$$

Let (V,π) be a complex representation of $\mathfrak{g}$. An element $\mu \in \mathfrak{h}$ is called a **weight** *for (V,π) (relative to $\mathfrak{h}$) if there is a nonzero vector $v \in V$ so that*

$$\pi(H)v = \langle \mu, H\rangle v, \qquad \forall\, H \in \mathfrak{h}.$$

Such a vector v is called a **weight** *vector for μ; the space of all weight vectors for μ is the* **weight space***, and its dimension is the* **multiplicity** *of μ.*

The **roots** *of $\mathfrak{g}$ (relative to $\mathfrak{h}$) are the nonzero weights of the adjoint representation of $\mathfrak{g}$; that is, $\alpha \in \mathfrak{h} \setminus \{0\}$ is a root if there is some nonzero $X \in \mathfrak{g}$ such that*

$$[H, X] = \langle \alpha, H\rangle X, \qquad \forall\, H \in \mathfrak{h}$$

The set of all roots is denoted $R(\mathfrak{g}|\mathfrak{h})$. Any such vector X is a **root vector** *for α. The space of all root vectors for a given root α is the* **root space** *$\mathfrak{g}_\alpha$.*

In fact, the roots necessarily lie on the "imaginary axis" in $\mathfrak{h}$.

PROPOSITION 17.65. *Each root $\alpha \in R(\mathfrak{g}|\mathfrak{h})$ belongs to $i\mathfrak{t} \subset \mathfrak{h}$.*

PROOF. As in the proof of Proposition 17.62, for each $H \in \mathfrak{t}$ is a skew self-adjoint operator on $\mathfrak{h}$, and so $\mathrm{ad}H$ has purely imaginary eigenvalues. Let $X \in \mathfrak{g}_\alpha$ be a root vector; the equation $[H, X] = \langle \alpha, H\rangle X$ shows that X is an eigenvector of $\mathrm{ad}H$ with eigenvalue $\langle \alpha, H\rangle$. It follows that this inner product is purely imaginary for $H \in \mathfrak{t}$. Since the inner product is real-valued on $\mathfrak{t}$, this can only happen if $\alpha \in i\mathfrak{t}$. □

Now, since the vectors $H \in \mathfrak{h}$ have adjoint representations $\mathrm{ad}(H)$ that are simultaneously diagonalizable, there is a basis of $\mathfrak{g}$ of common eigenvectors for all $\mathrm{ad}H$. Among these are elements of H, which are eigenvectors with eigenvalue 0; the others (nonzero elements with nonzero eigenvectors) are root vectors for various roots. Moreover, eigenvectors with distinct eigenvalues are linearly independent. In summary, we have the following.

PROPOSITION 17.66. *Let $\mathfrak{g}$ be a semisimple Lie algebra, and let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$. Then*

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in R(\mathfrak{g}|\mathfrak{h})} \mathfrak{g}_\alpha.$$

(This is a vector space sum, not necessarily a Lie algebra direct sum: the elements of $\mathfrak{h}$ typically do not commute with the elements of $\mathfrak{g}_\alpha$.)

Now, from the finite set $R(\mathfrak{g}|\mathfrak{h})$, we select a subset $R_+(\mathfrak{g}|\mathfrak{h})$ with the property that every root is an integer linear combination of elements of $R(\mathfrak{g}|\mathfrak{h})$, with all coefficients of the same sign. (It takes significant work to show that the discrete subset $R(\mathfrak{g}|\mathfrak{h}) \subset i\mathfrak{t}$ has the right symmetries to make such a spanning condition possible.) We call the elements of $R_+(\mathfrak{g}|\mathfrak{h})$ the **simple positive roots**; $R_+(\mathfrak{g}|\mathfrak{h})$ is sometimes called a **base** for $R(\mathfrak{g}|\mathfrak{h})$.

We define on $\mathfrak{h}$ the same partial order from Definition 17.49: say that $\mu_1 \succcurlyeq \mu_2$ iff $\mu_1 - \mu_2$ is a nonnegative linear combination of simple positive roots. This makes the weights of any representation into a finite partial order.

Now, for any root $\alpha \in R(\mathfrak{g}|\mathfrak{h})$, its **coroot** H_α is the rescaled vector

$$H_\alpha = \frac{2\alpha}{\langle \alpha, \alpha \rangle}.$$

DEFINITION 17.67. *A weight $\mu \in \mathfrak{h}$ is called* **integral** *if $\langle \mu, H_\alpha \rangle \in \mathbb{Z}$ for all $\alpha \in R(\mathfrak{g}|\mathfrak{h})$. It is called* **dominant** *if $\langle \alpha, \mu \rangle \geq 0$ for all $\alpha \in R_+(\mathfrak{g}|\mathfrak{h})$.*

For example: in $\mathfrak{sl}(3, \mathbb{C})$ with its standard Cartan subalgebra, and positive simple roots $\alpha_1 = H_1, \alpha_2 = H_2$, the associated coroots are also $H_{\alpha_j} = H_j$ (since $\langle H_j, H_j \rangle = 2$), and so integral weights are those whose inner products with H_1, H_2 are integers. It is easy to check that this gives precisely the integer lattice $\mathbb{Z}^2$. The conditions $\langle H_j, \mu \rangle \geq 0$ then specify the nonnegative integer lattice $\mathbb{N}^2$.

We can now state (superficially) the exact analog of Theorem 17.51.

THEOREM 17.68 (The Theorem on Highest Weights). *Let $\mathfrak{g}$ be a complex semisimple Lie algebra. Fix a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, and a base of simple positive roots.*

(1) *Every irrep of $\mathfrak{g}$ has a unique highest weight, which is dominant integral weight.*
(2) *Every dominant integral weight is the highest weight of some irrep of $\mathfrak{g}$. Two irreps are isomorphic iff they have the same highest weight.*
(3) *Every irrep of $\mathfrak{g}$ is the direct sum of its weight spaces.*

We have most of the technology needed to prove Theorem 17.68 in place already, but the exercise would take a minimum of 5 more weeks. What's more, the theorem doesn't tell us much unless we have some kind of machinery to actually compute the roots, weights, and dominant integral elements, and use them (together with some sophisticated raising and lowering type operators as in the $\mathfrak{sl}(2, \mathbb{C})$ case) to actually understand the irreps. This can be done nearly completely, and would take us another solid 5 weeks. So, a real understanding of the representation theory of semisimple Lie algebras would require another full quarter.

For each semisimple Lie algebra, its collection of roots forms what is abstractly known as a **root system**. For example, the root system of $\mathfrak{sl}(n+1, \mathbb{C})$ is called A_n. There are four infinite families of root systems, creatively called A_n, B_n, C_n, and D_n. Three other "families" of root systems exists, called "sporadic": E_6, E_7, E_8, along with F_4 and G_2. We will make no attempt to describe all of them here. They are classified by *Dynkin diagrams*. The interested reader is urged to consult the book [**2**] for further details.

One more note: a somewhat more sophisticated version of the proof of Proposition 17.33 shows that all semisimple Lie algebras have the complete reducibility property. Hence, once we understand the irreps, we understand all representations.

6. Schur's Lemma

We close out this chapter with an elementary tool we will need in the sequel. The main result of this section (Schur's Lemma) applies simultaneously to groups and Lie algebras. In order to state it effectively, we use a standard abuse of notation here, and refer to "the representation V" instead of the representation (V, Π) or (V, π) – highlighting the fact that we are talking about a linear action of the group / algebra on V.

THEOREM 17.69 (Schur's Lemma). *Let V and W be irreducible (real or complex) representations of a Lie group or Lie algebra.*

(1) *If $\phi\colon V \to W$ is an intertwining map, then either $\phi = 0$ or ϕ is an isomorphism.*
(2) *If V is a complex irrep, and $\phi\colon V \to V$ is an intertwining map, then $\phi = \lambda I$ for some $\lambda \in \mathbb{C}$.*
(3) *If V, W are complex irreps and $\phi, \psi\colon V \to W$ are nonzero intertwining maps, then $\phi = \lambda\psi$ for some $\lambda \in \mathbb{C}$.*

PROOF. As usual, we will provide the proofs in the case of representations of a group G; the Lie algebra case is identical modula the necesssary changes in notation.

(1) Let $v \in \ker\phi$. Since ϕ is an intertwining map, for any $g \in G$

$$\phi(g \cdot v) = g \cdot \phi(v) = g \cdot 0 = 0$$

because the action of G is linear. Hence $g \cdot v \in \ker\phi$ as well, and so $\ker\phi$ is an invariant subspace. Since the representation is irreducible, it follows that either $\ker\phi = V$ or $\ker\phi = 0$, meaning that either $\phi = 0$ or ϕ is one-to-one. In the latter case, the image $\phi(V) \subseteq W$ is a non-zero subspace. It is, in fact, an invariant subspace: if $w = \phi(v) \in \phi(V)$ then, for all $g \in G$, $g \cdot w = g \cdot \phi(v) = \phi(g \cdot v) \in \phi(V)$. Thus, since $\phi(V) \neq 0$ and the representation is irreducible, it follows that $\phi(V) = W$, and so ϕ is onto. Thus, if $\phi \neq 0$, then ϕ is a linear isomorphism. Since it is an intertwining map, it is therefore an isomorphism of the two representations.
(2) Since $\phi\colon V \to V$ is a $\mathbb{C}$-linear map, it has eigenvalues, so let $\lambda \in \mathbb{C}$ be an eigenvalue. Denote the representation as (V, Π). Since ϕ is an intertwiner of this representation (with itself), this means that $\phi(\Pi(g)v) = \Pi(g)\phi(v)$ for all $v \in V$ and $g \in G$, meaning that the matrices $\phi\colon V \to V$ and $\Pi(g)\colon V \to V$ commute. Hence, if U is the eigenspace of ϕ for eigenvalue λ, $\Pi(g)$ leaves U invariant: for $u \in U$,

$$\phi(\Pi(g)u) = \Pi(g)\phi(u) = \Pi(g)\lambda u = \lambda\Pi(g)u$$

showing that $\Pi(g)u$ is also an eigenvector of ϕ with eigenvalue λ. Thus, U is an invariant subspace for the irrep (V, Π), and hence either $U = 0$ or $U = V$. Since λ is an eigenvalue, $U \neq 0$, and so $U = V$. Thus, for all $v \in V = U$, $\phi(v) = \lambda v$, and so $\phi = \lambda I$.
(3) Since $\psi \neq 0$, by (1) it is an isomorphism, and (as we can easily check) $\psi^{-1}\colon V \to W$ is an intertwining map. Thus $\phi \circ \psi^{-1}\colon V \to V$ is an intertwining map from V to itself, and so by (2) it is constant, $\phi \circ \psi^{-1} = \lambda I$ for some $\lambda \in \mathbb{C}$. This shows that $\phi = \lambda\psi$ as claimed.

□

Schur's lemma is one of the most important elementary tools in representation theory; we will use it frequently from now on. We conclude this short section with two immediate corollaries.

COROLLARY 17.70. *If (V, Π) is a complex irrep of a Lie group G, and if $g \in Z(G)$ is in the center of G, then $\Pi(g) = \lambda I$ for some $\lambda \in \mathbb{C}$. Simialry, if (V, π) is a complex irrep of a Lie algebra $\mathfrak{g}$, and if $X \in Z(\mathfrak{g})$ is in the center of $\mathfrak{g}$, then $\pi(g) = \lambda I$ for some $\lambda \in \mathbb{C}$.*

PROOF. Since $g \in Z(G)$, for any $h \in G$ we have $\Pi(g)\Pi(h) = \Pi(gh) = \Pi(hg) = \Pi(h)\Pi(g)$, which shows that $\Pi(g)$ is an intertwiner of (V, Π) with itself. The result now follows from Schur's Lemma, part (2). The Lie algebra case is analogous. □

COROLLARY 17.71. *Every complex irrep of an abelian group or Lie algebra is one-dimensional.*

PROOF. Let (V, Π) denote the complex irrep. If G is abelian then $G = Z(G)$, and so Corollary 17.70 shows that, for every $g \in G$, there is a $\lambda = \lambda(g) \in \mathbb{C}$ with $\Pi(g) = \lambda(g)I$. But this means that *every* subspace of V is invariant, since every vector is an eigenvector for all $\Pi(g)$. Hence, since the representation is irreducible, the only way it can fail to have nontrivial invariant subspaces is if there are no nontrivial subspaces, meaning V is one-dimensional (since we do not include the 0-dimensional space in the definition of representations). □

CHAPTER 18

Representations of Compact Lie Groups

Throughout this chapter, K will be a compact Lie group. From Proposition 12.6, we know that each such K possesses a unique left-invariant Haar measure. It will be useful to note that this measure is *also right invariant*.

PROPOSITION 18.1. *For compact Lie group K, the left-invariant Haar measure is in fact the unique bi-invariant probability measure.*

PROOF. For the purposes of this proof, let μ denote the left Haar measure of K; so $\mu(K) = 1$ and, for any fixed $y \in K$ and $f \in C(K)$,

$$\int_K f(yx)\,\mu(dx) = \int_K f(x)\,\mu(dx).$$

Now, fix $z \in K$ and let μ_z be the Borel measure on K given by $\mu_z(B) = \mu(Bz)$ for any Borel set $B \subseteq K$, or equivalently

$$\int_K f(x)\,\mu_z(dx) = \int_K f(xz)\,\mu(dx), \qquad f \in C(K).$$

Note that $Kz = K$, and so $\mu_z(K) = 1$, so it is still a probability measure. Also, for any $y \in K$ and $f \in C(K)$, we have

$$\int_K f(yx)\,\mu_z(dx) = \int_K f(yxz)\,\mu(dx) = \int_K (R_z^* f)(yx)\,\mu(dx)$$

where $R_z^* f(a) = f(az)$. Since group multiplication is continuous, $R_z^* f \in C(K)$, and therefore by left-invariance of μ, this equals

$$\int_K (R_z^* f)(x)\,\mu(dx) = \int_K f(xz)\,\mu(dx) = \int_K f(x)\,\mu_z(dx).$$

Hence, we have shown that, for all $f \in C(K)$ and all $z, y \in K$, $\int_K L_y^* f\,d\mu_z = \int_K f\,d\mu_z$, which means that μ_z is a left-invariant measure. Since it is also a probability measure, by the uniqueness clause of Proposition 12.6, we must have $\mu_z = \mu$ for all $z \in K$. This shows that μ is also right-invariant. □

REMARK 18.2. The preceding proof used, in a fundamental way, the fact that the Haar measure on a compact group can be normalized. If one tries to make the same argument work without normalization, on a noncompact Lie group G, knowing only that the Haar measure is unique up to scale, then one finds instead that μ and μ_z are related by a function $\mu_z = \Delta(z)\mu$. The function $\Delta\colon G \to \mathbb{R}_+$ is called the *modular function* of the group. It is a group homomorphism. Compact groups are *unimodular*: the function Δ is constantly equal to 1. The same is true for abelian groups, discrete Lie groups, and in fact all semisimple Lie groups (i.e. Lie groups whose Lie algebras are semisimple). There are, however, many examples of non-unimodular Lie groups. Their representation theory turns out to be a lot harder to understand.

1. Representative Functions and Characters

Fix a finite-dimensional representation (V, Π) of K, with dimension d. Thus $\Pi\colon K \to \mathrm{GL}(V)$ is a group homomorphism, and so for each $x \in K$, $\Pi(x)$ is a linear operator on V. If we fix a basis $\{e_j\}_{j=1}^d$ for V, then we can think of $\Pi(x)$ as a matrix. Hence, note that, for $x, y \in K$,

$$[\Pi(xy)]_{jk} = [\Pi(x)\Pi(y)]_{jk} = \sum_{\ell=1}^{d} [\Pi(x)]_{j\ell}[\Pi(y)]_{\ell k}. \tag{18.1}$$

Denote by $\Pi_{jk}\colon K \to \mathbb{C}$ the *matrix entry functions* associated to the representation Π. The above calculation says that, for fixed j, k, and fixed $y \in K$,

$$\text{the function } x \mapsto \Pi_{jk}(xy) \text{ is a linear combination of the functions } \{\Pi_{j\ell}\colon 1 \le \ell \le d\}.$$

Thus, the $\le d^2$-dimensional space of matrix entry functions $\{\Pi_{jk}\colon 1 \le j, k \le d\}$ is invariant under right-translation by the group (on the variable of the function).

To (apparently) generalize this kind of function, we introduce the *left* and *right regular representations* (which are not finite dimensional in general).

DEFINITION 18.3. *Let $L^2(K)$ denote the Hilbert space of L^2-functions on K with respect to the Haar measure. The* **left** *and* **right regular representations** $(L^2(K), \mathcal{L})$ *and* $(L^2(K), \mathcal{R})$ *are defined by*

$$[\mathcal{L}(x)f](y) = f(x^{-1}y), \quad [\mathcal{R}(x)f](y) = f(yx).$$

It is a quick exercise to check that these are indeed representations. In fact, note that for each $x \in K$

$$\|\mathcal{R}(x)f\|_{L^2(K)} = \int_K |f(yx)|^2\, dy = \int_K |f(y)|^2\, dy = \|f\|_{L^2(K)}$$

holds for $f \in C(K)$ and thus in general by standard density theorems; thus $\mathcal{R}(x) \in \mathrm{U}(L^2(G))$. A similar argument applies to the left regular representation – they are both unitary representations.

Now, for a given finite dimensional representation (V, Π), with a given basis for V, the matrix-entry functions Π_{jk} are smooth, and hence since K is compact they are in $L^2(K)$. So in the language of Definition 18.3, the above discussion shows that the space of matrix entry functions $\{\Pi_{jk}\colon 1 \le j, k \le \dim(V)\}$ is an invariant subspace of the representation $(L^2(K), \mathcal{R})$, all of whose elements are continuous functions.

DEFINITION 18.4. *A continuous function $f \in L^2(K)$ is called a* **representative function** *if there is a finite dimensional invariant subspace of $(L^2(K) \cap C(K), \mathcal{R})$ containing f.*

It might make sense to call such functions *right* representative functions, as the same notion for the left-regular representation makes good sense. We will see below that the two coincide.

In general, one would expect the orbit of a random function $f \in L^2(K)$ under the action of all $\mathcal{R}(x)$ for $x \in K$ to be infinite-dimensional. On the other hand, any linear combination of representative functions is immediately seen to be a representative function. In fact, up to linear combinations over potentially different representations, all representative functions have this form.

PROPOSITION 18.5. *Any representative function of K is a linear combination of matrix entries of irreducible representations of K.*

PROOF. Let f be a representative function. By definition, there is a finite dimensional subspace $V \subset L^2(K) \cap C(K)$ containing f that is invariant under the right regular representation $\mathcal{R}$. In other words, $(V, \mathcal{R}|_V)$ is a finite dimensional representation of K. It is therefore completely reducible

(cf. Corollary 17.34), so $V = \bigoplus_j V_j$ is a finite direct sum of irreducible invariant subspaces for $\mathcal{R}$; thus $(V_j, \mathcal{R}|_{V_j})$ are irreps of K.

Now, the evaluation functional $V_j \to \mathbb{C}$ given by $g \mapsto g(e)$ is a well-defined linear functional on the finite dimensional space V_j of continuous functions. It is therefore continuous. Hence, by the Riesz-Fisher theorem, there is a vector $\xi_j \in V_j$ such that

$$g(e) = \langle \xi_j, g \rangle = \int \overline{\xi_j(x)} g(x)\, dx.$$

But then we have, for each $g \in V_j$ and $x \in K$,

$$g(x) = [\mathcal{R}(x)g](e) = \langle \xi_j, \mathcal{R}(x)g \rangle. \tag{18.2}$$

In particular, since V_j is irreducible (and therefore not trivial), the vector ξ_j is not 0. Let $\hat{\xi}_j = \xi_j / \langle \xi_j, \xi_j \rangle^{1/2}$. We can therefore select an orthonormal basis $\{h_k \colon 1 \le k \le \dim(V_j)\}$ for V_j with $h_1 = \hat{\xi}_j$. Expanding g in this basis $g = \sum_k a_k h_k$, we therefore have

$$g(x) = \sum_k a_k \langle \xi_j, \mathcal{R}(x) h_k \rangle = \sum_k a_k \langle \xi_j, \xi_j \rangle^{1/2} \langle h_1, \mathcal{R}(x) h_k \rangle.$$

This shows that every $g \in V_j$ is a linear combination of the matrix entries $[\mathcal{R}|_{V_j}]_{1k}$ of an irrep. Since $f \in V = \bigoplus_j V_j$ is a sum of such g_j, it too is a linear combination of matrix entries of irreps. □

COROLLARY 18.6. *Any finite dimensional invariant subspace of $(L^2(K), \mathcal{R})$ is also invariant for $\mathcal{L}$; hence, left and right representative functions coincide.*

PROOF. By Proposition 18.5, if V is a finite dimensional invariant subspace for $\mathcal{R}$, then there is a finite collection of irreducible representations whose matrix entries span V. A calculation akin to (18.1) shows that the space of matrix entries of any representation is also invariant for $\mathcal{L}$, and so V is as well. □

Now, let us focus on a very important class of representative functions.

DEFINITION 18.7. *For a given finite dimensional representation (V, Π) of K, let $\chi = \chi_V \colon K \to \mathbb{C}$ be the function $\chi = \Pi_{11} + \Pi_{22} + \cdots + \Pi_{dd}$ where $d = \dim(V)$; that is: $\xi(x) = \operatorname{Tr}(\Pi(x))$. The function ξ is called the* **character** *of (V, Π). It is a linear combination of matrix entries, and so it is a representative function. Note that it is also a class function:*

$$\chi(yxy^{-1}) = \operatorname{Tr}[\Pi(yxy^{-1})] = \operatorname{Tr}[\Pi(y)\Pi(x)\Pi(y)^{-1}] = \operatorname{Tr}[\Pi(x)] = \chi(x).$$

If (V, Π) is an irreducible representation, its character is called an **irreducible character**.

These characters (especially irreducible ones) play a very important role in harmonic analysis on compact Lie groups, as we will see a bit later. The following examples gives an indication why.

EXAMPLE 18.8. Consider the special case $K = \mathbb{S}^1$, and look at all irreducible representations. Since $\mathbb{S}^1$ is abelian, by Schur's lemma any complex irrep Π is one-dimensional, so we may take the representation space to be $\mathbb{C}$, and $\Pi \colon \mathbb{S}^1 \to \mathrm{GL}(\mathbb{C})$; the trace is the identity map on $\mathrm{GL}(\mathbb{C})$, and so the irreducible characters are just the irreps themselves. Moreover, since $\mathbb{S}^1$ is compact, wlog Π is a unitary representation, which in this case simply means $\Pi \colon \mathbb{S}^1 \to \mathbb{S}^1$. A homework exercise showed that the only such group homomorphisms are of the form $\Pi(u) = u^m$ for some $m \in \mathbb{Z}$. Thus, the irreducible characters of $\mathbb{S}^1$ are precisely all the functions $\chi(e^{i\theta}) = e^{im\theta}$. These are precisely the standard Fourier basis for the circle.

In a sense, much of this chapter is devoted to generalizing the basic ideas of Fourier series to general compact Lie groups. The following lemma gets this started. First, let (V, Π) be a representation of K. We denote by V^K the subspace of V that is (pointwise) invariant under the action of V (via Π):

$$V^K \equiv \{v \in V \colon \Pi(x)v = v \text{ for all } x \in K\}.$$

LEMMA 18.9. *Let (V, Π) be a finite dimensional representation of K, and let $P \colon V \to V$ be the operator given by*

$$P = \int_K \Pi(x)\, dx; \quad \textit{i.e. } P(v) = \int_K \Pi(x)v\, dx.$$

Fix an inner product on V with respect to which Π is a uniary representation. Then P is the orthogonal projection onto V^K.

PROOF. Fix $y \in K$ and $v \in V$; then

$$\Pi(y)Pv = \Pi(y)\int_K \Pi(x)v\, dx = \int_K \Pi(yx)v\, dx = \int_K \Pi(z)v\, dz = Pv$$

by the left-invariance of the Haar measure. Hence, $Pv \in V^K$. On the other hand, if $v \in V^K$, then

$$Pv = \int_K \Pi(x)v\, dx = \int_K v\, dx = v$$

and so P maps V onto V^K and restricts to the identity on V^K. It is therefore a projection onto V^K. What's more, for $v, w \in V$,

$$\langle Pv, w\rangle = \int_K \langle \Pi(x)v, w\rangle\, dx = \int_K \langle v, \Pi(x^{-1})w\rangle\, dx$$

because the representation Π is unitary. The bi-invariant Haar measure is also invariant under the inversion map $x \mapsto x^{-1}$ (this follows anaologously to the proof of Proposition 18.1), and so

$$\langle Pv, w\rangle = \int_K \langle v, \Pi(z)w\rangle\, dz = \langle v, Pw\rangle.$$

Hence $P = P^*$. This identifies P as the orthogonal projection onto V^K. □

Next, we will state a few important properties of characters. This will require a few new notions of how to put different representations together.

DEFINITION 18.10. *Let (V, Π) and (W, Σ) be representations of K. The* **tensor product representation** *$(V \otimes W, \Pi \otimes \Sigma)$ determined by*

$$\Pi \otimes \Sigma \colon K \to \mathrm{GL}(V \otimes W), \quad [(\Pi \otimes \Sigma)(x)](v \otimes w) = \Pi(x)v \otimes \Sigma(x)w$$

for $v \in V$ and $w \in W$. In other words: as linear operators, $(\Pi \otimes \Sigma)(x) = \Pi(x) \otimes \Sigma(x)$. It is a simple matter to check that it is, indeed, a representation.

REMARK 18.11. The tensor product representation also makes sense for representations of different groups G and H: we then have $\Pi \otimes \Sigma$ is a representation of $G \times H$ given by $(\Pi \otimes \Sigma)(x, y) = \Pi(x) \otimes \Sigma(y)$. In the applications at hand, we only need the "diagonal" case of this where $G = H = K$.

DEFINITION 18.12. *Let* (V, Π) *be a representation of* K. *The* **dual** *or* **contragredient representation** (V^*, Π^*) *is defined by*

$$\Pi^* \colon K \to \mathrm{GL}(V^*), \quad [\Pi^*(x)\lambda](v) = \lambda(\Pi(x^{-1})v)$$

for $v \in V$ *and* $\lambda \in V^*$. *It is a simple matter to check that it is, indeed, a representation (due to the transpose and inverse together).*

Using standard abuse of notation, we may refer to the tensor product of the representations simply as $V \otimes W$, and contragredient simply as V^* (just as we refer to the direct sum representation simply as $V \oplus W$).

Following are some basic computationally-useful properties of characters.

PROPOSITION 18.13. *Let* (V, Π) *and* (W, Σ) *be representations of* K, *and let* χ_V *and* χ_W *denote their characters.*

(1) χ_V *is a* C^∞ *class functions.*
(2) *If* $(V, \Pi) \cong (W, \Sigma)$, *then* $\chi_V = \chi_W$.
(3) $\chi_{V \oplus W} = \chi_V + \chi_W$.
(4) $\chi_{V \otimes W} = \chi_V \cdot \chi_W$.
(5) $\chi_{V^*}(x) = \chi_V(x^{-1})$, *for* $x \in G$.
(6) $\chi_V(e) = \dim(V)$.

PROOF. Item (1) follows immediately from Definition 18.7, and item (2) is a simple calculation using the trace property. Items (3)-(5) are elementary calculations untwisting the definitions; we give the proof of (4) here and leave the others to the reader. Fix bases $\{v_j\}$ for V and $\{w_k\}$ for W; then $\{v_j \otimes w_k\}$ is a basis for $V \otimes W$. For any linear operators A on V and B on W with matrices $[A]_{j_1 j_2}$ and $[B]_{k_1 k_2}$ in the given bases, it is easy to check that the matrix of $A \otimes B$ in the tensor product basis is $[A \otimes B]_{(j_1,k_1)(j_2,k_2)} = A_{j_1 j_2} B_{k_1 k_2}$. Thus

$$\mathrm{Tr}\,(A \otimes B) = \sum_{j,k} [A \otimes B]_{(j,k)(j,k)} = \sum_{j,k} A_{jj} B_{kk} = \sum_j A_{jj} \cdot \sum_k B_{kk} = \mathrm{Tr}\,(A)\,\mathrm{Tr}\,(B).$$

In particular, for $x \in K$, taking $A = \Pi(x)$ and $B = \Sigma(x)$,

$$\chi_{V \otimes W}(x) = \mathrm{Tr}\,[(\Pi \otimes W)(x)] = \mathrm{Tr}\,[\Pi(x) \otimes \Sigma(x)] = \mathrm{Tr}\,[\Pi(x)]\,\mathrm{Tr}\,[\Sigma(x)] = \chi_V(x)\chi_W(x).$$

Finally, for imem (6), $\chi_V(e) = \mathrm{Tr}\,[\Pi(e)] = \mathrm{Tr}\,(\mathrm{Id}_V) = \dim(V)$. □

This brings us to the orthogonality relations. Consider, again, Example 18.8, where we saw that the characters of irreps of S^1 are the functions $\chi_m(e^{i\theta}) = e^{im\theta}$ for $m \in \mathbb{Z}$. These are the standard Fourier basis for the circle: with respect to the uniform probability measure on $\mathbb{S}^1$ (which is, of course, the Haar measure), they form an orthonormal basis. This turns out to be true for irreducible characters on all compact groups. The easy half (that they form an orthonormal set) is the content of the next theorem.

THEOREM 18.14. *The irreducible characters of* K *are an orthonormal set in* $L^2(K)$. *More precisely: if* (V, Π) *and* (W, Σ) *are two irreps of* K, *then*

$$\int_K \overline{\chi_V(x)} \chi_W(x)\, dx = \begin{cases} 1 & \text{if } V \cong W, \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. Fix an inner product on V with respect to which (V, Π) is a unitary representation. Now, note that for $x \in K$

$$\overline{\chi_V(x)} = \overline{\operatorname{Tr}[\Pi(x)]} = \operatorname{Tr}[\Pi(x)^*] = \operatorname{Tr}[\Pi(x^{-1})] = \chi_V(x^{-1}) = \chi_{V^*}(x)$$

where the second equality is an easy fact about traces, the second is due to the unitarity of Π, and the third is Proposition 18.13(5). Hence, using Proposition 18.13(4),

$$\overline{\chi_V(x)}\chi_W(x) = \chi_{V^*}(x)\chi_W(x) = \chi_{V^*\otimes W}(x), \quad x \in K.$$

Thus

$$\int_K \overline{\chi_V(x)}\chi_W(x)\,dx = \int_K \chi_{V^*\otimes W}(x)\,dx = \int_K \operatorname{Tr}[(\Pi^*\otimes\Sigma)(x)]\,dx = \operatorname{Tr}\left(\int_K (\Pi^*\otimes\Sigma)(x)\,dx\right).$$

This integral is the orthogonal projection onto the space $(V^*\otimes W)^K$ fixed by $\Pi^*\otimes\Sigma$, cf. Lemma 18.9, and thus its trace is the dimension of the space:

$$\int_K \overline{\chi_V(x)}\chi_W(x)\,dx = \dim((V^*\otimes W)^K). \tag{18.3}$$

Now, we use the canonical linear isomorphism $\vartheta\colon V^*\otimes W \to \operatorname{End}(V, W)$, where $\vartheta(\lambda\otimes w)(v) = \lambda(v)w$. Let us now define a representation $(\operatorname{End}(V, W), \Phi)$ of K by

$$\Phi(x)A = \Sigma(x)A\Pi(x)^{-1}.$$

Then ϑ intertwines $\Pi^*\otimes\Sigma$ and Φ:

$$\vartheta[(\Pi^*\otimes\Sigma)(x)(\lambda\otimes w)](v) = \vartheta[\Pi^*(x)\lambda\otimes\Sigma(x)w](v) = (\Pi(x)^*\lambda)(v)\Sigma(x)w = \lambda(\Pi(x^{-1})v)\Sigma(x)w$$

while

$$[\Phi(x)\vartheta(\lambda\otimes w)](v) = \Sigma(x)\varphi(\lambda\otimes w)\Pi(x)^{-1}(v) = \Sigma(x)\lambda(\Pi(x)^{-1}v)w = \lambda(\Pi(x^{-1})v)\Sigma(x)w.$$

So the representation $(V^*\otimes W, \Pi^*\otimes\Sigma)$ is isomorphic to $(\operatorname{End}(V, W), \Phi)$. Under this isomorphism, $(V^*\otimes W)^K$ is therefore mapped to the set of all endomorphisms A with the property that $\Sigma(x)A\Pi(x)^{-1} = \Phi(x)A = A$ for all $x \in K$, meaning that A is an intertwining map from (V, Π) to (W, Σ). Since both representations are irreducible, by Schur's lemma any two intertwining maps are scalar multiples of each other, so the image of the space $(V^*\otimes W)^K$ has dimension 0 or 1. If $(V, \Pi) \cong (W, \Sigma)$ then there is an isomorphism which is an intertwining map, so the dimension is 1; if they are not isomorphic, then by Schur's lemma all intertwining maps are 0, hence the dimension is 0. The theorem now follows from (18.3). □

Let us conclude this section by noting one obvious consequence of the orthogonality relations: if two irreps are non-isomorphic, they have orthogonal, and hence different, characters.

COROLLARY 18.15. *The isomorphism class of an irreducible representation is completely determined by its character.*

2. Group Convolution

DEFINITION 18.16. *If $f, g \in L^2(K)$, their* **convolution** *$f * g$ is a new function on K defined by*

$$(f * g)(x) = \int_K f(xy^{-1})g(y)\,dy.$$

Since the Haar measure is bi-invariant (and invariant under inversion), the function $\check{f}\colon y \mapsto f(xy^{-1})$ is also in $L^2(K)$ with norm $\|\check{f}\|_{L^2(K)} = \|f\|_{L^2(K)}$; thus the product $\check{f} \cdot g$ is in $L^1(G)$: by the Cauchy-Schwarz inequality, it satisfies

$$\int_K |f(xy^{-1})g(y)|\,dy = \int_K |\check{f}(y)g(y)|\,dy \leq \|f\|_{L^2(K)}\|h\|_{L^2(K)}$$

which shows that $(f * g)(x)$ is well defined and satisfies the pointwise bound

$$|(f * g)(x)| \leq \|f\|_{L^2(K)}\|g\|_{L^2(K)}, \quad x \in K. \tag{18.4}$$

Hence, for all $f, g \in L^2(K)$, $f * g \in L^\infty(K)$. By squaring both sides of (18.4) and integrating, we see that also

$$\|f * g\|_{L^2(K)} \leq \|f\|_{L^2(K)}\|g\|_{L^2(K)} \tag{18.5}$$

which shows that $f * g \in L^2(K)$ (although we already knew this since the measure is finite, so $L^\infty(K) \subset L^2(K)$). In fact, convolution is a "smoothing" operation: the convolution of L^2 functions is, in fact, continuous.

PROPOSITION 18.17. *If $f, g \in L^2(K)$, then $f * g \in L^2(K) \cap C(K)$.*

PROOF. First, assume that $f, g \in C(K)$ to start. Since K is compact, f, g are uniformly continuous and bounded. Fix $x \in K$ and let $x_n \to x$. Then

$$\begin{aligned}|(f * g)(x_n) - (f * g)(x)| &\leq \int_K |f(x_n y^{-1}) - f(xy^{-1})||g(y)|\,dy \\ &\leq \|g\|_{L^\infty(K)} \int_K |f(x_n y^{-1}) - f(xy^{-1})|\,dy.\end{aligned}$$

Since f is uniformly continuous, $f(x_n y^{-1}) \to f(xy^{-1})$ uniformly in y, and hence the above integral converges to 0. Thus $f * g$ is continuous.

Now for general $f, g \in L^2(K)$. The space $C(K)$ of continuous functions is dense in $L^2(K)$ (generically true in measure theory), so let (f_n) and (g_m) be continuous functions on K with $f_n \to f$ and $g_m \to g$ in $L^2(K)$. Then note that, for $x \in K$ and $m \in \mathbb{N}$,

$$|(f_n * g_m)(x) - (f * g_m)(x)| = |((f_n - f) * g_m)(x)| \leq \|f - f_n\|_{L^2(K)}\|g_m\|_{L^2(K)}$$

by (18.5). The quantity on the right tends to 0 and does not depend on x, so $f_n * g_m$ converges to $f * g_m$ uniformly. By the above argument, $f_n * g_m$ is continuous, and hence $f * g_m$ is a uniform limit of continuous functions, which is therefore continuous. An analogous argument now exhibits $f * g$ as the uniform limit of these continuous functions, proving that it is continuous. □

We now consider some invariance properties of the convolution operation, in particular with regard to the left and right regular representations of K.

LEMMA 18.18. *Convolution on the left commutes with right translation, and convolution on the right commutes with left translation. That is: for $f, g \in L^2(K)$ and $x \in K$,*

$$(\mathcal{L}(x)f) * g = \mathcal{L}(x)(f * g), \quad \textit{and} \quad f * (\mathcal{R}(x)g) = \mathcal{R}(x)(f * g).$$

PROOF. We prove only the statement for $\mathcal{R}$, leaving the similar calculation for $\mathcal{L}$ to the reader. For $z \in K$,

$$[f * (\mathcal{R}(x)g)](z) = \int f(zy^{-1})(\mathcal{R}(x)g)(y)\,dy = \int f(zy^{-1})g(yx)\,dy.$$

Now make the substitution $u = yx$; the right-invariance of the Haar measure means that $du = dy$, and since $y = ux^{-1}$ we have $y^{-1} = xu^{-1}$ and the above integral equals

$$\int f(zxu^{-1})g(u)\,du = (f * g)(zx) = [\mathcal{R}(x)(f * g)](z).$$

□

On $\mathbb{R}^n$, convolution is commutative. That is not generally true for group convolution (on non-abelian groups). However, it remains true if one of the convolved functions is a class function.

PROPOSITION 18.19. *Let $f \in L^2(K)$ be a class function. Then for all $g \in L^2(K)$, $f*g = g*f$.*

PROOF. The trick is to make the change of variables $z = y^{-1}x$ for fixed x. Then $y = xz^{-1}$ and $y^{-1} = zx^{-1}$, and by the invariance of the Haar measure $dy = dz$. Thus

$$(f * g)(x) = \int_K f(xy^{-1})g(y)\,dy = \int_K f(xzx^{-1})g(xz^{-1})\,dz = \int_K f(z)g(xz^{-1})\,dz$$

where we used the fact that f is a class function in the last equality. The last integral is, by definition, $(g * f)(x)$. □

We will now use convolution to construct an *approximate identity*: a sequene of smooth functions that act in the (L^2-)limit like the delta function. The idea, as in $\mathbb{R}^n$, is to construct non-negative bump functions φ_n suppored in small neighborhoods of the identity (shrinking as n grows), and normalized to each have area 1; then $f * \varphi_n$ will converge to f in $L^2(K)$. In fact, we can do this with class functions φ_n. To understand what is meant by "small neighborhoods", we will need a bi-invariant Riemannian metric, cf. Section 14.6.

LEMMA 18.20. *There exists a bi-invariant Riemannian metric g on K. The resultant Riemannian distant function $\mathrm{d_g}$ is also bi-invariant.*

It is important to note that the compactness of K is needed here, or more precisely the existence of an Ad-invariant inner product on its Lie algebra.

PROOF. Fix an $\mathrm{Ad}(K)$-invariant inner product on $\mathfrak{k} = \mathrm{Lie}(K)$, and (as in Definition 14.30) extend it to a Riemannian metric on K by defining, for any $x \in K$ and $X_x, Y_x \in T_xK$

$$\mathrm{g}_x(X_x, Y_x) = \langle dL_{x^{-1}}|_x(X_x), dL_{x^{-1}}|_x(Y_x)\rangle.$$

I.e. translate X_x, Y_x to vectors at e and take the inner product there. Since $L_{x^{-1}}$ is a diffeomorphism, this map is smooth: i.e. is $X, Y \in \mathscr{X}(K)$ then $x \mapsto \mathrm{g}_x(X_x, Y_x)$ is a $C^\infty(K)$ function. It is evidently positive definite on each tangent space, so it is a Riemannian metric. We need to check that it is bi-invariant. It is more or less obviously left-invariant. On the other side, for $x, y \in K$,

$$\begin{aligned}(R_y)^*(\mathrm{g})(X, Y)(x) &= \mathrm{g}_{xy}(dR_y|_x(X_x), dR_y|_x(Y_x)) \\ &= \langle dL_{(xy)^{-1}}|_{xy} \circ dR_y|_x(X_x), dL_{(xy)^{-1}}|_{xy} \circ dR_y|_x(Y_x)\rangle.\end{aligned}$$

Now $dL_{(xy)^{-1}} \circ dR_y = d(L_{(xy)^{-1}} \circ R_y) = d(L_{y^{-1}} \circ L_{x^{-1}} \circ R_y) = d(C_{y^{-1}} \circ L_{x^{-1}}) = \mathrm{Ad}(y^{-1}) \circ dL_{x^{-1}}$. Thus

$$\begin{aligned}(R_y)^*(\mathrm{g})(X,Y)(x) &= \langle \mathrm{Ad}(y^{-1}) \circ dL_{x^{-1}}|_x(X_x), \mathrm{Ad}(y^{-1}) \circ dL_{x^{-1}}|_x(Y_x)\rangle \\ &= \langle dL_{x^{-1}}|_x(X_x), dL_{x^{-1}}|_x(Y_x)\rangle \\ &= \mathrm{g}(X,Y)(x)\end{aligned}$$

where the second equality is the Ad-invariance of the inner product.

The fact that the resultant Riemannian distance is bi-invariant follows exactly as the left-invariant version of this statement does in Proposition 14.31. □

Thus, we fix a bi-invariant Riemannian metric g on K. We may then talk about balls in K: $B_\epsilon(e) = \{x \in K \colon d_{\mathrm{g}}(e,x) < \epsilon\}$ is the Riemannian ball of radius $\epsilon > 0$ centered at the identity $e \in K$. As the Riemannian distance function is left invariant, this ball is symmetric, meaning that $B_\epsilon(e) = B_\epsilon(e)^{-1}$; indeed, $d(e,x^{-1}) = d(x,e) = d(e,x)$ for $x \in K$ by left-invariance. A similar argument shows that $y^{-1}(B_\epsilon(e))y = B_\epsilon(e)$ for each $y \in K$.

THEOREM 18.21. *There exists a sequence of nonnegative C^∞ class function $(\varphi_n)_{n\in\mathbb{N}}$ on K with the following properties.*

(1) *supp* $(\varphi_n) \subseteq B_{1/n}(e)$,
(2) $\varphi_n(x^{-1}) = \varphi_n(x)$ *for all* $x \in K$, *and*
(3) $\int_K \varphi_n(x)\,dx = 1$.

For any such sequence,

$$\lim_{n\to\infty} \|f * \varphi_n - f\|_{L^2(K)} = 0$$

for all $f \in L^2(K)$. If f is continuous, then additionally $f * \varphi_n \to f$ uniformly.

PROOF. Fix a C^∞ bump function ψ_n that is supported in $B_{1/n}(e)$. Define

$$\xi_n(x) = \int_K \psi_n(yxy^{-1})\,dy.$$

Then ξ_n is a nonnegative C^∞ class function, and it is still supported in $B_{1/n}(e)$: if $x \notin B_{1/n}(e)$ then $yxy^{-1} \notin B_{1/n}(e)$ for all $y \in K$ (since $B_{1/n}(e)$ is invariant under conjugation) and so the integrant of $\xi_n(x)$ is the 0 function. Now, define

$$\check{\varphi}_n(x) = \frac{1}{2}(\xi_n(x) + \xi_n(x^{-1})).$$

Again, $\check{\varphi}_n$ is supported in $B_{1/n}(e)$ since the ball is symmetric, and it manifestly satisfies properties (1) and (2). It is a $C^\infty(K)$ function that is positive on some neighborhood (as ψ_n is), and so its integral is positive. Thus, if we define

$$\varphi_n = \frac{\check{\varphi}_n}{\int_K \check{\varphi}_n(x)\,dx},$$

these functions have properties (1)-(3), as desired.

Now, fix any such sequence φ_n, and let $g \in C(K)$. Since K is compact, f is uniformly continuous. Fix $\epsilon > 0$; then for all sufficiently large $n \in \mathbb{N}$ $|g(x) - g(y)| < \epsilon$ for $y \in B_{1/n}(x)$. Then we have

$$(\varphi_n * g)(x) - g(x) = \int_K \varphi_n(xy^{-1})g(y)\,dy - g(x) = \int_K [\varphi_n(xy^{-1})g(y) - g(x)]\,dy$$

where the last equality just comes from the fact that the Haar measure is normalized. Thus

$$|(\varphi_n * g)(x) - g(x)| \leq \int_K \varphi_n(xy^{-1})|g(y) - g(x)|\, dy.$$

The integrand is 0 whenever xy^{-1} is outside the support of φ_n, which is contained in $B_{1/n}(e)$; hence, the integral is really over those y for which $xy^{-1} \in B_{1/n}(e)$; right invariance of the distance functions shows that this is precisely the set $y \in B_{1/n}(x)$. Thus, for n sufficiently large that $|g(y) - g(x)| < \epsilon$ when $y \in B_{1/n}(x)$, we have

$$|(\varphi_n * g)(x) - g(x)| < \epsilon \int_K \varphi_n(xy^{-1})\, dy = \epsilon \int_K \varphi_n(y^{-1})\, dy = \epsilon.$$

We have thus shown that, for all sufficiently large n, $|(\varphi_n * g)(x) - g(x)| < \epsilon$ for all x. That is: $\varphi_n * g$ converges uniformly to g, proving the second claim. Since uniform convergence is stronger than L^2-convergence, it follows that $\varphi_n * g$ converges to g in $L^2(K)$ as well, in the case that $g \in C(K)$. The extension from this to all $L^2(K)$ is a standard approximation argument that is left to the reader. □

Hence, we can approximate (uniformly or in L^2) any (continuous) function f by convolutions $\varphi * f$. In typical analysis arguments, the benefit of this approach is that $\varphi * f$ is nicer than f: for example, since φ is smooth, so is $\varphi * f$, regardless of how rough f is. In the present situation, we are motivated by a different notion of "niceness", stemming from our desire to approximate any function by representative functions. The key here is to realize that the linear operator

$$\mathcal{C}_\varphi \colon f \mapsto \varphi * f$$

is the closest kind of object there is to an "infinite dimensional matrix". Indeed, we have

$$\mathcal{C}_\varphi(f)(x) = \int_K \varphi(xy^{-1})f(y)\, dy = \int_K \kappa(x,y)f(y)\, dy$$

where $\kappa \colon K \times K \to \mathbb{C}$ is a "continuous matrix", better known as an integral kernel. In this case, since $\varphi(x) = \varphi(x^{-1})$ (Theorem 18.21(2)), we have

$$\kappa(x,y) = \varphi(xy^{-1}) = \varphi((xy^{-1})^{-1}) = \varphi(yx^{-1}) = \kappa(y,x)$$

so κ is a "symmetric continuous matrix". As this κ is real-valued, this is equivalent to "Hermitian continuous matrix". Finally, note that

$$\int_{K\times K} |\kappa(x,y)|^2\, dx\, dy = \int_K dy \int_K dx\, \varphi(xy^{-1})^2 = \int_K dy \int_K dx\, \varphi(x)^2 = \int_K dy(1) = 1$$

and so κ is a square-integrable function on $K \times K$. These properties qualify the integral operator $\mathcal{C}_\varphi$, whose integral kernel is κ, as a Hermitian *Hilbert-Schmidt operator* – a special example of a Hermitian *compact operator*. Compact operators are the closest infinite dimensional operators to matrices. There is a spectral theorem for them that resembles the finite dimensional one very closely. We state it here with no proof; the reader who has not seen these ideas is referred to the excellent treatment in [**1**].

THEOREM 18.22 (The Spectral Theorem for Compact Operators). *There is a (Hilbert space) orthonormal basis $\{e_n\}_{n=1}^\infty$ of $L^2(K)$ consisting of eigenvectors for $\mathcal{C}_\varphi$. The corresponding sequence of eigenvalues $\{\lambda_n\}_{n=1}^\infty$ are all real, and $\lambda_n \to 0$ as $n \to \infty$.*

COROLLARY 18.23. *Let $\lambda \neq 0$ be a nonzero eigenvalue of $\mathcal{C}_\varphi$. Its eigenspace $E_\lambda \subset L^2(K)$ is a finite-dimensional invariant subspace of $(L^2(K), \mathcal{R})$.*

PROOF. As the sequence of eigenvalues converge to 0, $|\lambda_n| < |\lambda|$ for all sufficiently large n. Hence, there are only finitely-many n for which $\lambda = \lambda_n$, so there are only finitely-many n for which e_n is an eigenvector of λ. As the e_n form an orthonormal basis for $L^2(K)$, any eigenvector of λ is a linear combination of these finitely-many e_n, and thus the eigenspace E_λ is finite dimensional.

Now, let $f \in L^2(K)$, and let $x \in K$. Then

$$\mathcal{C}_\varphi(\mathcal{R}(x)f) = \varphi * (\mathcal{R}(x)f) = \mathcal{R}(x)(\varphi * f) = \mathcal{R}(x)C_\varphi f$$

by Lemma 18.18. In particular, if $f \in E_\lambda$, then

$$\mathcal{C}_\varphi(\mathcal{R}(x)f) = \mathcal{R}(x)C_\varphi f = \mathcal{R}(x)\lambda f = \lambda\mathcal{R}(x)f$$

showing that $\mathcal{R}(x)f \in E_\lambda$. Thus, E_λ is an invariant subspace. □

3. The Peter-Weyl Theorem

We now state and prove the central result of this chapter: the Peter-Weyl theorem, published by Herman Weyl and his PhD student Fritz Peter in 1927. We will state it in two parts, and to understand the second part we first need a bit more notation, given in the following two lemmas.

LEMMA 18.24. *Let $L^2(K)_{\text{class}}$ denote the subspace of class function in $L^2(K)$. Then $L^2(K)_{\text{class}}$ is a Hilbert subspace.*

PROOF. A function $f \in L^2(K)$ is a class function iff $f(u^{-1}xu) = f(x)$ for all $x, u \in K$. We can write this in the form $[\mathcal{L}(u)\mathcal{R}(u)]f(x) = f(x)$, so

$$L^2(K)_{\text{class}} = \bigcap_{u\in K} \ker(\mathcal{L}(u)\mathcal{R}(u) - I).$$

Now $\mathcal{L}(u)$ and $\mathcal{R}(u)$ are unitary operators, and therefore so is their composition. Thus $\mathcal{L}(u)\mathcal{R}(u) - I$ is a bounded operator, and its kernel is therefore a closed subspace. An intersection of closed subspaces is closed, and hence $L^2(K)_{\text{class}}$ is a closed subspace of a Hilbert space, therefore it is a Hilbert subspace. □

LEMMA 18.25. *Let $C(K)_{\text{class}}$ denote the subspace of class function in $C(K)$. Then $C(K)_{\text{class}}$ is a closed subspace of $C(K)$ in the uniform topology.*

PROOF. It is easy to check that $C(K)_{\text{class}}$ is a subspace. Let (f_n) be a sequence in this subspace, and suppose $f \in C(K)$ is its uniform limit. Then for any $x, u \in K$,

$$|f(x) - f(uxu^{-1})| \le |f(x) - f_n(x)| + |f_n(x) - f_n(uxu^{-1})| + |f_n(uxu^{-1}) - f(uxu^{-1})|.$$

The middle terms is equal to 0 since $f_n \in C(K)_{\text{class}}$. The first and third terms are both less than $\|f - f_n\|_\infty$ which tends to 0. Hence $|f(x) - f(uxu^{-1})| = 0$. □

And now, the main event.

THEOREM 18.26 (Peter-Weyl). *Let K be a compact Lie group.*

(1) *The space of representative functions on K is dense in both $C(K)$ and $L^2(K)$.*
(2) *The linear span of irreducible characters of K is dense in both $C(K)_{\text{class}}$ and $L^2(K)_{\text{class}}$. In particular, the characters of all pairwise non-isomorphic irreducible representations form an orthonormal basis for $L^2(K)_{\text{class}}$.*

REMARK 18.27. In fact, Peter and Weyl proved this theorem, as stated, for the much larger class of compact topological groups (i.e. groups with a Hausdorff compact topology making the group operations continuous).

Since $C(K)$ is dense in $L^2(K)$ (and ergo the same is true for the subspaces of class functions), it would be enough to prove the uniform density results for continuous functions. In fact, we will go the other direction, proving first density in $L^2(K)$ and then (with the help of the Weierstrass approximation theorem) deducing uniform density in $C(K)$.

For ease of reading, we prove each of the four statements (two in each of (1) and (2)) as separate propositions.

PROPOSITION 18.28. *The representative functions are dense in $L^2(K)$.*

PROOF. Suppose, to the contrary, that $g \in L^2(K)$ is not in the closed linear span of the representative functions; then it is in the orthogonal complement of this closed subspace, and hence g is orthogonal to every representative function f. Now, let (φ_n) be an approximate identity sequence as in Theorem 18.21; then $\varphi_n * g \to g$ in $L^2(K)$ by that theorem. Now, by Corollary 18.23, the eigenspaces of nonzero eigenvalues of $\mathcal{C}_{\varphi_n}$ are finite-dimensional invariant subspaces of $(L^2(K), \mathcal{R})$, meaning that these eigenvectors are representative functions. Hence, g is orthogonal to all eigenvectors of C_{φ_n} with nonzero eigenvalues. By Theorem 18.22 (the Spectral Theorem), the eigenvectors of C_{φ_n} form an orthonormal basis for $L^2(K)$, and hence g is in the closed linear span of the eigenvectors with eigenvalue 0. This means precisely that $\varphi_n * g = C_{\varphi_n} g = 0$, for all n. Thus $g = 0$ (which was, of course, in the closed linear span of the representative functions). This contradiction proves the proposition. □

COROLLARY 18.29. *The space of representative functions separates points: if $x \neq y \in K$, there is a representative function f with $f(x) \neq f(y)$.*

PROOF. Let β be any continuous function with $\beta(x) = 1$ and $\beta(y) = 0$ (for example, take a smooth bump function centered at x whose support does not contain y). Then β is in $L^2(K)$, and by Proposition 18.28 there is a sequence f_n of representative functions with $f_n \to \beta$ in $L^2(K)$. We can then choose a subsequence f_{n_k} that converges pointwise to β. (This is a standard result in measure theory for L^p spaces: convergence in L^p norm implies convergence almost everywhere for a subsequence, and since all the involved functions are continuous, this implies pointwise convergence.) Hence, $f_{n_k}(x) \to \beta(x) = 1$ and $f_{n_k}(y) \to \beta(y) = 0$, and hence for all large k $f_{n_k}(x) \neq f_{n_k}(y)$. □

REMARK 18.30. Since all matrix entries are representative functions, an immediate corollary is that the set of all irreducible representations separates points: if $x \neq y$ then there is an irrep Π with $\Pi(x) \neq \Pi(y)$. Otherwise all matrix entries would agree on x and y, and hence also all linear combinations of them; Proposition 18.5 would then show that all representative functions agree on x and y.

PROPOSITION 18.31. *The representative functions are dense in $C(K)$.*

PROOF. The set of representative functions is, by Proposition 18.5, the same as the set of linear combinations of matrix entries of representations of K. This set is therefore a complex unital $*$-algebra:

- It is a $\mathbb{C}$-vector space by definition.
- It is closed under product: the product of two matrix entries is a matrix entry for the tensor product of the two representations.

- It is closed under complex conjugation: for any representation Π, $\overline{\Pi_{jk}} = \Pi^*_{kj}$ is a matrix entry of the contragredient representation.

By Corollary 18.29, this algebra separates points of K. Since K is a compact Hausdorff space, it follows from the Stone-Weierstrass theorem that the algebra of representative functions is dense in $C(K)$. □

Now moving to the second item of Theorem 18.26, we need the following lemma.

LEMMA 18.32. *If g is a representative function, then*

$$h(x) = \int_K g(uxu^{-1})\,du$$

is a linear combination of irreducible characters.

PROOF. Since any representative function is a linear combination of matrix entries of irreps (by Proposition 18.5), by linearity of the integral it suffices to prove the statement for matrix entries g. So let (V, Π) be a representation and take $g = \Pi_{jk}$. Note that $\Pi_{jk}(x) = \operatorname{Tr}(\Pi(x)E_{kj})$ (where E_{kj} is the matrix with a 1 in the kj position and 0s everywhere else, in the basis used to determine Π_{jk}). Hence

$$h(x) = \int_K \operatorname{Tr}(\Pi(uxu^{-1})E_{kj})\,du = \int_K \operatorname{Tr}(\Pi(x)\Pi(u)^{-1}E_{kj}\Pi(u))\,du = \operatorname{Tr}(\Pi(x)A)$$

where $A = \int_K \Pi(u)^{-1}E_{kj}\Pi(u)\,du$. Note that, for $x \in K$,

$$\Pi(x)^{-1}A\Pi(x) = \int_K \Pi(ux)^{-1}E_{kj}\Pi(ux)\,du = \int_K \Pi(y)^{-1}E_{kj}\Pi(y)\,du = A.$$

So A commutes with the representation, meaning that it is an intertwining map. Since the representation is irreducible, by Schur's lemma, $A = \lambda I$ for some constant $\lambda \in \mathbb{C}$. Hence, $h(x) = \lambda \operatorname{Tr}(\Pi(x)) = \lambda\chi_\Pi(x)$ is a scalar multiple of an irreducible character. □

PROPOSITION 18.33. *The linear span of irreducible characters is dense in $C(K)_{\text{class}}$.*

PROOF. Let $f \in L^2(K)_{\text{class}}$. By Proposition 18.31, there is a sequence of representative functions g_n with $\|g_n - f\|_\infty \to 0$. Now, let

$$f_n(x) = \int_K g_n(uxu^{-1})\,du.$$

By Lemma 18.32, f_n is a linear combination of irreducible characters. We can compute that, for each x,

$$\begin{aligned}|f_n(x) - f(x)| = \left|\int_K g_n(uxu^{-1})\,du - f(x)\right| &= \left|\int_K [g_n(uxu^{-1}) - f(uxu^{-1})]\,du\right| \\ &\le \int_K |g_n(uxu^{-1}) - f(uxu^{-1})|\,du \\ &\le \|g_n - f\|_\infty\end{aligned}$$

where we used the fact that f is a class function in the second equality. Hence, $\|f_n - f\|_\infty \le \|g_n - f\|_\infty \to 0$, which shows that $f_n \to f$ uniformly. Since f_n is a linear combination of irreducible characters, this proves the result. □

PROPOSITION 18.34. *The linear span of irreducible characters is dense in $L^2(K)_{\text{class}}$; in particular, the characters of all pairwise non-isomorphic irreducible representations form an orthonormal basis for $L^2(K)_{\text{class}}$.*

PROOF. Since K is compact and the Haar measure is normalized, the L^2 norm is $\leq$ the uniform norm; thus, if f_n is a sequence of linear combinations of characters that converges uniformly to a given $f \in L^2(K)_{\text{class}}$ (cf. Proposition 18.33), it also converges in L^2, proving the first statement. The second is just the statement that the given set is orthonormal and its span is dense; this follows from the first statement and the orthogonality relations of Theorem 18.14. □

EXAMPLE 18.35. Consider the Lie group $\mathrm{SU}(2)$. As we showed in Section 3, the irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$ are (up to isomorphism) precisely the representations $(V_m, (\pi_m)_{\mathbb{C}})$ of Example 17.12. This, in turn, shows that the representations (V_m, Π_m) of Example 17.6 are (up to isomorphism) all of the irreducible representations of $\mathrm{SU}(2)$ (this follows from Lemmas 17.16, 17.22, 17.24, and 17.25). As a reminder, the irreps (V_m, Π_m) are given by $V_m = \mathbb{C}^m[z_1, z_2]$ (the $m+1$ $\mathbb{C}$-dimensional space of degree m homogeneous polynomials in two complex variables) and

$$[\Pi_m(U)f](z_1, z_2) = f(U^{-1}[z_1, z_2]^\top), \qquad U \in \mathrm{SU}(2).$$

Let us now compute the characters $\chi_m(U) = \operatorname{Tr}[\Pi_m(U)]$ of these representations. First, fix the standard maximal torus T of $\mathrm{SU}(2)$:

$$T = \left\{ t_\theta \equiv \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} : \theta \in [0, 2\pi) \right\}.$$

By the torus Theorem 16.17, for every $U \in \mathrm{SU}(2)$ there is some $V \in \mathfrak{GU}(2)$ with $U = V t_\theta V^{-1}$ for some $t_\theta \in T$. (In this case, this follows easily from the spectral theorem the spectral theorem.) Since characters are class functions, it follows that $\chi_m(U) = \chi_m(t_\theta)$ – i.e. they only depends on U through their eigenvalues. So we need only compute the characters restricted to T. We have

$$[\Pi_m(t_\theta)f](z_1, z_2) = f\left(\begin{bmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) = f(e^{-i\theta} z_1, e^{i\theta} z_2).$$

Fixing the usual basis $\{f_k(z_1, z_2) = z_1^{m-k} z_2^k : 0 \leq k \leq m\}$ for $\mathbb{C}^m[z_1, z_2]$, the action of $\Pi_m(t_\theta)$ on f_k is

$$[\Pi_m(t_\theta)f_k](z_1, z_2) = (e^{-i\theta} z_1)^{m-k} (e^{i\theta} z_2)^k = e^{i(2k-m)\theta} z_1^{m-k} z_2^k,$$

i.e. $\Pi_m(t_\theta) f_k = e^{i(2k-m)\theta} f_k$. So the standard basis diagonalizes $\Pi_m(t_\theta)$, and so its trace is the sum of the eigenvalues

$$\chi_m(t_\theta) = \operatorname{Tr}[\Pi_m(t_\theta)] = \sum_{k=0}^{m} e^{i(2k-m)\theta} = e^{-im\theta} + e^{-i(m-2)\theta} + \cdots + e^{i(m-2)\theta} + e^{im\theta}.$$

We can write this more simply by summing the geometric series:

$$\chi_m(t_\theta) = e^{-im\theta} \sum_{k=0}^{m} e^{2ik\theta} = e^{-im\theta} \frac{(e^{2i\theta})^{m+1} - 1}{e^{2i\theta} - 1}.$$

Cleverly writing $e^{-im\theta} = e^{-i(m+1)\theta}/e^{-i\theta}$, we have

$$\chi_m(t_\theta) = \frac{e^{i(m+1)\theta} - e^{-i(m+1)\theta}}{e^{i\theta} - e^{-i\theta}} = \frac{\sin((m+1)\theta)}{\sin\theta}.$$

These are the irreducible characters of $\mathrm{SU}(2)$. Hence, by the Peter-Weyl theorem, these from a Hilbert space basis for all L^2 class functions on $\mathrm{SU}(2)$.

4. Compact Lie Groups are Matrix Groups

We now conclude the course with one application if the Peter-Weyl theorem, stated as the subject of this section. First, we need the following lemma, which applies to Lie groups in general.

LEMMA 18.36. *Let G be a Lie group. There is a neighborhood U of the identity $e \in G$ such that $\{e\}$ is the only subgroup contained in U.*

PROOF. Fix any norm on $\mathfrak{g} = \mathrm{Lie}(G)$, and let $\epsilon > 0$ be small enough that $\exp$ is a diffeomorphism from $B_\epsilon(0) \subset \mathfrak{g}$ onto its image in G. Let $0 < \delta < \epsilon/2$, and set $U = \exp(B_\delta(0))$; since $\exp$ is a diffeomorphism on a neighborhood of $B_\delta(0)$, U is an open neighborhood of $\exp(0) = e$. Assume, for a contradiction, that there is a nontrivial subgroup H of G with $H \subseteq U$. Let $h \in H \setminus \{e\}$; then there is a unique $X \in B_\delta(0) \subset \mathfrak{g}$ with $\exp X = h$. Since $h \neq e$, $X \neq 0$, and so for all sufficiently large n, $nX \notin B_\delta(0)$. Let n_0 be the largest integer with $n_0 X \in B_\delta(0)$; note $n_0 \geq 1$. Then $(n_0+1)X \notin B_\delta(0)$, but

$$\|(n_0+1)X\| = \frac{n_0+1}{n_0}\|n_0 X\| < \frac{n_0+1}{n_0}\delta \leq 2\delta < \epsilon$$

and so $(n_0+1)X \in B_\epsilon(0)$. Now, $\exp((n_0+1)X) = h^{n_0+1} \in H \subseteq U$, and $U = \exp(B_\delta(0))$, so there is some $Y \in B_\delta(0)$ such that $\exp((n_0+1)X) = \exp Y$. Since both $(n_0+1)X$ and Y are in $B_\epsilon(0)$ where $\exp$ is a one-to-one, it follows that $(n_0+1)X = Y$. But this is a contradiction, since $Y \in B_\delta(0)$ by $(n_0+1)X \notin B_\delta(0)$. □

Now, on the other hand, in a compact Lie group, we can use the Peter-Weyl theorem to construct a representation whose kernel is contained in any given neighborhood of e.

LEMMA 18.37. *Let K be a compact Lie group, and let U be a neighborhood of $e \in K$. There is a representation (V, Π) of K such that $\ker \Pi \subseteq U$.*

PROOF. Let $x \in K \setminus U$. Since $e \notin K \setminus U$, Corollary 18.29 (which states that the representative functions of K separate points) shows that there is a representative function f such that $f(x) \neq f(e)$. By Proposition 18.5, f is a linear combination of matrix entries, and hence there is at least one matrix entry that separates x and e, and so it follows that there is a representation (which we call Π_x) with $\Pi_x(x) \neq \Pi_x(e) = I$. By continuity of Π_x, it follows that there is a neighborhood V_x of x such that $\Pi^x(y) \neq I$ for $y \in V_x$.

The neighborhoods $\{V_x \colon x \in K \setminus U\}$ cover $K \setminus U$. Since K is compact and U^c is a closed subset of K, $K \setminus U = K \cap U^c$ is compact. Hence, there is a finite subcover $\{V_{x_1}, \ldots, V_{x_n}\}$, and so there are finitely many representations $\Pi_{x_1}, \ldots, \Pi_{x_n}$ so that, for each $y \in K \setminus U$, there is at least one Π_{x_j} with $\Pi_{x_j}(x) \neq I$. Thus, we may take $\Pi = \Pi_{x_1} \oplus \cdots \oplus \Pi_{x_1}$; for each $y \in K \setminus U$, $\Pi(y) \neq I \oplus \cdots \oplus I$ (which is the identity for the direct sum), so $y \notin \ker(\Pi)$. □

Hence, we are positioned to prove:

THEOREM 18.38. *Any compact Lie group K has a faithful unitary representation $K \hookrightarrow \mathrm{U}(n)$ for some finite n.*

PROOF. Fix a neighborhood U of the identity in K as in Lemma 18.36. By Lemma 18.37, there is a representation (V, Π) of K with $\ker \Pi \subseteq U$. But $\ker \Pi$ is a subgroup of K, and hence, by construction of U, $\ker \Pi = \{e\}$. Thus, (V, Π) is a faithful representation. Since K is compact, by Lemma 17.32 there is an inner product on V with respect to which Π is a unitary representation. Identifying V with $\mathbb{C}^n$ for $n = \dim_{\mathbb{C}}(V)$ by sending some orthonormal basis of V to the standard basis of $\mathbb{C}^n$, we identify K as a closed subgroup of $\mathrm{U}(n)$. □

REMARK 18.39. Virtually everything we've done in this chapter before the present section applies much more generally than to Lie group: the Peter-Weyl theorem holds for compact Hausdorff topological groups (and continuous representations). This is also true of Lemma 18.37: continuous representations of compact Hausdorff topological groups can have kernels contained in arbitrarily small neighborhoods of the identity.

However, Lemma 18.36 explicitly used Lie theory, and this *cannot* be avoided: Theorem 18.38 does not hold for *any* non-Lie group! After all, by the closed subgroup Theorem 14.5, since $\mathrm{U}(n)$ is a Lie group, if $K \hookrightarrow \mathrm{U}(n)$ is a closed subgroup, it is a Lie group.

Hence, we see that Lemma 18.36 must fail for any compact Hausdorff topological group that is not a Lie group: all such groups have arbitrarily small nontrivial subgroups.

Bibliography

[1] CONWAY, J. B. *A course in functional analysis*, second ed., vol. 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1990.

[2] HALL, B. C. *Lie groups, Lie algebras, and representations*, vol. 222 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003. An elementary introduction.

[3] LEE, J. M. *Introduction to Smooth Manifolds*, Second ed., vol. 218 of *Graduate Texts in Mathematics*. Springer, New York, 2013.

[4] MILNOR, J. Analytic proofs of the "hairy ball theorem" and the Brouwer fixed-point theorem. *Amer. Math. Monthly 85*, 7 (1978), 521–524.

[5] MUNKRES, J. R. *Topology: a first course*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

[6] RUDIN, W. *Principles of Mathematical Analysis*, Third ed. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976. International Series in Pure and Applied Mathematics.